

Comparing the Outcomes of Face-to-Face and Synchronous Online Research Mentor Training Using Propensity Score Matching

Jenna Rogers,^{1,2,*} Xue Gong,^{2,3} Angela Byars-Winston,^{4,5} Melissa McDaniels,¹ Nancy Thayer-Hart,⁵ Philip Cheng,⁶ Kelly Diggs-Andrews,⁷ Kermin J. Martínez-Hernández,⁸ and Christine Pfund,^{1,9,10}

¹Wisconsin Center for Education Research, University of Wisconsin–Madison, Madison, WI 53706;

²Department of Educational Leadership and Policy Analysis, University of Wisconsin–Madison, Madison, WI 53706;

³Department of Medicine, University of Wisconsin–Madison, Madison, WI 53715;

⁴Center for Women's Health Research, University of Wisconsin–Madison, Madison, WI 53715;

⁵Thomas Roth Sleep Disorders and Research Center, Henry Ford Health, Detroit, MI 48202;

⁶Diggs-Andrews Consulting LLC, Ashburn, VA 20147; ⁷Chemistry Department, St. John Fisher University, Rochester, NY 14618; ⁸Institute for Clinical and Translational Research, University of Wisconsin–Madison, Madison, WI 53705

ABSTRACT

In this study, propensity score matching (PSM) was conducted to examine differences in the effectiveness of research mentor training (RMT) implemented using two modes—face-to-face or synchronous online training. This study investigated each training mode and assessed participants' perceived gains in mentoring skills, ability to meet mentees' expectations, and overall quality of mentoring, as well as intention to make changes to their mentoring practices. Additional factors that may contribute to participant outcomes were also examined. In total, 152 mentors trained using a synchronous online platform and 655 mentors trained in in-person workshops were analyzed using the PSM method. Mentors were matched based on similar characteristics, including mentee's career stage, mentor's title, mentor's prior mentoring experience, mentor's race/ethnicity and sex, and mentor's years of experience; results show that both face-to-face and synchronous online modes of RMT are effective. Findings indicated that the training mode did not significantly impact the mentors' perceived training outcomes. Factors associated with the reported training outcomes included dosage (hours of training), facilitator effectiveness, race/ethnicity, and previous mentoring experience. The results of this study demonstrate that mentors' perceived training outcomes are comparable regardless of the training modality used—online versus face-to-face.

COMPARING THE OUTCOMES OF FACE-TO-FACE AND SYNCHRONOUS ONLINE RESEARCH MENTOR TRAINING USING PROPENSITY SCORE MATCHING

Mentoring relationships play a critical role in the talent development of emerging professionals (Nagda *et al.*, 1998; Dolan and Johnson, 2009; Junge *et al.*, 2010; Poodry and Asai, 2018). In academic research labs, in which training is largely grounded in the cognitive apprenticeship model (Brown *et al.*, 1989; McGee, 2016), research mentors help mentees form science identity (Chemers *et al.*, 2011), build research efficacy (Byars-Winston *et al.*, 2015), and contribute to academic persistence (Haeger and Fresquez, 2016). For example, junior and senior undergraduates who report better-quality mentorship are more likely to also report stronger science efficacy, identity, and values (Estrada *et al.*, 2018). Similarly, for graduate students, mentor support is associated with positive academic self-concept, low levels of stress, and career commitment (Ulku-Steiner *et al.*, 2000).

Mark Barsoum, *Monitoring Editor*

Submitted Dec 14, 2021; Revised Jun 28, 2022; Accepted Jul 27, 2022

CBE Life Sci Educ December 1, 2022 21:ar62

DOI:10.1187/cbe.21-12-0332

²These authors contributed equally to this work.

¹⁰In full disclosure, there is a connection between one of the authors and a commercial product being used and reported. Specifically, Christine Pfund, is an author of one or more of the published research mentor training curriculum described in this article.

*Address correspondence to: Jenna Rogers (jrogers22@wisc.edu).

© 2022 J. Rogers, X. Gong, *et al.* CBE—Life Sciences Education © 2022 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 4.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/4.0>).

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

Despite the importance of effective mentoring relationships in the research experience, most mentors receive no type of formal training, leading them to develop a variety of mentoring approaches that vary in consistency, intensity, and effectiveness (Straus *et al.*, 2013). This lack of standards, necessary knowledge and skills, and mentoring professional development on the part of the mentors has led to negative experiences for mentees (Dolan and Johnson, 2010; Thiry and Laursen, 2011; Limeri *et al.*, 2019; Tuma *et al.*, 2021). In a national survey of faculty, only 7% of mentors reported significant training in mentoring students (Stolzenberg *et al.*, 2019). In recognition of this unmet need, the National Academies of Sciences, Engineering, and Medicine (NASEM) produced a consensus study report, *The Science of Effective Mentorship in STEMM* (science, technology, engineering, mathematics, and medicine), that lays out the scholarship on mentorship to inform and guide effective mentorship in academia (NASEM, 2019). The report defines mentorship as “a professional, working alliance in which individuals work together over time to support the personal and professional growth, development, and success of the relational partners through the provision of career and psychosocial support” (NASEM, 2019, p. 37). Among the nine recommendations in the report is a call for institutional leaders to use evidence-based mentoring practices, which include tested mentorship education.

Mentorship Education: Entering Mentoring

One model of evidence-based mentorship education is *Entering Mentoring* (Handelsman *et al.*, 2005; Pfund *et al.*, 2006, 2015a), a process-based research mentor training (RMT) curriculum developed to improve mentorship behaviors of both new and seasoned researchers. The *Entering Mentoring* curriculum features six primary competencies, including 1) maintaining effective communication; 2) establishing and aligning expectations; 3) assessing mentees’ understanding of scientific research; 4) addressing diversity within mentor–mentee relationships; 5) fostering mentee independence; and 6) promoting mentee career development. The duration of this curriculum ranges typically between 4 and 8 hours. Because of its process-based approach, in which a series of steps are systematically followed, successful implementation of *Entering Mentoring* requires a skilled facilitator to cultivate a learning environment where participants feel connected to and are open to learning from one another. *Entering Mentoring* (Pfund *et al.*, 2015a) was established at the University of Wisconsin–Madison and is widely available and used nationwide in institutions of higher education, organizations, and non-college workplaces. The curriculum has since been adapted for faculty mentors of graduate students, junior faculty, and postdoctoral scholars. It has been adapted multiple times for research mentors across the STEM disciplines, as well as within medicine and public health, for mentors of mentees at various career stages (Branchaw *et al.*, 2011; Asquith *et al.*, 2014; House *et al.*, 2014; Pfund *et al.*, 2013a,b).

The efficacy of the *Entering Mentoring* curriculum has also been tested. In a 2014 double-blind trial (Pfund *et al.*, 2014), mentors were randomized to either an intervention group (*Entering Mentoring* curriculum) or a control group. Compared with the control group, mentors in the intervention group reported larger perceived gains in mentorship skills across the

six primary competencies of the entering mentoring curriculum (listed earlier), as measured by the Mentoring Competency Assessment (MCA; Fleming *et al.*, 2013). In addition, qualitative results from the study showed a greater number of mentors in the treatment group reporting specific changes in their mentoring behavior, 123/141 (87%) compared with only 57/136 (42%) in the control group, indicating an impact of the training beyond reported skill gains. Consistent results were also reported by the mentees, with 95/140 (68%) of mentees whose mentors were in the treatment group reporting at least one positive change in their mentors’ behavior as compared with 77/135 (57%) in the control group. Mentees also reported better experiences working with the trained mentors. These results suggest that a significant perceived change in mentorship skill assessment may translate to meaningful improvements in mentoring practices (Pfund *et al.*, 2014).

The *Entering Mentoring* curriculum and its subsequent adaptations have been used to train thousands of mentors across the country (Pfund *et al.*, 2015b; Rogers *et al.*, 2018; Spencer *et al.*, 2018). Evaluations of these trainings consistently show an increase in self-reported mentoring skills as well as overall mentoring quality, despite the length of the training or previous mentoring experience of participants (Rogers *et al.*, 2020). Studies that assessed mentoring behaviors a few months post-training found that mentors still report using some of the suggested practices and behaviors from RMT, including strategies recommended in the training such as articulating expectations; developing a written mentoring philosophy; and using tools such as mentoring compacts and individual development plans to help improve communication, address diversity, and align expectations (Pfund *et al.*, 2014; House *et al.*, 2018; Trejo *et al.*, 2022).

Since its publication, *Entering Mentoring* has been converted to a synchronous online learning experience to enhance its scalability relative to face-to-face implementation. The efficacy of *Entering Mentoring* when delivered in a synchronous online environment has also been evaluated (McDaniels *et al.*, 2016). In addition to perceived gains in mentorship skills, the impact of the synchronous online technologies on participant experience of the learning community was examined. Participants reported both satisfaction and significant confidence gains in their ability to engage in high-quality mentorship, address diversity and difference in mentoring relationships, and cultivate learning communities among mentees on a research team. Participants also noted that the online learning environment was inclusive and found that their peers played essential roles in the learning experience. Additionally, participants reported that the technological tools in the synchronous environment (chat rooms, whiteboards, etc.) increased course engagement, noting that these tools are lacking in face-to-face contexts. Barriers and drawbacks reported by participants mirror those identified in the online learning literature, including technological access and reliability.

Comparison of Online and Face-to-Face Training

Though the research literature on online or distance education in higher education is broad, much of the extant literature focuses on asynchronous training (Mallonee *et al.*, 2017). Contrasted with asynchronous training, synchronous training allows instructors and students to engage in real time without

geographic limitations and leverages the affordances of chat and/or audio- and/or video-conferencing. There is less published on the efficacy of synchronous online learning and an even smaller base of research comparing participant outcomes of face-to-face versus synchronous online courses. This paper seeks to expand the research in this particular area.

