Received: 1 May 2020 | Revised: 12 January 2022 | Accepted: 6 February 2022

DOI: 10.1002/tea.21761

RESEARCH ARTICLE

JRST | WILEY

Student and advisor gender identity in STEM doctoral programs: Examining longitudinal and mediating effects with latent growth models

Jennifer M. Blaney[1] | **David F. Feldon**[2] | **Kaylee Litson**[2]

[1]Department of Educational Leadership, Northern Arizona University, Flagstaff, Arizona, USA

[2]Department of Instructional Technology and Learning Sciences, Utah State University, Logan, Utah, USA

Correspondence
Jennifer M. Blaney, Northern Arizona University, College of Education, Flagstaff, AZ 86011, USA.
Email: jennifer.blaney@nau.edu

Funding information
National Science Foundation, Grant/Award Numbers: 1431234, 1431290, 1760894

Abstract

While previous research documents that women in STEM doctoral programs tend to fare better when their advisor shares their gender identity, this study provides new insights into the role of student–advisor gender identity congruence, relying on a longitudinal sample of doctoral students in biology and using structural equation and latent growth curve modeling. Findings show that advisor gender played an inconsistent and typically indirect role in predicting student outcomes. Further, all students, regardless of gender, tended to report higher quality advising when their advisor was a woman, pointing to potential gender inequities in advising expectations of faculty. Implications for research, theory, and practice are discussed.

KEYWORDS

advisor gender, biological sciences, doctoral students

1 | INTRODUCTION

Despite the great strides women have made toward parity in higher education, gender inequity persists across science, technology, engineering, mathematics (STEM) fields and manifests in a myriad of ways. While women's representation remains markedly low in certain STEM disciplines like computer science and engineering, in the biological sciences, women earning graduate degrees now outnumber men, making up 58% and 53% of master's and PhD degree recipients, respectively (National Science Foundation, 2018). Yet, women are significantly less

likely to be hired as postdoctoral researchers in the top laboratories that account for nearly all new faculty hired at elite research institutions (Sheltzer & Smith, 2014). Accordingly—and despite being the majority of PhD recipients—women make up 42% of new assistant professors and 30% of tenured faculty in life sciences (National Science Foundation, 2018). Other evidence suggests that women are less frequently promoted to full professor, relative to men in the biological sciences (Gumpertz et al., 2017; Nelson et al., 2007), collectively revealing ways in which gender inequities persist.

One perspective on ways to remedy such gender inequities emphasizes the importance of having women as faculty mentors and role models (Etzkowitz et al., 2000). More specifically, researchers have documented reduced gender gaps in STEM spaces when instructors are women (Solanki & Xi, 2018). Further, faculty gender may take on a unique importance among STEM doctoral students whose interactions with faculty are often characterized by the apprenticeship relationship between student and advisor (Austin & McDaniels, 2006; Sallee, 2011). In fact, student–advisor interactions have been identified as a key component of doctoral socialization and success (Nettles & Millett, 2006; Paglis et al., 2006; Weidman et al., 2001). Of particular importance to the present work, one recent study found that having a dissertation chair with the same gender identity positively predicted degree completion for women pursuing science and engineering doctorates (Main, 2018). Still, much remains unknown about how, why, and under what conditions an alignment between student and faculty gender identities impacts doctoral student outcomes.

The present study addresses this gap in the literature by examining the role of advisor gender identity among doctoral students in the biological sciences, taking other aspects of student–advisor interactions into account. The focus on the biological sciences in the present work is an important one, because it permits the examination of gender dynamics where interactions are not necessarily skewed by the rarity of women as students and faculty (see Fox, 2006). Specifically, this study focuses on the role of advisor gender identity alignment between doctoral students and their advisors in predicting key graduate school outcomes, paying careful attention to how the quality of student–advisor interactions might mediate or moderate these relationships. The following questions frame this study:

1. To what extent do student–advisor interactions differ by advisor gender and student–advisor gender identity congruence?
2. To what extent do advisor gender and student–advisor gender congruence predict desirable outcomes for doctoral students (i.e., sense of belonging, self-efficacy, and research skill development)?
3. To what extent do student–advisor interactions mediate or moderate the relationship between advisor gender/student–advisor gender congruence and student outcomes?