Synchronous Online Training and Instruction

Synchronous online training or instruction enables instructors and students to learn at the same time while in different places by taking advantage of video-conferencing platforms. With the increasing accessibility and use of Web-conferencing platforms such as Blackboard, Zoom, and Webex both before and since the COVID-19 pandemic, the need for studies of training outcomes of synchronous online training modules has increased. In their systematic review of published reports on synchronous online learning, Martin *et al.* (2017) examined scholarly articles of original research published from 1994 to 2014 that: 1) examined the use of synchronous online learning for professional development or teaching purposes, 2) had identifiable methods and results sections, and 3) were written in English. Of the 157 articles in the final sample, the median sample size was 34 (range: 1–6321 participants), with the majority of studies having fewer than 100 participants. Dependent variables across these studies included perceptions/attitudes of the tool or course, engagement, and willingness to communicate. Additionally, questionnaires, session transcripts, exams, interviews/focus groups, and observations were used as the primary data sources in these studies. Of the 157 studies examined in this systematic review, none directly compared face-to-face and synchronous online learning.

Efficacy of Face-to-Face versus Synchronous Online Training

Since the publication of Martin *et al.*'s (2017) systematic review, a handful of studies have been conducted, that specifically examine the similarities and differences in training outcomes between synchronous online and face-to-face delivery methods. Mullin *et al.* (2016) evaluated outcomes of synchronous online versus face-to-face delivery of a 20+-hour-long motivational interviewing (MI) course offered for clinicians by the University of Massachusetts Medical School. All 34 participants completed all course requirements, with 20 being in-person and 14 online. As a part of this training, participants received 2 hours of individual MI practice and feedback, which provided the comparative data for analysis for this study. Using a motivational interviewing treatment integrity observational protocol and closed-ended self-evaluation questionnaire, the authors no significant differences in MI skills in the outcomes for synchronous online participants and face-to-face participants. However, the authors noted that the small sample size might not have had the power to identify these differences, and the participants were not randomized between treatment groups.

Faulconer *et al.* (2018) compared outcomes for 1964 college students in a face-to-face, synchronous video, and asynchronous introductory physics course. Students who took the course through synchronous and asynchronous video passed at higher rates (96.80% and 95.98%, respectively) than those who took the class in-person (90.99%). However, the authors noted limitations, including small effect sizes and the inability to control

all moderating variables, such as student age as a potential factor influencing these observations.

Propensity Score Matching

To examine the differences in face-to-face and synchronous online learning and overcome some of the limitations of previous studies, we undertook a propensity score matching (PSM) method to explore the effectiveness of RMT between two different training modes. In the absence of randomization, when participants are placed into conditions, quasi-experimental designs are usually subject to selection bias by comparing differences in the treatment effect between groups. As such, the observed effect might be due to differences in participants from different conditions rather than or in addition to the intervention (Shadish and Steiner, 2010). Thus, groups from different conditions may not be comparable at baseline. PSM helps account for this bias by using regression techniques to predict group assignments from predetermined and theoretically relevant covariates, followed by matching participants in different groups based on these predicted propensity scores (Lane *et al.*, 2012). Many reports have empirically justified the benefit of using PSM. Dehejia and Wahba (1999, 2002) used the PSM method to analyze data from the National Supported Work experiment and examine the effect of a labor training program on postintervention earnings. From the methodological lenses, the researchers paired the experimental-treated units with a small subset of nonexperimental comparison units that were most comparable in observed characteristics to the treated units, alleviating bias due to systematic differences between the treated and comparison units. Subsequently, they compared estimates of the treatment effect obtained using the PSM method to benchmark results from the experiment. The results showed that the PSM method is viable to yield accurate estimates of the treatment effect in nonexperimental settings, indicating that PSM can reasonably replicate experimental impact estimates.

The Present Study

This study uses a PSM method to examine differences in the effectiveness of RMT between two different training modes—face-to-face or synchronous online. The efficacy of online training as a feasible alternative option or supplementation for in-person mentor training is especially important given the growing need for online mentorship education due to COVID-19.

The specific research questions addressed in this study include:

1. Do mentors' perceived gains in mentoring skills, overall quality of mentoring provided, and ability to meet mentees' expectations differ by training platform—online versus in-person?
2. How do intended changes to their mentoring practices differ by training modality for mentors who took online versus in-person training?

METHODS

Data Collection

Face-to-Face Training Modality. At the end of each *Entering Mentoring* training implementation, a survey link was sent to participants through Qualtrics, an online survey platform,

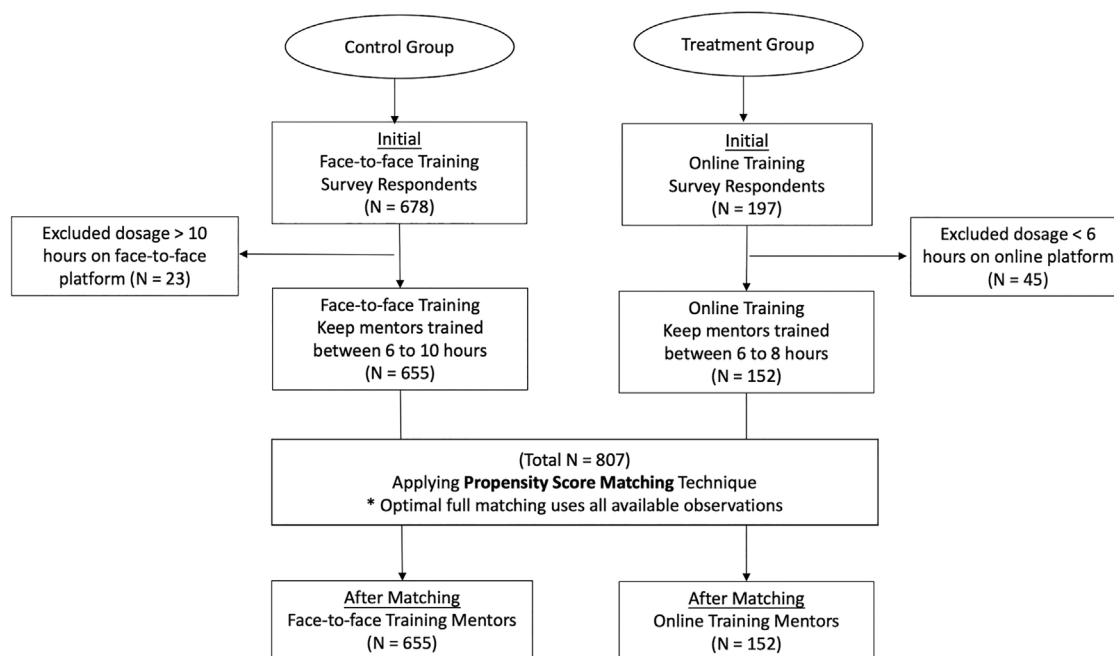


FIGURE 1. Flowchart of the changes in sample size.

asking them about their perceived skill gains and satisfaction. Participants were given 2–3 weeks to complete the survey, with two reminders sent during this period. Beginning in October 2015, data were collected using a more systematic approach overseen by the Research, Tracking, Assessment, and Evaluation team at this university. This team offers free, customized evaluation services to all who implement mentor training. Through the use of this evaluation service and the earlier website with the open survey link, data were collected from hundreds of different nationwide implementations of RMT since 2015, with an average response rate of 72%. Data for the present study came from 40 different RMTs with a minimum length of 6 hours implemented across the country at different universities between 2015 and 2018. The 6-hour duration was chosen to be comparable to the online training duration described in the next section. There were 678 mentors who attended the face-to-face training and completed the survey.

Online Training Modality. The online training study population consisted of research mentors (either faculty members or co-mentors; e.g., postdoctoral fellows, graduate students) who work closely with an undergraduate student (any stage/year) participating in a summer 2017 STEM research-oriented program (National Science Foundation REU, National Institutes of Health [NIH] MARC, etc.). Mentors self-selected to participate in four 2-hour sessions (8 hours total) of online RMT led by experienced facilitators trained in the curriculum. Mentors were eligible if they were actively mentoring an undergraduate student researcher during June and July of 2017 and if their relationships with their mentees began after January 1, 2017. There were 216 mentors enrolled in the training at the start of the study. However, attrition before the first training session resulted in 197 participants ultimately being included in the study.

The online training was offered in 16 sections across a range of days of the week and times of day, beginning the week of May 22, 2017, and ending the week of July 17, 2017. Each training section had its own Blackboard Ultra “classroom,” and materials for each training session were posted on the section’s Moodle site. In addition to the pair of facilitators leading each 2-hour training session, there was a technology support person online to help participants resolve any technical issues encountered during the session. The Blackboard Ultra technology enabled audio and video participation, chat conversations, whiteboard notes, slide presentations, and breakout groups.

Time was set aside at the end of the fourth and final training session for participants to respond to an online questionnaire about their experiences in the RMT.

Sample

A flowchart is provided to illustrate the changes in sample size at each stage of data analysis (Figure 1). The initial sample contains 678 survey respondents from a group of mentors who were trained in person during the years 2015–2018 and 197 survey respondents from a group of mentors who were trained through the online platform in Summer 2017. As the hours of training vary in the face-to-face training group and the online training group, we further constrained this sample to mentors who were trained between 6 to 10 hours through the face-to-face platform and mentors who were trained between 6 to 8 hours through the online platform. The hours of training were set to 6 hours and above as this represents a “full” RMT in which all six of the competencies in the curriculum were covered. As a result, we ended up with a final sample of 655 mentors who were trained in person and 152 mentors who were trained online. Of the sample, only 12% of mentors who were trained in person and 19% of mentors who were trained online had prior mentoring experience.

Measures

Treatment Indicator. We set the treatment variable as the mode of training in order to examine the differences in training outcomes between face-to-face training and online training—whether a mentor was trained in person or through the online platform. We specified the RMT participants trained online as the treatment group, and those who were trained in person were constructed as the comparison group.

Outcome Measures. We examined the effects of the mode of training on the effectiveness of RMT using four mentor training outcomes: perceived mentoring skill gains, perceived gains in overall mentoring quality, perceived gains in ability to meet mentees' expectations and intended changes to mentoring practices. See Table A1 in the Appendix in the Supplemental Material for the survey items used to construct the outcome measures.