2 | THEORETICAL FRAMEWORK

The present work is broadly guided by identity-based motivation theory (see Oyserman, 2009), merged with graduate socialization theory (Weidman et al., 2001; Weidman & Stein, 2003). While identity-based motivation theory guided our focus on student–advisor gender identity congruence, graduate socialization provided a framework for examining the role of gender in the larger context of how students are impacted by socializing interactions with advisors. More specifically, and as discussed below, graduate socialization theory guided our decision to focus

on advisor interaction quality as a potential mediating and moderating variable. Both theories guided our selection of dependent variables, our hypotheses, and the interpretation of results, which we discuss below. In our application of these theories, our goal was not to make causal inferences. Rather, we sought to deconstruct and document complexity in how gender identity congruence might operate within the larger context of graduate socialization experiences. Below, we describe how identity-based motivation theory and graduate socialization theory provided a framework for this inquiry.

2.1 | Identity-based motivation theory

Research on women's graduate training experiences document that women graduate students often fare better when they have an advisor who shares their gender identity (e.g., Main, 2018). Existing theories of identity can provide a framework for understanding *how* an advisor's gender might shape doctoral student outcomes, particularly for women in STEM contexts. Building on the broader literature on homophily in social networks—which documents ways in which individuals form relationships with others who share their salient identity characteristics (see McPherson et al., 2001)—identity-based motivation theory posits that "people use identities to prepare for action and to make sense of the world around them" (Oyserman, 2009, p. 252). Two key tenets underlie this theory: (1) individuals are motivated to act in ways that are congruent with their salient identities and (2) identity salience is highly situational and malleable (Oyserman, 2009, 2014). Further, the more an individual perceives an action as identity-congruent, the more they may be motivated to persist through difficulty, interpreting difficult actions as "important and meaningful;" on the other hand, an individual who perceives an action as identity-incongruent might be deterred by experienced difficulty, leading them to feel "that the behavior is pointless" (Oyserman, 2014, p. 216). As such, gender identity is generally expected to shape the way students experience environments dominated by men, like STEM doctoral training. In the case of our study, we focus on the biological sciences, where women are well-represented among doctoral students but remain underrepresented in the professoriate.

At the same time, identity-based motivation theory highlights the ways in which identities are socially constructed and dynamic, suggesting that changes to the environment (e.g., the presence of women role models) can change individuals' understanding of identity-congruence. Though some literature suggests that having a greater proportion of women faculty may benefit all students (e.g., Pezzoni et al., 2016; Sax, 2008), within the context of many STEM fields, faculty and advisor gender identity may take on unique importance for women students' motivation and outcomes. For example, Solanki and Xi (2018) applied identity-based motivation theory to study student-instructor gender matching in undergraduate STEM contexts to explain why women students seem to benefit from the presence of women instructors. Their findings support a "role model effect" such that women who had instructors of the same gender identity demonstrated greater self-efficacy, STEM interests, help-seeking behaviors, and classroom engagement, relative to those who had an instructor who was a man.

While identity-based motivation theory has primarily been applied to K-12 (e.g., Carlone et al., 2014) and undergraduate settings (e.g., Jackson & Seiler, 2013), Burt (2018) recently found that shared identity characteristics shape graduate student success and intention to pursue faculty careers, at least for students from minoritized groups. Burt posited that having an advisor with a shared underrepresented identity helps students to "navigate some challenges and pressures of being racially underrepresented in the academy" (Burt, 2018, p. 303). While Burt

focused on social identities broadly, and racial/ethnic identity more specifically, other research has considered gender homophily among women doctoral students, revealing that same-gender advising relationships positively predict degree completion for women, but are non-significant for men (Main, 2014). While the present study focuses specifically on the role of gender identity, doctoral student outcomes are shaped by much more than identity congruence with advisors; for this reason, we turned to graduate socialization theory to provide additional framing for our work.

2.2 | Graduate student socialization theory

Graduate socialization theory—used to explain the processes by which graduate students develop field-relevant skills and internalize the norms and behaviors of their discipline—has been widely applied to understand doctoral student development (Weidman et al., 2001; Weidman & Stein, 2003). Graduate socialization theory points to a range of socializing agents, which include interactions with other graduate students, experiences with disciplinary associations, and family support. Most relevant to the present work, Weidman and colleagues emphasized the importance of student–advisor relationships in predicting desirable outcomes (e.g., self-efficacy, research skills). In terms of doctoral training specifically, advisors have long been considered the primary socializing agent, conveying field norms and expectations, developing skills, and providing other forms of support to doctoral students (Baird, 1995; Barnes & Austin, 2009).