The first outcome measure is based on mentors' self-reported mentoring skill gains and is measured on the previously validated MCA (Fleming *et al.*, 2013; Pfund *et al.*, 2014; Hyun *et al.*, 2022). This 26-item research mentoring skills inventory was designed to evaluate research mentoring skills in six areas (maintaining effective communication, aligning expectations, assessing understanding, addressing diversity, promoting professional development, and fostering independence). The questions for mentors were framed as “please rate how skilled you feel you are in the following areas,” and the response was based on a seven-point Likert-type scale in which 1 = not at all skilled, 4 = moderately skilled, and 7 = extremely skilled. The MCA was administered postintervention to all mentors. This posttest version of the MCA also included a “retrospective pretest” that asked respondents to reconsider and rerate mentors' baseline skills. Specifically, mentor participants were asked to rate their mentoring skills before (designated as “before”) and after the RMT (designated as “now”) for each MCA item (Pfund *et al.*, 2014). We then constructed the perceived gains in MCA composite scores using individual retrospective pretest and posttest MCA scores. This survey design follows the methodology used for assessing skills gains in the randomized controlled trial testing the *Entering Mentoring* curriculum (Pfund *et al.*, 2014). The MCA was also recently revalidated on a sample of RMT participants, including those in this study, and was found to accurately capture skill gains across the six competencies of the *Entering Mentoring* curriculum (Hyun *et al.*, 2022).

The second outcome measure was based on mentors' self-reported scores on the overall quality of the mentoring they were able to provide to mentees, thinking back to before the training and now, after the training. Participants responded using a seven-point Likert-type scale in which 1 = very low, 4 = average, and 7 = very high. The difference in the retrospective pretest quality score and the posttest quality score was again used to measure the perceived gains in the overall quality of mentoring provided by mentors.

The third outcome measure was based on mentors' self-rating of their ability to meet their mentees' expectations before and after the training. A similar seven-point Likert scale was used, in which 1 = not at all, 4 = moderately, and 7 = completely. Similarly, we computed the difference in the retrospective pretest expectation score and the posttest expectation score to measure the perceived gains in their ability to meet mentees' expectations.

The fourth outcome was based on a single question asking mentors “Have you made any, or do you plan to make any changes in your mentoring as a result of this training?,” with a response of “yes” or “no.” We used this binary variable to measure whether the training affects mentors' intended behavioral changes.

Table 1 provides the descriptive statistics of all variables from the treatment group (online training) and the comparison group (face-to-face training). The bivariate correlation matrix of all the outcome variables is presented in Table 2. These measures allow for comparisons to be made between different modes of training. The specific measures used on the RMT evaluations, although they are subjective, short-term measures, have also been shown to be powerful predictors of overall mentoring skill gains and actual behavioral changes (as reported by mentors themselves as well as their mentees), both in the short and long term (Pfund *et al.*, 2014; House *et al.*, 2018; Trejo *et al.*, 2022). Thus, they are important to examine in terms of differences in the effectiveness of RMT across modalities.

Covariates. We also collected information on mentors' experience and backgrounds, such as mentors' race/ethnicity, sex, title, previous mentoring experience, and number of years of mentoring experience. The race/ethnicity variable was recoded into two groups due to sample size, mentors from well-represented backgrounds and mentors from historically excluded racial/ethnic groups. The well-represented racial/ethnic group includes participants who reported as White or Asian, and not Hispanic. The historically excluded racial/ethnic group contains mentors who reported being in any of the following racial and ethnic categories: 1) American Indian or Alaska Native, 2) Black or African American, 3) Native Hawaiian or Other Pacific Islander, 4) other, and 5) any Hispanic or Latino group.

Other training-related information was available as well, including hours of training, the effectiveness of facilitators, and mentees' career stages (faculty, graduate, undergraduate, and none of the above three). Specifically, the effectiveness of facilitators is a five-point Likert scale anchored with 1 = very effective and 5 = very ineffective. When interpreting the impact of the effectiveness of facilitators in the *Results*, the very effective category (coded as 1) serves as the reference group. Thus, the coefficient of the other facilitator effectiveness categories (2 = effective, 3 = neither, 4 = ineffective, and 5 = very ineffective) are the differences in the predicted outcomes compared with the reference group.

Analytical Strategies

A propensity score is a conditional probability of a participant being assigned to a particular treatment or comparison group given a set of observed covariates (Rosenbaum and Rubin, 1983). In the context of this study, online mentor training served as the “treatment” group. For individuals who underwent in-person training, a propensity score provides a measure of the likelihood they would have participated in the online training, based on shared characteristics with those trained online. Mentors who were trained online and mentors who were trained in-person with similar characteristics, as summarized by the propensity score, can be matched by these means to create a comparison group. As a result, groups can be compared with one another, because systematic differences have

TABLE 1. Descriptive statistics of all variables from the treatment group (online training) and the comparison group (face-to-face training)

Categorical variables	Online		Face-to-face		All				
	N	%	N	%	N	%			
Race/ethnicity ^a									
Well represented	9	5.9	430	65.6	439	54.4			
Historically excluded	139	91.4	150	22.9	289	35.8			
Prefer not to answer	4	2.6	75	11.5	79	9.8			
Sex									
Female	82	53.9	326	49.8	408	50.6			
Male	62	40.8	253	38.6	315	39.0			
Other	3	2.0	2	0.3	5	0.6			
Prefer not to answer	2	1.3	12	1.8	14	1.7			
Title									
Faculty member	59	38.8	217	33.1	276	34.2			
Graduate student	50	32.9	223	34.0	273	33.8			
Postdoc	22	14.5	83	12.7	105	13.0			
Researcher or scientist	12	7.9	63	9.6	75	9.3			
Other	8	5.3	56	8.5	64	7.9			
Mentor's previous mentoring experience									
Yes	29	19.1	76	11.6	105	13.0			
No	109	71.7	322	49.2	431	53.4			
Facilitator effectiveness									
Very effective	80	52.6	282	43.1	362	44.9			
Effective	52	34.2	163	24.9	215	26.6			
Neither	6	3.9	16	2.4	22	2.7			
Ineffective	–	–	6	0.9	6	0.7			
Very ineffective	–	–	3	0.5	3	0.4			
Mentee's career stage									
Faculty	13	8.6	213	32.5	226	28.0			
Graduate student	57	37.5	257	39.2	314	38.9			
Undergraduate student	134	88.2	233	35.6	367	45.5			
None	1	0.7	29	4.4	30	3.7			
Continuous variables	Online			Face-to-face			All		
	N	Mean	SD	N	Mean	SD	N	Mean	SD
Covariates									
Dosage	152	7.36	0.94	655	8.58	1.34	807	8.35	1.36
Mentor's years of experience	138	7.19	7.59	409	7.47	8.61	547	7.40	8.36
Outcome measures									
Mentoring skill gains	130	1.01	0.59	545	0.87	0.71	675	675	0.89
Overall quality gains	128	1.14	0.68	404	1.17	0.87	532	532	1.16
Meeting expectation gains	129	0.91	0.80	357	0.93	0.90	486	486	0.92
Intended changes	138	1.07	0.25	418	1.06	0.23	556	556	1.06

^aThe race/ethnicity variable was computed using the original racial and ethnic categories. Well-represented racial/ethnic group includes participants who reported as White or Asian, and not Hispanic. Historically excluded racial/ethnic group contains mentors who reported as any of the following racial and ethnic categories: 1) American Indian or Alaska Native, 2) Black or African American, 3) Native Hawaiian or Other Pacific Islander, 4) other, and 5) any Hispanic or Latino group.

been controlled through the experimental nature of the design (Lane *et al.*, 2012).

An alternative approach to adjusting for group differences (face-to-face vs. online) on covariates is to include all covariates in the regression model when testing for group differences in the outcomes. In the Appendix in the Supplemental Material, we included the estimates using ordinary least squares (OLS) regression with all covariates (see Table A2). We also provided a comparison of the estimates from the OLS regression models

with selected covariates to those from the PSM models (see Table A3). Although results may not significantly differ in the two approaches, we decided to report the PSM technique and the corresponding results in the subsequent sections, as there are characteristics of propensity score analysis that make it an overall more appealing approach compared with OLS regression analysis. Propensity score approaches can separate design and analysis in the first step of estimating propensity scores and then adjust for group differences in the second step when

TABLE 2. Correlations of outcome variables

	Mentoring skill gains	Overall quality gains	Meeting expectation gains	Intended changes
Mentoring skill gains	1			
Overall quality gains	0.706**	1		
Meeting expectation gains	0.646**	0.648**	1	
Intended changes	-0.154**	-0.108*	-0.113*	1

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

analyzing the main outcome (Reeve *et al.*, 2008; Amoah *et al.*, 2020). Therefore, researchers can add many more variables in the propensity score model during the first step, even those of no relationship to the outcome of interest, to increase the likelihood of similar distributions of measured covariates across the groups. This allows researchers to simplify the regression model during the second step by only including the key factors of interest (Reeve *et al.*, 2008). Also, the greater reduction in confounding afforded by propensity score methods increases the probability of more valid estimates of the relationship between the treatment and outcome (Amoah *et al.*, 2020). When comparing the outcomes of mentors who were trained in different platforms, a limitation of the OLS regression analysis is that the observed differences are the result of both varying mentor characteristics as well as differences related to the assigned treatment (face-to-face vs. online platforms), making it challenging to distinguish the true impact of the exposure to treatment from other varying factors. However, this constraint is not applied to propensity score methods, as they are more likely to achieve a similar distribution of the observed baseline variables and more closely mimic what would be expected in a randomized experiment (Amoah *et al.*, 2020).

Implementing PSM requires the following steps: 1) data preparation, 2) selection of covariates, 3) estimation of propensity score, 4) application of selected PSM technique, 5) balancing tests to assess matching quality, and 6) estimation of treatment effects (Dehejia and Wahba, 2002; Schafer and Kang, 2008; Rojewski *et al.*, 2010; Shadish and Steiner, 2010; Melguizo *et al.*, 2011). Following data preparation and selection of appropriate covariates, a propensity score is estimated for each case using a logistic regression (e.g., logit model), wherein a binary outcome variable indicates treatment status. These PSM estimates are then used to match units from the treatment and comparison groups using specific matching techniques, such as optimal full matching, inverse propensity weighting, and regression estimation using propensity score (Shadish and Steiner, 2010). Further, the matching quality is assessed by examining the degree of overlap and common support between the propensity score distributions of treatment and comparison groups, as well as testing the balance on the distribution of relevant variables in both treatment and comparison groups (Rojewski *et al.*, 2010). Finally, the treatment effects are estimated. Unlike an ordinary randomized experiment that estimates the average treatment effect across the entire population of treated and untreated units, PSM methods only estimate the average treatment effect for the treated population (Shadish and Steiner, 2010). The six steps of implementing PSM, excluding an overview of data preparation (described in *Methods*), are described in the following sections. All analyses were conducted using the R program (R Core Team, 2019).