Applications of graduate socialization theory to doctoral training contexts show that advisor interaction quality can play a significant role in shaping the extent to which students feel that they belong within disciplinary and academic spaces (Curtin et al., 2013). Similarly, advisor interactions have been shown to predict research self-efficacy among doctoral students (Curtin et al., 2016), which is unsurprising when assuming that advisors are responsible for developing students' research and other field-relevant skills. However, there is little empirical evidence demonstrating that advisors directly influence doctoral student skill development (Feldon et al., 2019). Despite this limitation, graduate socialization theory continues to be the dominant framework in the United States for understanding graduate student experiences and outcomes (see Weidman & DeAngelo, 2020).

Guided by Weidman et al.'s (2001) model of graduate socialization, we posit that student–advisor interaction quality may predict key doctoral student outcomes and may function as a mediating factor between advisor gender and student outcomes. Graduate socialization theory further guided our consideration of both affective and skill-based outcomes. That is, while much of the literature on doctoral education typically focuses on self-ratings and affective outcomes, we chose to also incorporate a measure of scholarly performance (i.e., demonstrated research skills), in keeping with recent recommendations that researchers utilize tangible measures of skill development in studies of graduate student socialization (Feldon et al., 2010).

3 | LITERATURE REVIEW

In addition to the guiding theories above, our work is informed by broader studies of graduate training. Below we review this relevant literature, focusing first on key student outcomes within

extant literature on STEM doctoral training. Next, we discuss how faculty advisors may shape doctoral student outcomes, before presenting literature on the role of student–faculty gender identity congruence. We conclude with a brief discussion of the ways in which gender is socially constructed within STEM training environments.

3.1 | Key doctoral student outcomes

To understand the impact of gender and advisor interactions on doctoral student success, we must first identify and define key outcomes within the literature on doctoral training. Across studies, researchers have focused on a range of outcomes, both affective and skill-based in nature. Recognizing this, the present study examined sense of belonging and self-efficacy outcomes, as well as research skill development. Below, we define and synthesize key literature on these three outcomes.

3.1.1 | Sense of belonging

Sense of belonging—defined as the extent to which one feels that they "fit in" and are valued within a given domain (Goodenow, 1993)—is an important outcome for doctoral students, as it is a documented predictor of retention, satisfaction with graduate school, and other measures of success (Freeman et al., 2007; Ostrove et al., 2011; Pittman & Richmond, 2008; Strayhorn, 2012). As such, sense of belonging has been a focus within the literature on doctoral socialization (see Curtin et al., 2013; Gardner & Holley, 2011; Ostrove et al., 2011; White & Nonnamaker, 2008). Some of this work reveals that sense of belonging is associated with advisor support (Curtin et al., 2013), and that advisors may play a particularly important role in fostering sense of belonging among women graduate students (Le et al., 2016). Studies of sense of belonging in graduate school tend to rely on students' self-reported feelings that they belong within their graduate program or the larger field (see Bollen & Hoyle, 1990; Curtin et al., 2013; Ostrove et al., 2011). Because we are particularly interested in the role of advisors within the biological sciences, we decided to focus specifically on doctoral students' sense of belonging within their research group or lab where the advisor holds a unique supervisory role (see Maher et al., 2020).

3.1.2 | Research self-efficacy

Self-efficacy, or one's belief about their ability to be successful in a given domain (Bandura, 1997), is associated with future academic achievement (Lent et al., 1986; Multon et al., 1991; Zimmerman et al., 1992) and identity as a scientist (Robnett et al., 2015). Self-efficacy has been explored in studies of gender and student–faculty interactions in STEM contexts. For example, Solanki and Xi (2018) found that undergraduate women in STEM reported greater self-efficacy when they were exposed to women instructors. Notably—and in keeping with the present work—studies of doctoral training tend to focus more specifically on *research* self-efficacy (e.g., Litson et al., 2021), documenting the importance of faculty mentorship in supporting more positive research self-efficacy outcomes among students (Overall et al., 2011).