Selection of Covariates. For the analysis, propensity scores represent the probability of a mentor participating in the online training, given the variables selected to predict participation. It is crucial to choose the appropriate covariates for matching purpose, because failure to include important covariates can result in those variables being unbalanced between treatment and comparison groups, thus leading to biased estimation of treatment effects (Rosenbaum and Rubin, 1983; Shadish and Steiner, 2010). It is recommended that only variables related to both treatment selection (the decision to participate, but not participation itself) and outcomes of interest should be included in the propensity score model (Caliendo and Kopeinig, 2008; Shadish and Steiner, 2010). As the appropriate covariates should be those unaffected by participation, we need to choose variables that are either fixed over time or measured before participation (Caliendo and Kopeinig, 2008). In this study, six pretreatment covariates are available, including the mentor's sex, the mentor's race/ethnicity, the title of a mentor, the mentor's previous mentoring experience, the mentor's years of experience, and the mentee's career stage. All six covariates were selected in the propensity score model (see Table 3).

Because the inclusion of missing values results in an incomplete matching process, the first step after selecting appropriate covariates for matching was to deal with missing data for covariates. We applied different imputation methods to deal with missing values of different types of covariates. Specifically, we generated a new category for missing values of categorical variables to indicate missing values (NA). These variables include the title of a mentor, the sex of a mentor, and a mentor's previous mentoring experience. For missing values of the continuous covariate (a mentor's years of experience), we generated two new variables of years of experience: the first new variable replaced all missing values as 0, and the second new variable was a binary variable, with 1 indicating the missingness of years of experience and 0 representing that years of experience is not missing. A combination of the two new variables was used later in the propensity score model. In the Appendix in the Supplemental Material (Table A4), we included more details of missing data and the strategies to impute missing values of covariates.

We then checked on the imbalance of selected baseline covariates that were established before the matching process by examining whether statistically significant differences existed on covariates between the treatment (online) and comparison (face-to-face) groups. The results (see Table 4) indicated that mentors who were trained online and mentors who were trained in person were significantly different in terms of their race/ethnicity, sex, years of previous mentoring experience, years of experience, and mentees' career stages.

TABLE 3. Covariates for PSM

Variable	Variable type	Values
Mentee’s career stage: faculty	Binary	1 = yes, 0 = no
Mentee’s career stage: graduate	Binary	1 = yes, 0 = no
Mentee’s career stage: undergraduate	Binary	1 = yes, 0 = no
Mentee’s career stage: none	Binary	1 = yes, 0 = no
Title of mentor	Categorical	1 = faculty member 2 = graduate student 3 = postdoc 4 = researcher of scientist 5 = other 6 = NA
Mentor’s previous mentoring experience	Binary	1 = yes 2 = no
Race/ethnicity of mentor ^a	Categorical	1 = well represented 2 = historically excluded 3 = prefer not to answer
Sex of mentor	Categorical	1 = female 2 = male 3 = other 4 = prefer not to answer
Mentor’s years of experience	Continuous	

^aThe race/ethnicity variable was computed using the original racial and ethnic categories. Well-represented racial/ethnic group includes participants who reported as White or Asian, and not Hispanic. Historically excluded racial/ethnic group contains mentors who reported as any of the following racial and ethnic categories: 1) American Indian or Alaska Native, 2) Black or African American, 3) Native Hawaiian or Other Pacific Islander, 4) other, and 5) any Hispanic or Latino group.

Estimating of Propensity Score. We estimated the propensity score by running a logit model in which the outcome variable is a binary variable indicating treatment status. After calculating the propensity score for individual cases in both groups, we examined the level of common support, which requires that any combination of individual characteristics observed in the treatment group is also observed in the comparison group. This can be inferred by the overlap of the distribution of propensity score estimates of the two groups. The minima–maxima comparison method was used to enforce the common support requirement by deleting cases in each group that had a propensity score (in

the form of propensity score logits) outside the range evident in the other group. That is to say, for nonoverlapping cases, the treatment effect cannot be estimated (Rojewski *et al.*, 2010).

The range of propensity score logits was –8.414 to 4.041 for the online group and –5.433 to 4.041 for the face-to-face group. Deleting values without corresponding overlap resulted in a region of common support that ranged between –5.433 and 4.041 in terms of propensity score logits.

Application of a Selected PSM Technique. We chose the optimal full matching technique to match cases from control and

TABLE 4. Initial balance check and balance check after PSM of online versus face-to-face groups on selected covariates

Covariate	Initial balance				After balance			
	Mean diff ^a	SE	Std mean diff	Var ratio	Mean diff ^a	SE	Std mean diff	Var ratio
Mentee’s career stage (faculty)	–0.240***	0.040	–0.620	0.358	–0.132***	0.040	–0.318	0.713
Mentee’s career stage (graduate)	–0.017	0.044	–0.036	0.988	–0.121***	0.044	–0.250	0.964
Mentee’s career stage (undergraduate)	0.526***	0.041	1.286	0.458	–0.005	0.045	–0.009	1.145
Mentee’s career stage (none)	–0.038**	0.017	–0.241	0.155	–0.006	0.017	–0.030	0.982
Title	–0.225*	0.119	–0.176	0.781	–0.269**	0.114	–0.226	0.615
Mentor’s previous mentoring experience	–0.375***	0.057	–0.630	0.635	0.165***	0.062	0.228	1.291
Mentor’s years of experience	1.868***	0.690	0.245	0.954	–0.507	0.667	–0.068	0.951
Mentor’s years of experience (NA) ^b	–0.283***	0.041	–0.710	0.358	0.170***	0.043	0.338	1.323
Sex	–0.260**	0.101	–0.260	0.433	0.286***	0.101	0.245	1.167
Race/ethnicity	0.509***	0.057	0.960	0.178	0.076	0.061	0.105	1.311

^aMean difference refers to the difference in means of online and face-to-face groups (face-to-face is the reference group).

^bThis binary variable was generated to address the missing values of a mentor’s years of experience, with 1 indicating the missingness of years of experience and 0 representing that years of experience is not missing. Therefore, the mean of this variable refers to the percent of missing values.

**p* < 0.1.

***p* < 0.05.

****p* < 0.01.

comparison groups. Optimal full matching constructs strata consisting of one treated unit and at least one control unit or one control unit and at least one treated unit, without replacement (Hansen, 2004, 2007; Hansen and Klopfer, 2006; Stuart, 2010; Austin and Stuart, 2017). It is optimal in the sense that it minimizes the average within-stratum differences in the propensity score between treated and control units. Optimal full matching is a stratification approach and can be implemented by using individual stratum weights; thus the optimal full matching falls at the intersection of matching, subclassification, and weighting. Like subclassification, it involves the formation of strata with varying numbers of treated and control subjects (Austin and Stuart, 2017; Stuart, 2010). To be more specific, the optimal full matching technique allows us to subclassify all observations into homogeneous strata and then match treatment and comparison cases within each single stratum (Shadish and Steiner, 2010). Optimal full matching can also be seen as a form of propensity score weighting, as it incorporates weights that are derived from the stratification (Austin and Stuart, 2017). Therefore, optimal full matching has attractive features compared with other matching approaches and can avoid the drawbacks of other matching methods. It includes all subjects in the analytical sample without discarding any cases. This avoids bias due to incomplete matching in some conventional matching methods, such as nearest neighbor matching, when some subjects are excluded from the final matched sample (Austin and Stuart, 2017; Hansen, 2004). The subclassification approach can result in some residual bias in the treatment effect, as the within-stratum distributions of propensity scores slightly differ between the treatment and control groups, which is a drawback that optimal full matching can avoid (Steiner and Cook, 2013). Compared with weighting, it is less sensitive to the form of the propensity score model, because the original propensity scores are used to create the strata instead of forming the weights directly (Stuart, 2010). Also, unlike weighting, it is less sensitive to outliers that can lead to extremely large weights (Steiner and Cook, 2013).

To help understand how treated subjects were matched to control subjects using optimal full matching in our analysis, we provide a table in the Appendix in the Supplemental Material to illustrate the number of treated and control subjects in each stratum (Table A5). Analysis using the optional full matching technique was done using the *optmatch* package (v. 0.9–13; Hansen and Klopfer, 2006).

Balancing Tests to Assess Matching Quality. We examined how well the comparison group matched the treatment group through a series of balancing tests, including a check on balance, the balance plot after PSM in observed covariates, and the density plot for the logit of PSM. When assessing the balance check results (Table 4) and comparing the corresponding plots before and after applying optimal full matching (Figure 2A–C), we noticed that, after using optimal full matching, most observed covariates achieve the absolute standardized differences of means that are less than 0.25, and the variance ratios are between 0.5 and 2 (Rubin, 2001; Stuart and Rubin, 2008). Although two covariates, mentee’s career stage (faculty) and mentor’s years of experience (NA) ended up with the absolute standardized differences in means slightly higher than 0.25, their balance was greatly enhanced after applying optimal full

matching: from -0.620 to -0.318 for mentee’s career stage (faculty), and from -0.710 to 0.338 for mentor’s years of experience (NA). Thus, we are confident that the balance between treatment and comparison groups has been dramatically improved after the application of propensity score full matching.

Estimation of the Treatment Effects. After specifying the propensity score model using optimal full matching, we estimated the average treatment effects on the treated (ATT) with 1) propensity score weighting only and 2) propensity score weighting plus additional covariates. We specified the two types of models below to examine the treatment effects on mentor outcomes.

$$Y_i = \beta_0 + \tau Z + \varepsilon_i \quad (1)$$

(corresponding weights are W_{T_i} , W_{T_q} and W_{T_m})

$$Y_i = \beta_0 + \tau Z + \beta_1 * \text{dosage} + \beta_2 * \text{Facilitator Effect} + \beta_3 * \text{race} + \beta_4 * \text{gender} + \beta_5 * \text{Previous MentorExp} + \varepsilon_i \quad (2)$$

(corresponding weights are W_{T_i} , W_{T_q} and W_{T_m})

where Y_i is mentor i ’s outcome (MCA skill gains, the gains in overall quality of mentoring provided mentors, the gains in meeting mentees’ expectations, and mentors’ behavioral changes) and z is the treatment indicator, with 1 = “Online” and 0 = “In-Person (face-to-face).”

RESULTS

Perceived Gains in Mentoring Skills

Perceived gains in mentoring skills are shown in columns 2 and 3 of Table 5 as the ATT group with and without additional covariates. Results showed no significant differences in perceived gains in mentoring skills between mentors trained online or in-person ($\beta = 0.087$, $p = 0.087$). Following quantification with additional covariates, the training platform (online vs. in-person) was not a significant factor contributing to the perceived mentoring skill gains ($\beta = 0.114$, $p = 0.206$). However, the hours of training (dosage), facilitator effectiveness, and the mentor’s race/ethnicity, sex, and previous mentoring experience were highly predictive in estimating perceived gains in mentoring skills. Specifically, higher gains were found with increased facilitator effectiveness. Regarding race/ethnicity, historically excluded racial/ethnic groups showed greater perceived gains in mentoring skills than well-represented groups ($\beta = 0.102$, $p = 0.064$). In addition, mentors without previous mentoring experience had higher reported mentoring skill gains than mentors with previous mentoring experience ($\beta = 0.319$, $p < 0.001$). In terms of sex, participants who did not report their sex had significantly lower perceived gains in mentoring skills than those who reported as female ($\beta = -0.569$, $p = 0.027$).

Perceived Gains in Overall Quality of Mentoring

ATT group in terms of perceived gains in overall quality of mentoring with and without additional covariates are shown in columns 4 and 5 of Table 5. No significant differences in perceived gains of overall quality of mentoring were found between mentors trained online or in-person (without covariates: $\beta = 0.044$, $p = 0.676$; with covariates: $\beta = 0.128$, $p = 0.247$). Considering additional covariates, the hours of training (dosage), facilitator effectiveness, and the mentor’s race/ethnicity and previous

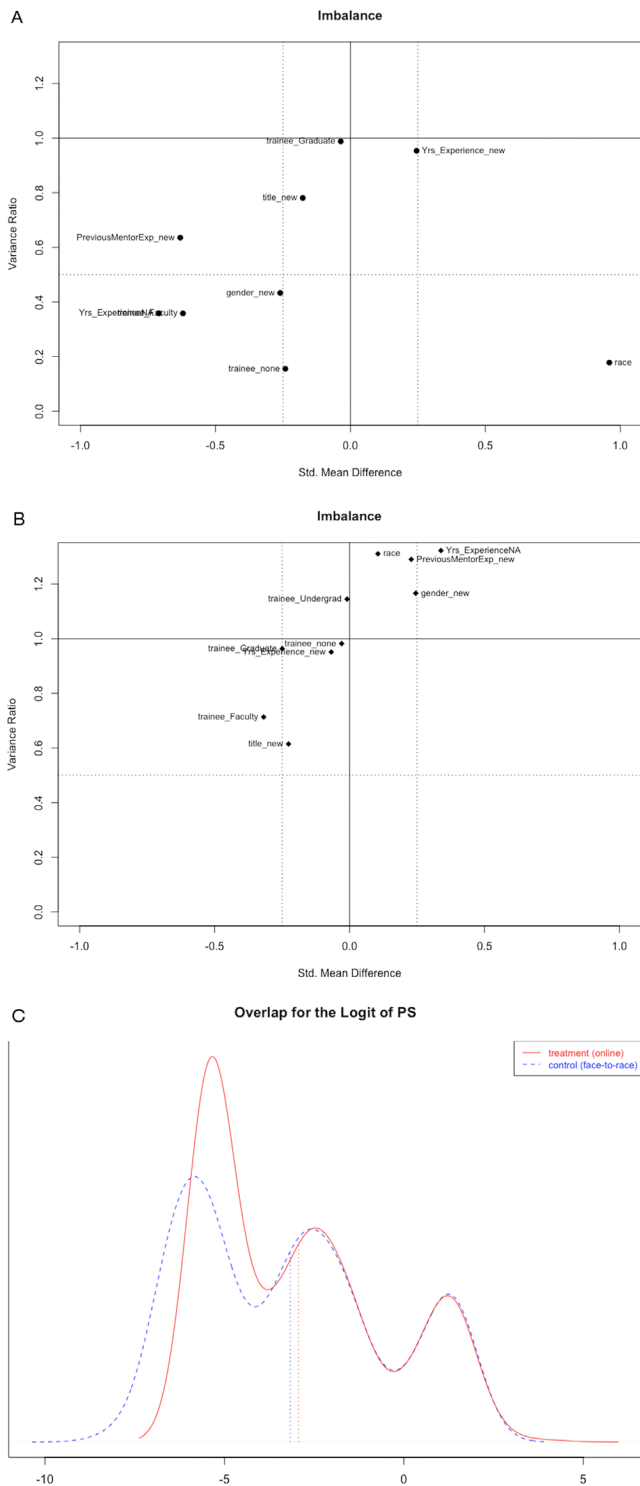


FIGURE 2. (A) Initial balance in observed covariates. The dashed lines represent Rubin’s benchmark: 1) standard differences in means less than 0.25; and 2) the variance ratios between 0.5 and 2 (Rubin, 2001; Stuart and Rubin, 2008). The covariates represented in the figure are mentees’ career stages: faculty (trainee_Faculty), graduate (trainee_Graduate), undergraduate (trainee_Undergrad), and none (trainee_none); mentor’s title (title_new); mentor’s previous mentoring experience (PreviousMentorExo_new); mentor’s years of experience (Yrs_Experience_new) and mentor’s

mentoring experience were significantly associated with perceived gains in overall quality of mentoring. Specifically, the hours of training and the facilitator’s effectiveness were positively associated with perceived gains in overall quality (dosage: $\beta = 0.113, p = 0.005$). Historically excluded racial/ethnic groups showed higher perceived gains in the overall quality of mentoring than well-represented groups ($\beta = 0.234, p = 0.002$). Mentors without previous mentoring experience reported higher gains in overall quality than mentors with previous mentoring experience ($\beta = 0.487, p < 0.001$).

Perceived Gains in Ability to Meet Mentees’ Expectations

Columns 6 and 7 of Table 5 show the results of treatment effects related to mentors’ perceived gains in ability to meet mentees’ expectations with and without additional covariates. In brief, perceived gains in ability to meet mentees’ expectations did not significantly differ for mentors trained online or in-person ($\beta = 0.168, p = 0.149$). After factoring in additional covariates, we found the training platform (online vs. in-person) was not a significant factor contributing to the perceived gains in ability to meet mentee’s expectations ($\beta = 0.098, p = 0.443$). Instead, facilitator effectiveness, race/ethnicity, and previous mentoring experience showed significant predictive power in estimating perceived gains in ability to meet mentees’ expectations. Specifically, greater facilitator effectiveness was associated with higher reported gains in ability to meet mentees’ expectations for mentors. For mentors’ race/ethnicity, historically excluded racial/ethnic groups demonstrated higher gains in ability to meet expectations than well-represented groups ($\beta = 0.275, p = 0.002$). Similarly, mentors without previous mentoring experience had higher reported gains than mentors with previous mentoring experience in meeting expectations ($\beta = 0.380, p < 0.001$).

Mentors’ Intended Changes to Mentoring Practices

In the last two columns of Table 5, the treatment effects in terms of the likelihood of making any changes in mentoring practices are shown. The training platform was found not to be a significant factor (without covariates: $\beta = 0.026, p = 0.305$; with covariates: $\beta = 0.031, p = 0.271$). However, facilitator effectiveness, race/ethnicity, and previous mentoring experience were significantly associated with intended behavioral changes. In particular, there was a greater likelihood of mentors intending to make changes in their mentoring with more effective facilitation. Regarding race/ethnicity, historically excluded racial/ethnic groups were more likely to intend to make changes

years of experience, NA (Yrs_ExperienceNA); mentor’s sex (gender_new); and mentor’s race (race). (B) The balance plot after PSM in observed covariates. The dashed lines represent Rubin’s benchmark: 1) standard differences in means less than 0.25; and 2) the variance ratios between 0.5 and 2 (Rubin, 2001; Stuart and Rubin, 2008). The covariates represented in the figure are mentees’ career stages: faculty (trainee_Faculty), graduate (trainee_Graduate), undergraduate (trainee_Undergrad), and none (trainee_none); mentor’s title (title_new); mentor’s previous mentoring experience (PreviousMentorExo_new); mentor’s years of experience (Yrs_Experience_new) and mentor’s years of experience, NA (Yrs_ExperienceNA); mentor’s sex (gender_new); and mentor’s race (race). (C) The density plot for the logit of PSM.

TABLE 5. Estimates of the ATT on mentor training outcomes using PSM^a

	Mentoring skill gains		Overall quality gains		Meeting expectation gains		Intended changes	
	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2	Model 1	Model 2
(Intercept)	0.921 *** (0.030)	0.344 (0.274)	1.227*** (0.041)	-0.087 (0.337)	1.029*** (0.046)	0.794** (0.391)	1.045*** (0.010)	1.040*** (0.086)
Platform: online	0.132 (0.087)	0.114 (0.090)	0.044 (0.106)	0.128 (0.110)	0.168 (0.116)	0.098 (0.128)	0.026 (0.026)	0.031 (0.028)
Dosage		0.064* (0.033)		0.113*** (0.040)		-0.008 (0.047)		-0.001 (0.010)
Facilitator effectiveness (compared with “very effective”)								
Effective		-0.134** (0.062)		-0.131 (0.081)		-0.152 (0.094)		0.085*** (0.021)
Neither		-0.630*** (0.175)		-1.073*** (0.281)		-0.811** (0.324)		0.292*** (0.066)
Ineffective		-0.398 (0.285)		-1.125** (0.469)		-1.007* (0.542)		0.426*** (0.121)
Very ineffective		-0.484 (0.414)		-0.265 (0.872)		-0.049 (1.007)		0.003 (0.225)
Missing (NA)		-0.540*** (0.113)		-0.816 (0.874)		-0.726 (1.008)		0.969*** (0.226)
Race/ethnicity (compared with “well represented”)								
Historically excluded		0.102* (0.055)		0.234*** (0.074)		0.275*** (0.089)		-0.014 (0.019)
Prefer not to answer		-0.012 (0.228)		-0.011 (0.254)		0.029 (0.321)		0.192*** (0.063)
Sex (compared with “female”)								
Male		-0.067 (0.054)		-0.038 (0.074)		-0.058 (0.089)		-0.013 (0.019)
Other		-0.382 (0.364)		-0.109 (0.443)		-0.075 (0.511)		-0.010 (0.114)
Prefer not to answer		-0.569** (0.257)		0.125 (0.293)		0.142 (0.362)		-0.124* (0.074)
Previous mentor experience (compared with “yes”)								
No		0.319*** (0.070)		0.487*** (0.087)		0.380*** (0.100)		-0.022 (0.022)
Missing (NA)		-0.097 (0.122)		1.296 (0.883)		-		-0.959*** (0.228)

^aRobust standard errors are in parentheses. Model 1 estimates the ATT on mentor training outcomes using PSM without additional covariates. Model 2 estimates the ATT on mentor training outcomes using PSM with additional covariates.