3.1.3 | Research skills

While much of the literature on doctoral student training has focused primarily on affective measures (e.g., sense of belonging and self-efficacy), developing research skills has been identified as a primary outcome of science education, both generally and in graduate education (Weidman et al., 2001). Within a STEM graduate training context, "research skills" encompass students' abilities across several key domains, including experimental design, data analysis, and presentation and communication of results. While several measures have been developed to assess specific subsets of skills related to research abilities (e.g., scientific reasoning; Lawson, 2008), other studies have used rubric-based assessment of research products to measure specific and collective research skills (e.g., Hackett & Rhoten, 2009; Timmerman et al., 2013). Although prior research using performance-based measures has not documented a direct connection between faculty advisors and research skill development among STEM doctoral students (Feldon et al., 2019), recent applications of graduate socialization theory stress the importance of considering tangible measures of skill development in addition to exploring affective student outcomes (e.g., Feldon, 2020; Gopaul, 2016).

3.2 | Advisor effects on students and the significance of faculty gender

While the importance of faculty has been well established within the broader literature on STEM education and career pathways (e.g., Pugh et al., 2021), faculty—particularly the faculty advisor—play a unique role within STEM doctoral training. Doctoral education has long been considered an apprenticeship model, in which doctoral students acquire disciplinary knowledge and an understanding of the norms of academia through one-on-one interactions with their faculty advisor (Flores-Scott & Nerad, 2012; Gardner, 2010a; Littlefield et al., 2015; Weidman et al., 2001). Researchers have identified the various roles advisors play in the doctoral training process, which include (1) providing intellectual and instrumental support, (2) helping students build their network in their chosen field, and (3) supporting students personally and psychosocially (Tenenbaum et al., 2001; Zhao et al., 2007). By serving in these different roles, advisors may impact doctoral student development and career outcomes. For example, Overall et al. (2011) surveyed doctoral students across disciplines and found that students tended to have the highest levels of research self-efficacy when they had faculty advisors who provided skill-development support coupled with opportunities for research autonomy. More recent research, also using quantitative data from doctoral students across fields, documents the importance of networking support, which predicts doctoral students' interest in an academic career (Curtin et al., 2016). While Curtin and colleagues also found psychosocial support to predict doctoral student academic career interests and self-efficacy, others found that students tended to have the lowest levels of research self-efficacy when their advisors focused *exclusively* on providing personal support (Overall et al., 2011). Taken together, existing literature suggests that students benefit from robust advising practices that include instrumental, networking, and psychosocial support.

In addition to the nature and quality of advising, other studies highlight the significance of faculty and advisor gender in shaping student outcomes. Within this literature, the importance of having a diverse faculty is broadly recognized and the negative impacts of inequitable gender representation are well-documented. Studies have found that students, regardless of gender,

generally benefit from being in programs with a greater proportion of women faculty (Sax, 2008). Other research focuses specifically on advisor gender identity congruence. For example, one case study of students at the California Institute of Technology revealed that women students saw the lowest publication rates when their advisor was a man, but that all students—regardless of their gender—generally published more frequently when their advisor was a woman (Pezzoni et al., 2016). In the field of chemistry, Gaule and Piacentini (2018) found that doctoral students tended to be more productive when they had an advisor who shared their gender identity, though this study was somewhat limited by their reliance on student and advisor names to determine gender. Another study—this one focusing on humanities and social science disciplines and relying on multi-institution survey and administrative data—revealed that having a woman as a dissertation chair predicted degree completion for women students (Main, 2014).

Recently, researchers have begun asking *how*, *why*, and *under what conditions* doctoral students might fare better with advisors who share their gender identity. Some literature has suggested that the answer may be related to faculty bias; that is, some men faculty may provide students with fewer opportunities if they are women (Rosser, 2018; Sheltzer & Smith, 2014) and may perceive women as less competent, even when they perform better than or equal to men (Moss-Racusin et al., 2012). Similarly, De Welde and Laursen (2011) analyzed interview data from doctoral students, revealing ways in which men faculty more frequently formed informal relationships with men students, which led to increased networking opportunities for men but not women. These findings are particularly disturbing, given that men faculty continue to serve as the primary gatekeepers to academia and make up the majority of senior STEM faculty (Rosser, 2018; Sheltzer & Smith, 2014).

3.3 | Contextualizing gender as a social construction

We must also contextualize our study and the studies cited above within broader understandings of gender and gender identity. Gender inequities and sexism have been well-documented in STEM doctoral training (e.g., Wofford & Blaney, 2021), among STEM faculty (e.g., Rosser, 2018), and in academia beyond STEM contexts (e.g., Croom, 2017). Though long understood to be a social construction (West & Zimmerman, 1987), sexism and gender inequities persist and often shape women's lived experiences in science and other academic contexts (see Kachchaf et al., 2015). More recent research has provided insights into the ways in which gender is often essentially constructed within STEM academic contexts, like the biological sciences, contributing to especially hostile environments for women (Lockhart, 2021). Other research shows that specific STEM disciplines vary in how gendered stereotypes manifest in the field, which, in turn, relates to the representation of women (see Leslie et al., 2015).