* $p < 0.1$.

** $p < 0.05$.

*** $p < 0.01$.

in their mentoring practices than well-represented groups ($\beta = -0.014$, $p = 0.002$). Compared with mentors with previous mentoring experience, mentors without any previous mentoring experience had a higher likelihood of intending to make changes in their mentoring ($\beta = 0.959$, $p < 0.001$).

DISCUSSION

In this study, differences in the effectiveness of RMT between in-person and synchronous online modalities were investigated using the PSM method. Consistent with previous reports (Handelsman *et al.*, 2005; Pfund *et al.*, 2014), mentors trained using *Entering Mentoring* reported perceived mentoring skill gains, further validating the efficacy of this training curriculum. Here, we report no differences in outcomes between training platforms for the first time, demonstrating that both in-person and synchronously online-trained mentors achieved similar perceived gains in mentorship skills, overall quality of mentoring, and the ability to meet mentees' expectations, as well as

intent to make changes in their mentoring practices. As such, online training may be a feasible alternative option or supplementation for in-person mentor training due to its time- and cost-efficiency (Bartley and Golek, 2004). This finding is especially important given the need to pivot to online mentorship education due to COVID-19.

This finding aligns with other studies that have suggested equivalent or greater learning gains in online learning environments compared with face-to-face classroom instruction (Nguyen, 2015), including public administration (Ni, 2013), evidenced-based psychotherapies (Mallonee *et al.*, 2017), and teacher education courses (Thirunarayanan and Perez-Prado, 2001; Hurlbut, 2018). Multiple pilot studies have recently reported significant mentoring skill improvements using hybrid in-person/online training interventions. For example, self-reported retrospective pre-post assessment of participants in the NIH Tobacco Regulatory Science (TRS) blended in-person and online training program indicated increased confidence in

assisting, guiding, and supporting their mentees' success in pursuing TRS careers (Di Frances *et al.*, 2020). Further, at the University of Minnesota, a pilot study evaluating the efficacy of a hybrid 90-minute self-paced, asynchronous online training and an approximately 4-hour in-person mentoring workshop modality was conducted (Weber-Main *et al.*, 2019). After completion of the asynchronous online module, researchers demonstrated an increase in knowledge gains and the likelihood of implementing behavioral changes in their mentoring practices. Yet these improvements and overall skill gains for trained mentors were significantly augmented following completion of both the online and in-person training to levels comparable to a full-day standard in-person workshop. Additional studies are currently underway to assess the efficacy, quality, and impact of asynchronous online training as a stand-alone RMT intervention.

The results also demonstrate that the number of training hours is positively associated with training outcomes, indicating higher dosage improves outcomes. These findings are consistent with a previous report showing that perceived increases in overall mentoring quality were significantly augmented with longer (4–10 hours) training, while gains were still evident in shorter training models (Rogers *et al.*, 2020). These results suggest that training duration should be optimized for online training modalities to maximize participant learning gains and avoid e-learning attrition due to cognitive overload (Tyler-Smith, 2006). While this study finds dosage to be significant in training outcomes, studies further examining how this relationship works are needed. Are the additional hours of training meaningful alone or are results dependent on the content of those hours of training? In addition, mentors without previous mentoring experience reported greater benefits from mentor training than mentors with some previous mentoring experience. Similarly, Rogers *et al.* (2020) reported that perceived overall mentoring quality and mentoring skill gains were significantly greater in participants without previous training. However, it is important to note that participants with prior mentoring experience reported higher baseline levels of perceived mentoring quality and skills on the pretraining evaluation, indicating that heightened perception of incoming skills may have impacted subsequent learning gains. This result suggests a need for tiered training based on experience (i.e., basic vs. advanced mentor training) to simulate a novice baseline for experienced mentors.

Facilitator effectiveness was found to be another significant contributing factor to mentors' training outcomes. Many studies have highlighted the importance of teacher effectiveness and facilitator style in both in-person and virtual environments to support student engagement, knowledge, and learning gains (Willis, 1994; Oncu and Cakir, 2011; Cacciamani *et al.*, 2012). For facilitator development and to meet growing training demands, a validated facilitator training infrastructure was developed and expanded for facilitators of the *Entering Mentoring* curriculum, training more than 200 facilitators (Pfund *et al.*, 2015b; Rogers *et al.*, 2018; Spencer *et al.*, 2018). Participants in the Train-the-Trainer model showed significant confidence gains in the ability to facilitate RMT and to successfully implement training in their individual contexts. However, the perceived effectiveness of trained facilitators has not been assessed in all contexts; thus, its impact on participant experience and learning gains must be further explored. Additionally, to ensure consistency in facilitation across mentor training

experiences, the quality of facilitator training for virtual educators will be assessed in the future.

Finally, this study demonstrates greater skill gains across all measures for mentors from historically excluded racial/ethnic groups compared with mentors from well-represented backgrounds, suggesting a difference in the perceived benefits of mentor training, which is of great interest. Other studies also have shown underrepresented learners having more favorable perceptions of instructor support compared with well-represented students (Ke and Kwak, 2013); however, the satisfaction rating is also lower in these studies. The difference in the perceived benefits of mentor training found in this study could also be an indicator of the need for changes in the training environment and advancement in culturally aware mentorship education. Perhaps, for mentors of historically excluded racial/ethnic groups, their own experiences in the training environment and the previous lack of adequate mentoring have made them more aware of the need for skilled mentoring and thus more open to the training. For example, Ginther *et al.* (2011) showed that White investigators are significantly more likely than Black and Hispanic investigators to win R01 awards. One hypothesis posed to explain this difference is that minority investigators are less likely to receive adequate mentoring (Ginther *et al.*, 2011). This finding has contributed to a national focus on advancing the science of mentorship. *The Science of Effective Mentorship in STEMM* report from the National Academies highlights the critical role mentorship plays in developing a science identity and calls for culturally aware mentorship education. Thus, the finding that RMT participants from historically excluded racial/ethnic groups gain more from the training is especially important, given the heightened focus on improving the training and training environment for scholars from historically excluded racial/ethnic groups. Future studies are needed to fully explore this finding. In particular, an examination of differences in pretraining levels for historically excluded racial/ethnic groups compared with well-represented groups is needed, as well as qualitative studies to fully understand the differences.

Limitations and Strengths of the PSM Approach

Quasi-experimental studies can be prone to misinterpretation of treatment effects due to pre group differences in the absence of randomized experiments. PSM allows researchers to balance nonequivalent groups with predetermined and theoretically relevant covariates. This methodology has been proved to reduce selection bias and provides a more precise estimation of treatment response. Additionally, PSM gives nonrandomized studies experimental design characteristics (Dehejia and Wahba, 2002; Schafer and Kang, 2008). This study revealed that mentors who participated in online training might significantly differ from those in face-to-face training in terms of their characteristics. However, the use of PSM revealed that the training mode did not significantly impact the measured training outcomes. This is due to the fact that the propensity score matched data set was balanced with respect to selected covariates; in other words, most of the selection bias influencing the measured training outcomes was removed using the propensity score methodology. Thus, we obtained more correct estimates of the measured training outcomes for mentors in both training modes.

As an approach to check the sensitivity of the results using the optimal full matching, the research team also considered

other possible propensity score techniques, including inverse propensity weighting and regression estimation using the propensity score. The idea of inverse propensity weighting is up-weighting underrepresented units and down-weighting overrepresented units in the treatment and control groups (Steiner and Cook, 2013). Regression estimation using propensity score is based on the idea of using the propensity score as an additional covariate in standard regression approaches (Steiner and Cook, 2013). The authors chose optimal full matching as the primary analytical strategy due to the following reasons. Compared with inverse propensity weighting, optimal full matching yielded a more balanced set of the selected covariates used for matching, meaning that the treatment and control groups were more comparable using optimal full matching. Also, the inverse propensity weighting is sensitive to large weights and thus results in larger standard errors than the optimal full matching (Steiner and Cook, 2013). On the other hand, regression estimation relies on functional form assumptions, whereas optimal full matching relaxes them. Admittedly, there are certain limitations to the current study. For example, the baseline covariates are not extensive enough to measure the pretreatment group differences, limiting the observed covariates' potential to balance the treatment and control groups and remove all the overt bias. One way to address this issue is to collect a large number of covariates from heterogeneous domains, for example, years of mentor's previous mentoring experience, field of expertise, age, and so on. The propensity score methods also cannot control for unmeasured confounding factors, such as the mentor's ability or other concurrent events that may contaminate the impacts of the training on the outcomes.

Summary

This study suggests that RMT participants perceived mentoring skill gains and overall quality of mentoring to be comparable, regardless of the training modality used—online versus in-person. As such, online mentoring training should be considered a viable option. Other variables such as dosage, facilitator effectiveness, race/ethnicity, sex, and previous mentoring experience were associated with perceived gains in mentoring skills and overall quality of mentoring. Similarly, both perceived gains in the ability to meet mentees' expectations and mentors' intention to make changes to their mentoring practices were comparable regardless of the training modality. Interestingly, a practical implication of the study is the effectiveness of the facilitator, suggesting that facilitator training is critical and thus should be standardized to ensure that facilitators of the *Entering Mentoring* curriculum have the requisite skills.

A limitation of this study is that the outcome measures are self-reported perceived gains and intended behavioral changes, which do not have specific evidence of validity and reliability in this study. While these measures have been proven to be strong predictors of actual skill gains and behaviors in both the short and long term, in other studies they still remain short-term, self-reported measures (Pfund *et al.*, 2014; House *et al.*, 2018; Trejo *et al.*, 2022). Additionally, there is some heterogeneity in survey data collection between the two modes of training. Future studies should look at long-term RMT outcomes, including actual behavioral changes and reported skill gains, and explore how these vary by in-person versus online training modes.

Additionally, future research should identify and assess specific domains of facilitators' skills. Concerning race/ethnicity, mentors from historically excluded racial/ethnic groups benefited more from the training, reported higher perceived gains in mentoring skills, overall quality of mentoring, and their ability to meet mentees' expectations. Mentors without previous mentoring experience benefited more from the training than those with some experience. This study has also highlighted the need for future studies to untangle the effects of dosage, type of participant, and mode of training. To optimize training outcomes for all groups, we need more in-depth analyses studies examining how different levels of dosage impact training outcomes, how this varies by types of participants, and also how this differs for online trainings. Finally, this study demonstrates that the PSM method is able to yield accurate estimates of the treatment effect in nonexperimental settings, indicating that PSM can reasonably replicate experimental impact estimates.

ACKNOWLEDGMENTS

This project was sponsored by the NIH under grants R01 GM094573 through the National Institute of General Medical Sciences and U54GM119023 (NRMN) through the National Institute of General Medical Sciences from the National Institutes of Health Common Fund and Office of Scientific Workforce Diversity. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Additional support was provided by the Institute for Clinical and Translational Research supported by the UW-Madison Clinical and Translational Science Award (CTSA) program, the National Center for Advancing Translational Sciences (NCATS), grant 1UL1TR002373. We would also like to thank Peter Steiner, Rachel Wolfson, and Aviva Klein for their insights and review of the article. We again thank Amanda Butz and Kim Spencer, who helped us implement the training assessed as part of this study and assisted with data collection.

REFERENCES

- Amoah, J., Stuart, E. A., Cosgrove, S. E., Harris, A. D., Han, J. H., Lautenbach, E., & Tamma, P. D. (2020). Comparing propensity score methods versus traditional regression analysis for the evaluation of observational data: A case study evaluating the treatment of gram-negative bloodstream infections. *Clinical Infectious Diseases*, 71(9), e497–e505. <https://doi.org/10.1093/cid/ciaa169>
- Asquith, P., Shapiro, E., Weber-Main, A. M., Jacobs, E., & Sorkness, C. A. (2014). Mentor training for clinical and behavioral researchers. Madison, WI: University of Wisconsin Institute for Clinical and Translational Research.
- Austin, P. C., & Stuart, E. A. (2017). The performance of inverse probability of treatment weighting and full matching on the propensity score in the presence of model misspecification when estimating the effect of treatment on survival outcomes. *Statistical Methods in Medical Research*, 26(4), 1654–1670. <https://doi.org/10.1177/0962280215584401>
- Bartley, S. J., & Golek, J. H. (2004). Evaluating the cost effectiveness of online and face-to-face instruction. *Journal of Educational Technology & Society*, 7(4), 167–175.
- Branchaw, J. L., Pfund, C., & Rediske, R. (2011). *Entering research: Workshops for students beginning research in science. Facilitator's manual*. New York, NY: Freeman.
- Brown, J. S., Collins, A., & Duguid, P. (1989). Situated cognition and the culture of learning. *Educational Researcher*, 18(1), 32–42. <https://doi.org/10.2307/1176008>
- Byars-Winston, A. M., Branchaw, J., Pfund, C., Leverett, P., & Newton, J. (2015). Culturally diverse undergraduate researchers' academic

- outcomes and perceptions of their research mentoring relationships. *International Journal of Science Education*, 37(15), 2533–2554. <https://doi.org/10.1080/09500693.2015.1085133>
- Cacciamani, S., Cesareni, D., Martini, F., Ferrini, T., & Fujita, N. (2012). Influence of participation, facilitator styles, and metacognitive reflection on knowledge building in online university courses. *Computers & Education*, 58(3), 874–884. <https://doi.org/10.1016/j.compedu.2011.10.019>
- Caliendo, M., & Kopeinig, S. (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys*, 22(1), 31–72. <https://doi.org/10.1111/j.1467-6419.2007.00527.x>
- Chemers, M. M., Zurbriggen, E. L., Syed, M., Goza, B. K., & Bearman, S. (2011). The role of efficacy and identity in science career commitment among underrepresented minority students. *Journal of Social Issues*, 67(3), 469–491. <https://doi.org/10.1111/j.1540-4560.2011.01710.x>
- Dehejia, R. H., & Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of the American Statistical Association*, 94(448), 1053–1062. <https://doi.org/10.1080/01621459.1999.10473858>
- Dehejia, R. H., & Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics*, 84(1), 151–161. <https://doi.org/10.1162/003465302317331982>
- Di Frances, C. D., Childs, E., Fetterman, J. L., Villanti, A. C., Stanton, C. A., Russo, A. R., ... & Benjamin, E. J. (2020). Implementing and evaluating a mentor training to improve support for early-career scholars in tobacco regulatory science. *Nicotine & Tobacco Research*, 22(6), 1041–1045. <https://doi.org/10.1093/ntr/ntz083>
- Dolan, E., & Johnson, D. (2009). Toward a holistic view of undergraduate research experiences: An exploratory study of impact on graduate/postdoctoral mentors. *Journal of Science Education and Technology*, 18(6), 487–500. <https://doi.org/10.1007/s10956-009-9165-3>
- Dolan, E., & Johnson, D. (2010). The undergraduate–postgraduate–faculty triad: Unique functions and tensions associated with undergraduate research experiences at research universities. *CBE—Life Sciences Education*, 9(2), 543–553. <https://doi.org/10.1187/cbe.10>
- Estrada, M., Hernandez, P. R., & Schultz, P. W. (2018). A longitudinal study of how quality mentorship and research experience integrate underrepresented minorities into STEM careers. *CBE—Life Sciences Education*, 17(1), ar9. <https://doi.org/10.1187/cbe.17-04-0066>
- Faulconer, E. K., Griffith, J., Wood, B., Acharyya, S., & Roberts, D. (2018). A comparison of online, video synchronous, and traditional learning modes for an introductory undergraduate physics course. *Journal of Science Education and Technology*, 27(5), 404–411. <https://doi.org/10.1007/s10956-018-9732-6>
- Fleming, M., House, S., Hanson, V. S., Yu, L., Garbutt, J., McGee, R., ... & Rubio, D. M. (2013). The Mentoring Competency Assessment: Validation of a new instrument to evaluate skills of research mentors. *Academic Medicine*, 88(7), 1002–1008. <https://doi.org/10.1097/ACM.0b013e318295e298>
- Ginther, D. K., Schaffer, W. T., Schnell, J., Masimore, B., Liu, F., Haak, L. L., & Kington, R. (2011). Race, ethnicity, and NIH research awards. *Science*, 333(6045), 1015–1019. <https://doi.org/10.1126/science.1196783>
- Haeger, H., & Fresquez, C. (2016). Mentoring for inclusion: The impact of mentoring on undergraduate researchers in the sciences. *CBE—Life Sciences Education*, 15(3), ar36. <https://doi.org/10.1187/cbe.16-01-0016>
- Handelsman, J., Pfund, C., Miller Lauffer, S., & Pribbenow, C. M. (2005). *Entering mentoring: A seminar to train a new generation of scientists*. Madison, WI: University of Wisconsin Press.
- Hansen, B. B. (2004). Full matching in an observational study of coaching for the SAT. *Journal of the American Statistical Association*, 99(467), 609–618. <https://doi.org/10.1198/016214504000000647>
- Hansen, B. B. (2007). Optmatch: Flexible, optimal matching for observational studies. *New Functions for Multivariate Analysis*, 7(2), 18–24.
- Hansen, B. B., & Klopfer, S. O. (2006). Optimal full matching and related designs via network flows. *Journal of Computational and Graphical Statistics*, 15(3), 609–627. <https://doi.org/10.1198/106186006X137047>
- House, S., Dearlove, A., Spencer, K., & Zeigahn, L. (2014). Mentor training for community engaged researchers (W.H. Freeman Entering Mentoring Series. Madison, WI: University of Wisconsin–Madison.
- House, S. C., Spencer, K. C., & Pfund, C. (2018). Understanding how diversity training impacts faculty mentors' awareness and behavior. *International Journal of Mentoring and Coaching in Education*, 7(1), 72–86. <https://doi.org/10.1108/IJMCE-03-2017-0020>
- Hurlbut, A. R. (2018). Online vs. traditional learning in teacher education: A comparison of student progress. *American Journal of Distance Education*, 32(4), 248–266. <https://doi.org/10.1080/08923647.2018.1509265>
- Hyun, S. H., Rogers, J. G., House, S. C., Sorkness, C. A., & Pfund, C. (2022). Re-validation of the mentoring competency assessment to evaluate skills of research mentors: The MCA-21. *Journal of Clinical and Translational Science*, 6(1), 1–23. <https://doi.org/10.1017/cts.2022.381>
- Junge, B., Quiñones, C., Kakietek, J., Teodorescu, D., & Marsteller, P. (2010). Promoting undergraduate interest, preparedness, and professional pursuit in the sciences: An outcomes evaluation of the SURE program at Emory University. *CBE—Life Sciences Education*, 9(2), 119–132. <https://doi.org/10.1187/cbe.09-08-0057>
- Ke, F., & Kwak, D. (2013). Online learning across ethnicity and age: A study on learning interaction participation, perception, and learning satisfaction. *Computers & Education*, 61, 43–51. <https://doi.org/10.1016/j.compedu.2012.09.003>
- Lane, F., To, Y., Shelley, K., & Henson, R. (2012). An illustrative example of propensity score matching with education research. *Career and Technical Education Research*, 37(3), 187–212. <https://doi.org/10.5328/cter37.3.187>
- Limeri, L. B., Asif, M. Z., Bridges, B. H. T., Esparza, D., Tuma, T. T., Sanders, D., ... & Dolan, E. L. (2019). "Where's my mentor?!" Characterizing negative mentoring experiences in undergraduate life science research. *CBE—Life Sciences Education*, 18(4), ar61. <https://doi.org/10.1187/cbe.19-02-0036>
- Mallonee, S., Phillips, J., Holloway, K., & Riggs, D. (2017). Training providers in the use of evidence-based treatments: A comparison of in-person and online delivery modes. *Psychology Learning & Teaching*, 17(1), 61–72. <https://doi.org/10.1177/1475725717744678>
- Martin, F., Ahlgrim-Delzell, L., & Budhrani, K. (2017). Systematic review of two decades (1995 to 2014) of research on synchronous online learning. *American Journal of Distance Education*, 31(1), 3–19. <https://doi.org/10.1080/08923647.2017.1264807>
- McDaniels, M., Pfund, C., & Barnicle, K. (2016). Creating dynamic learning communities in synchronous online courses: One approach from the Center for the Integration of Research, Teaching and Learning (CIRTL). *Online Learning*, 20(1), 110–129. <https://doi.org/10.24059/olj.v20i1.518>
- McGee, R. (2016). Biomedical workforce diversity: The Context for mentoring to develop talents and foster success within the "pipeline." *AIDS and Behavior*, 20(2), 231–237. <https://doi.org/10.1007/s10461-016-1486-7>
- Melguizo, T., Kienzl, G. S., & Alfonso, M. (2011). Comparing the educational attainment of community college transfer students and four-year college rising juniors using propensity score matching methods. *Journal of Higher Education*, 82(3), 265–291. <https://doi.org/10.1080/00221546.2011.11777202>
- Mullin, D. J., Saver, B., Savageau, J. A., Forsberg, L., & Forsberg, L. (2016). Evaluation of online and in-person motivational interviewing training for healthcare providers. *Families, Systems, and Health*, 34(4), 357–366. <https://doi.org/10.1037/fsh0000214>
- Nagda, B. A., Gregerman, S. R., Jonides, J., von Hippel, W., & Lerner, J. S. (1998). Undergraduate student-faculty research partnerships affect student retention. *Review of Higher Education*, 22(1), 55–72. <https://doi.org/10.1353/rhe.1998.0016>
- National Academies of Sciences, Engineering, and Medicine (NASEM). (2019). *The science of effective mentorship in STEM*. Washington, DC: National Academies Press. <https://doi.org/10.17226/25568>
- Nguyen, T. (2015). The effectiveness of online learning: Beyond no significant difference and future horizons. *Journal of Online Learning and Teaching*, 11(2), 309–319.
- Ni, A. Y. (2013). Comparing the effectiveness of classroom and online learning: Teaching research methods. *Journal of Public Affairs Education*, 19(2), 199–215. <https://doi.org/10.1080/15236803.2013.12001730>
- Oncu, S., & Cakir, H. (2011). Research in online learning environments: Priorities and methodologies. *Computers & Education*, 57(1), 1098–1108. <https://doi.org/10.1016/j.compedu.2010.12.009>
- Pfund, C., Brace, C., Branchaw, J., Handelsman, J., Masters, K. S., & Nanney, L. (2013a). Mentor training for biomedical researchers. New York, NY: Freeman.