Recognizing the socially constructed nature of gender, it is important to note that prior research documenting that women doctoral students may benefit from having a woman advisor (e.g., Main, 2014) does not indicate that gender identity is the determining factor that shapes student outcomes. Rather, these prior findings represent how gender might operate only within specific gendered contexts. The same logic holds for our study; that is, while we examine the effect of gender identity congruence on doctoral student outcomes, we recognize that gender identity is a social construction and seek to provide new insight into the complex role that constructions of gender might play in the specific context of mentorship in doctoral programs in the biological sciences.

3.4 | Gaps in the literature

As reviewed above, a large and growing body of literature has examined the role of student–advisor interactions and the role of gender in predicting key doctoral student outcomes. Collectively, these studies highlight the importance of high-quality advising interactions that include student support for skill development, networking, and psychosocial needs (e.g., Ostrove et al., 2011), as well as the ways in which women STEM students tend to benefit from having an advisor who shares their gender identity (e.g., Main, 2014, 2018). However, little is known about how student–advisor gender identity congruence might shape doctoral student success within the larger context of student–advisor interactions. Further, much of the existing literature on student–advisor relationships and gender has focused exclusively on affective outcomes. The present study addresses these gaps in the literature by exploring the direct and indirect relationships between student–advisor gender identity congruence, student–advisor interaction quality, and three outcomes: sense of belonging, research self-efficacy, and research skills.

4 | THEORY-DRIVEN HYPOTHESES

Our first research question focuses on how the quality of student–advisor interactions may differ as a product of advisor gender and student–advisor gender match. We hypothesized that students would report higher quality advising experiences when their advisor shared their gender identity, in keeping with the broader literature on identity-based motivation theory and homophilous relationships. Additionally, we hypothesized that advisor gender would predict doctoral student outcomes, such that students would have more positive outcomes when their gender identity was shared with their advisor (Research Question 2). However, in regard to our third research question, we expected that this relationship may be mediated or moderated by the quality of student–advisor interactions, given the theorized importance of advisor socialization experiences (Weidman et al., 2001). We also expected that the role of advisor gender may change over time as students' relationships with their advisors evolve. Finally, while we expected that gender identity congruence may matter for all students, we hypothesized that women students may more strongly benefit from same-gender advising relationships, due to their gender identity being salient within the context of STEM doctoral training.

5 | METHODS

5.1 | Sample and data collection procedures

This study relied on data from a larger mixed-methods study of doctoral student development in the biological sciences, focusing on 4 years of longitudinal data from a cohort of 336 students who began their PhD programs in the biological sciences (i.e., microbiology, cellular biology, molecular biology, developmental biology, or genetics) in fall of 2014 at one of 53 universities across the United States. Of those universities, 42 hold an R1 Carnegie classification, seven hold R2, and four hold other classifications. To recruit this sample, researchers contacted the program directors and department chairs of the 100 largest biological sciences doctoral programs in the United States to describe the study and request that they inform incoming PhD students about the research project. Following, to diversify the prospective pool of participants, all

similar programs at public flagship universities (research intensive), historically Black colleges and universities, and Hispanic-serving institutions offering PhD programs were contacted. Those who agreed forwarded recruitment information on behalf of the study to students, who then contacted the research team expressing interest in participating. In instances where incoming cohorts were six students or more, campus visits were arranged for a member of the research team to present information to eligible students and answer questions during program orientation or an introductory seminar meeting.

Surveys were administered to students each year of their program, focusing on their socialization experiences (e.g., faculty interactions), research lab experiences, and key outcomes. Additionally, we solicited sole-authored writing samples annually. For the purpose of the present study, we restricted the sample to 281 students, removing those who left their doctoral program during the first 4 years of data collection.[1] The sample included 62% women and 38% men. The majority of participants were from white (58%) or Asian/Asian American (22%) racial/ethnic groups. Students from underrepresented minoritized racial/ethnic groups (i.e., Black, Latinx, and Indigenous students) made up 16% of the sample, and the remaining 4% of the sample indicated an "other" racial/ethnic identity. More specifically, the sample included 10% underrepresented racially minoritized women and 6% underrepresented racially minoritized men.

Because precise data on doctoral enrollment disaggregated by demographic characteristics and field are not available either nationally or from many individual institutions, it is not possible to determine precisely the extent to which our sample is representative. Data from the Survey of Earned Doctorates provides information only on degree recipients in a given year, so data are missing from students who either attrite or differ in their time to degree (National Center for Science and Engineering Statistics [NCSES], 2016). To offer our closest approximation, we provide field-matched data from 2014 (year of matriculation for our sample) as a basis for comparison (see NCSES, 2016). These population estimates reflect 56% women and 14% students from underrepresented and racially minoritized groups, which is highly similar to the distribution within our study sample.

5.2 | Endogenous variables

To assess the role of advisor gender, we measured three outcome variables: sense of belonging in the research lab, research self-efficacy, and demonstrated research skills, which were each measured annually and are described below.[2] These three outcomes were selected in accordance with our guiding theoretical framework and because of their prominence within prior studies of doctoral education (see Curtin et al., 2013; Feldon et al., 2019; Litson et al., 2021). Collectively, these three outcomes capture key affective and skill-based outcomes.

5.2.1 | Sense of belonging to one's lab

Sense of belonging in the lab was measured using a three-item composite variable, adapted from Bollen and Hoyle (1990), where students indicated their agreement with the following: I feel a sense of belonging to my lab/research group; I feel that I am a member of the lab/research group community; I see myself as part of the lab/research group community. Students responded on a scale ranging from 1 (Strongly Disagree) to 11 (Strongly Agree). The three items had a Cronbach's alpha ranging from 0.95 to 0.96; thus, items were added together to create a scale ranging from 3 (low sense of belonging) to 33 (high sense of belonging).

5.2.2 | Research self-efficacy

Research self-efficacy was measured using a composite variable made up of 10 items developed by Kardash (2000), asking students to rate their abilities in the following areas: understand contemporary concepts in your field; make use of the primary scientific research literature in your field; identify a specific question for investigation based on the research in your field; formulate a research hypothesis based on a specific question; design an experiment or theoretical test of hypotheses; understand the importance of "controls" in research; observe and collect data; statistically analyze data; interpret data by relating results to original hypotheses; and reformulate your original research hypotheses (as appropriate). Students responded on a scale ranging from 1 to 5. Items were added together to create a scale ranging from 10 (low self-efficacy) to 50 (high self-efficacy). The composite variable had a Cronbach's alpha ranging from 0.91 to 0.92.

5.2.3 | Demonstrated research skills

To begin to capture demonstrated research skills, we collected sole-authored writing samples (e.g., manuscripts in progress, dissertation proposals, etc.), utilizing an established rubric for assessing doctoral student research skills in the biological sciences (Feldon et al., 2019). Each participant selected and provided only one sample each year. An overall rate of 95% of participants submitted a writing sample in at least one of the 4 years. The response rates for research skills (i.e., writing samples) were 73% in year 1, 77% in year 2, 69% in year 3, and 52% in year 4. A missing values analysis showed that the missing data across time met the assumption for missing completely at random (MCAR; Little, 1988), $x^2(312) = 346.41$, $p = 0.09$.

Pairs of expert raters, drawn arbitrarily from a pool of eight with PhDs in relevant subfields of biology, blindly (i.e., raters were blind to the participants' characteristics, including their gender and their advisor's gender) and independently scored each writing sample using the rubric to measure discrete research skills. This rubric yielded intraclass correlations (ICC; two-way, random effects) for individual planks between 0.818 and 0.969. The rubric measured the following research skills: setting context for a study (ICC = 0.881), generating testable hypotheses (ICC = 0.911), establishing appropriate controls (ICC = 0.888), research/experimental design (ICC = 0.942), appropriate selection of data for analysis (ICC = 0.893), presentation of data (ICC = 0.953), data analysis (ICC = 0.818), drawing conclusions based on data (ICC = 0.873), exploring alternative interpretations of data (ICC = 0.883), identifying research design limitations (ICC = 0.898) generating implications for findings (ICC = 0.894), effective use of primary literature (ICC = 0.969), and overall writing quality (ICC = 0.869). Raters evaluated all criteria on a scale from 0 to 3.25, and ratings were averaged across raters for each skill. Ratings across skills (excluding overall writing quality) were averaged to create the research skills score. This scale had high reliability across years, McDonald's omega ranged from 0.91 to 0.95. This rubric measuring research skill has been established and utilized within prior studies, and the full text of the rubric is available online (see Feldon et al., 2019, table S1).