- Pfund, C., Branchaw, J. L., & Handelsman, J. (2015a). *Entering mentoring* (2nd ed.). New York, NY: Freeman.
- Pfund, C., House, S. C., Asquith, P., Fleming, M. F., Buhr, K. A., Burnham, E. L., ... & Sorkness, C. A. (2014). Training mentors of clinical and translational research scholars: A randomized controlled trial. *Academic Medicine*, *89*(5), 774–782. <https://doi.org/10.1097/ACM.0000000000000218>
- Pfund, C., House, S., Spencer, K., Asquith, P., Carney, P., Masters, K. S., ... & Fleming, M. (2013b). A research mentor training curriculum for clinical and translational researchers. *Clinical and Translational Science*, *6*(1), 26–33.
- Pfund, C., Pribbenow, C. M., Branchaw, J., Lauffer, S. M., & Handelsman, J. (2006). The merits of training mentors. *Science*, *311*(5760), 473–474. <https://doi.org/10.1126/science.1123806>
- Pfund, C., Spencer, K. C., Asquith, P., House, S. C., Miller, S., & Sorkness, C. A. (2015b). Building national capacity for research mentor training: An evidence-based approach to training the trainers. *CBE—Life Sciences Education*, *14*(2), ar24. <https://doi.org/10.1187/cbe.14-10-0184>
- Poodry, C. A., & Asai, D. J. (2018). Questioning assumptions. *CBE—Life Sciences Education*, *17*(3), es7. <https://doi.org/10.1187/cbe.18-02-0024>
- R Core Team. (2019). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved March 15, 2022, from www.R-project.org
- Reeve, B. B., Smith, A. W., Arora, N. K., & Hays, R. D. (2008). Reducing bias in cancer research: Application of propensity score matching. *Health Care Financing Review*, *29*(4), 69–80.
- Rogers, J., Branchaw, J., Weber-Main, A. M., Spencer, K., & Pfund, C. (2020). How much is enough? The impact of training dosage and previous mentoring experience on the effectiveness of a research mentor training intervention. *Understanding Interventions*, *11*(1), 1–17.
- Rogers, J., Sorkness, C. A., Spencer, K., & Pfund, C. (2018). Increasing research mentor training among biomedical researchers at Clinical and Translational Science Award hubs: The impact of the facilitator training initiative. *Journal of Clinical and Translational Science*, *2*(3), 118–123. <https://doi.org/10.1017/cts.2018.33>
- Rojewski, J., Lee, I. H., & Gemici, S. (2010). Using propensity score matching to determine the efficacy of secondary career academies in raising educational aspirations. *Career and Technical Education Research*, *35*(1), 3–27. <https://doi.org/10.5328/cter35.102>
- Rosenbaum, P. R., & Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, *70*(1), 41–55. <https://doi.org/10.1093/biomet/70.1.41>
- Rubin, D. B. (2001). Using propensity scores to help design observational studies: Application to the tobacco litigation. *Health Services and Outcomes Research Methodology*, *2*(3), 169–188. <https://doi.org/10.1023/A:1020363010465>
- Schafer, J. L., & Kang, J. (2008). Average causal effects from nonrandomized studies: A practical guide and simulated example. *Psychological Methods*, *13*(4), 279–313. <https://doi.org/10.1037/a0014268>
- Shadish, W. R., & Steiner, P. M. (2010). A primer on propensity score analysis. *Newborn and Infant Nursing Reviews*, *10*(1), 19–26. <https://doi.org/10.1053/j.nainr.2009.12.010>
- Spencer, K. C., McDaniels, M., Utzerath, E., Rogers, J. G., Sorkness, C. A., Asquith, P., & Pfund, C. (2018). Building a sustainable national infrastructure to expand research mentor training. *CBE—Life Sciences Education*, *17*(3), ar48. <https://doi.org/10.1187/cbe.18-03-0034>
- Steiner, P. M., & Cook, D. (2013). Matching and propensity scores. In *The Oxford handbook of quantitative methods* (Vol. 1, pp. 237–259). New York, NY: Oxford University Press.
- Stolzenberg, E. B., Eagan, M. K., Zimmerman, H. B., Berdan Lozano, J., Cesar-Davis, N. M., Aragon, M. C., & Rios-Aguilar, C. (2019). *Undergraduate teaching faculty: The HERI Faculty Survey 2016–2017*. Los Angeles: Higher Education Research Institute, UCLA. Retrieved March 15, 2021, from <https://heri.ucla.edu/monographs/HERI-FAC2017-monograph.pdf>
- Straus, S. E., Johnson, M. O., Marquez, C., & Feldman, M. D. (2013). Characteristics of successful and failed mentoring relationships: A qualitative study across two academic health centers. *Academic Medicine*, *88*(1), 82–89. <https://doi.org/10.1097/ACM.0b013e31827647a0>
- Stuart, E. A. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, *25*(1), 1–21. <https://doi.org/10.1214/09-STS313>
- Stuart, E. A., & Rubin, D. B. (2008). Best practices in quasi-experimental designs: Matching methods for causal inference. In Osborne, J. (Ed.), *Best practices in quantitative methods* (pp. 155–176). Los Angeles, CA: Sage. <https://doi.org/10.4135/9781412995627.d14>
- Thirunaranayan, M. O., & Perez-Prado, A. (2001). Comparing Web-based and classroom-based learning: A quantitative study. *Journal of Research on Technology in Education*, *34*(2), 131–137. <https://doi.org/10.1080/15391523.2001.10782340>
- Thiry, H., & Laursen, S. L. (2011). The role of student-advisor interactions in apprenticing undergraduate researchers into a scientific community of practice. *Journal of Science Education and Technology*, *20*(6), 771–784.
- Trejo, J., Wingard, D., Hazen, V., Bortnick, A., Hoesen, K. V., Byars-Winston, A., ... & Reznik, V. (2022). A system-wide health sciences faculty mentor training program is associated with improved effective mentoring and institutional climate. *Journal of Clinical and Translational Science*, *6*(1), E18. <https://doi.org/10.1017/cts.2021.883>
- Tuma, T. T., Adams, J. D., Hultquist, B. C., & Dolan, E. L. (2021). The dark side of development: A systems characterization of the negative mentoring experiences of doctoral students. *CBE—Life Sciences Education*, *20*(2), ar16. <https://doi.org/10.1187/cbe.20-10-0231>
- Tyler-Smith, K. (2006). Early attrition among first time learners: A review of factors that contribute to drop-out, withdrawal and non-completion rates of adult learners undertaking learning programmes. *Journal of Online Learning and Teaching*, *2*(2), 73–85.
- Ülkü-Steiner, B., Kurtz-Costes, B., & Kinlaw, C. R. (2000). Doctoral student experiences in gender-balanced and male-dominated graduate programs. *Journal of Educational Psychology*, *92*(2), 296–307. <https://doi.org/10.1037/0022-0663.92.2.296>
- Weber-Main, A. M., Shanedling, J., Kaizer, A. M., Connett, J., Lamere, M., & El-Fakahany, E. E. (2019). A randomized controlled pilot study of the University of Minnesota mentoring excellence training academy: A hybrid learning approach to research mentor training. *Journal of Clinical and Translational Science*, *3*(4), 152–164. <https://doi.org/10.1017/cts.2019.368>
- Willis, B. (1994). Enhancing faculty effectiveness in distance education. In Willis, B. (Ed.), *Distance education: Strategies and tools* (pp. 277–290). New Jersey, NJ: Educational Technology Publications.