This approach to measuring research skills has a high level of face validity, as it entails experts examining authentic scholarly products according to shared criteria. The specific rubric used evolved from a previous version published by Timmerman et al. (2011) over multiple iterations used in a variety of studies (e.g., Feldon et al., 2011). Timmerman and colleagues established validity based on an extensive review of research on component scientific inquiry skills and scientific reasoning, other relevant published rubrics, comparisons with professional

research quality criteria, and multiple rounds of recursive feedback from stakeholders who also served as content experts. Across these studies, and in the current one, the measurement of research skills through written products may impact the interpretation of research skills. For this reason, overall writing ability was included as a criterion on the rubric, but writing ability ratings were not included in the composite measure of research skills used in the present study. If research skills were confounded with writing ability, the plank measuring writing skills discretely would have significant relationships with the same variables as the research skill planks. In our study, measured research skills do have significant relationships with relevant variables but writing skill scores do not (i.e., included 0 in the 95% confidence interval). Details of this analysis are reported in the results section.

5.3 | Exogenous variables

5.3.1 | Advisor characteristics

A standard practice in lab-based biological sciences doctoral programs is for students to participate in lab rotations during the first year before they select a permanent lab (and corresponding advisor) by the start of their second year (see Maher et al., 2018). Thus, after students indicated that they entered a permanent lab, we administered a short survey asking them to provide the name and rank (1 = non-tenure track, 2 = assistant, 3 = associate, 4 = full) of their advisor. To determine advisor gender, members of the research team reviewed each advisor's professional webpage, using gender pronouns from their biographical statement to identify the advisor's gender as a woman (pronouns she/her/hers) or man (pronouns he/him/his). This was used to compute a dichotomous measure of gender (0 = woman, 1 = man). Approximately a quarter of advisors in our sample were women and three-quarters were men.

5.3.2 | Advisor interaction quality

To better understand the role of advisor gender, we also account for the quality of student–advisor interactions. To measure this, we relied on a 12-item composite measure adapted from Barnes et al.'s (2011)[3] measure of graduate advisor function where students indicated their agreement with the following: my primary advisor is readily available to talk with me when needed; my primary advisor treats me with respect; my primary advisor teaches me strategies for succeeding in my field; my primary advisor cares about me as a person; my primary advisor helps me get to know other faculty members at the university; my primary advisor helps me get to know other faculty or professionals in my field; my primary advisor helps me navigate departmental politics; my primary advisor initiates discussions about the progress I am making toward my degree; my primary advisor is knowledgeable about formal degree requirements; my primary advisor provides information about career paths open to me; my primary advisor gives me constructive feedback on my progress toward degree completion; overall, my advisor performs the advising role to my expectations. For each item, students responded on a 3-point scale (1 = Disagree; 2 = Neither agree nor disagree; 3 = Agree), and items were summed to create a scale ranging from 12 (low advising quality) to 36 (high advising quality). Advisor interactions were measured in years 2, 3, and 4 (i.e., each year after students would have selected their

permanent advisors). Cronbach's alphas for this 12-item composite measure ranged from 0.83 to 0.86, indicating reliability.

5.3.3 | Student gender

Students reported their gender on the survey as a woman, man, or non-binary identity. No students in our analytic sample selected a non-binary gender identity, so we relied on a dichotomous measure of gender (0 = woman; 1 = man).

5.4 | Analytical procedures

5.4.1 | Research question one

In order to examine descriptive differences in student–advisor interactions by advisor and student gender, we first ran independent sample *t*-tests to examine mean differences in advisor interaction quality scores in years 2, 3, and 4 as a product of advisor gender. To determine the appropriateness of our analytical approach, we first examined homogeneity of variance of advisor interaction across advisor gender using Levene's test (Gastwirth et al., 2020), finding that variance across advisor gender for the *t*-tests were equal in the second (Levene's test = 2.45, $p = 0.12$) and fourth (Levene's test = 0.07, $p = 0.80$) years; there was less evidence for equality of variance in third year (Levene's test = 5.37, $p = 0.02$) of doctoral study, suggesting that a test assuming non-equal variances is more appropriate for this time point, so we adjusted the analysis accordingly by allowing variances across groups in the Mplus models to be freely estimated. Visually, data showed some skewness ($-1.68 < \text{skew} < 0.44$) and kurtosis ($-1.13 < \text{kurtosis} < 3.59$), which was accounted for in the analysis by using full-information maximum likelihood estimator with robust standard errors to account for non-normality.

Next, we used two-way analysis of variance (ANOVA) to examine whether advisor interaction quality differed as a product of student gender and advisor gender together (i.e., student–advisor gender interaction), evaluating the main effects of both student and advisor gender and further evaluating the interactive effect between student and advisor gender to evaluate gender congruence versus incongruence. Simple effects examining gender congruence and incongruence were calculated and interpreted only for significant *F*-statistics. These analyses were conducted by using model constraints in M*plus* version 8.3 (Muthén & Muthén, 1998–2019) to use a sandwich estimator (i.e., Type = Complex in M*plus*) to control for students nested within universities (code provided in Appendix A). This method of handling nested data was selected because our research questions were not related to level-two variables, and our goal was to simply control for variability at the university-level (Muthén & Satorra, 1995).

5.4.2 | Research questions two and three

The remaining two research questions ask about the extent to which advisor gender statistically predicts doctoral student outcomes (i.e., sense of belonging, self-efficacy, and research skills) and the extent to which those relationships might be mediated or moderated by student–advisor interactions. To examine these questions, we used structural equation modeling, which combined latent growth curve modeling, mediation analysis, and controlled for the hierarchical

nature of students nested within universities using a sandwich estimator via the command type = complex in M*plus* version 8.3 (Muthén & Muthén, 1998–2019). All mediation models used bootstrapping methods to account for the non-normal distribution of the mediated effects (Williams & MacKinnon, 2008). Importantly, this structural equation modeling approach was used to simultaneously (1) examine the impact of the potentially changing nature of both advisor interactions and the outcomes of interest using latent growth models and (2) test for indirect relationships using mediation analysis. Because doctoral students in lab-based biological sciences typically declare their permanent advisor at the beginning of their second year in the program (i.e., after completing lab rotations in the first year), we structured the analyses so that estimates are centered on year 2 (i.e., the first year after students would have selected their permanent advisor).

Analyses were conducted separately for each of the endogenous variables.[4] All models included faculty rank as a covariate, given prior research demonstrating how faculty and advisor interactions differ by rank (see Lechuga, 2011; Maher et al., 2013). To more closely examine the significance of student–advisor gender alignment, we used a multigroup framework to generate separate estimates by student gender. In instances where the estimates for women and men students were similar (i.e., all relationships were in the same direction and of similar magnitude), we removed the multigroup framework and reported estimates for all students in the aggregate. The code for these procedures is included in Appendix B.

To ensure the appropriateness of these procedures, statistical power[5] was evaluated by reviewing literature on latent growth mediation models. One large simulation study that considered many different model population parameters (Cheong, 2011) showed that a sample size of 200 had adequate statistical power to detect medium and large effects in latent growth mediation, if a large proportion of variance (0.80) was attributed to the growth factors. Given the high reliabilities of the scales and generally strong within construct correlations with our measures (Appendix C, Table C1), this condition holds for our models.

5.4.3 | Missing data

Some data were missing in the study, as is expected in longitudinal research (e.g., Laird, 1988). To reduce estimation bias, we used full information maximum likelihood estimation with robust standard errors to account for missing data in all analyses (Enders, 2010). No patterns of missingness were detected using Little's (1988) MCAR test, indicating that missing data would not introduce any bias into the analyses. Missingness for each variable is also shown in Appendix C.

6 | FINDINGS

Descriptive statistics were evaluated for all variables and are shown in Appendix C.

6.1 | Research question one: Gender differences in advisor interactions

We found that all students, both women and men, had significantly higher mean scores on advisor interaction quality when their advisor was a woman ($31.52 \leq M \leq 33.17$) relative to

when their advisor was a man ($30.27 \leq M \leq 31.67$); these differences were significant in both the second, $t(246) = 3.04$, $p = 0.002$, Cohen's $d = 0.39$, and third, $t(233) = 4.77$, $p < 0.001$, Cohen's $d = 0.62$, years of the doctoral program (i.e., the first 2 years after students would have identified their permanent advisor). However, this difference was just beyond the bounds of statistical significance in the fourth year of the program, $t(232) = 1.957$, $p = 0.05$.

Next, results of the 2 (student gender) × 2 (advisor gender) ANOVA revealed no significant interaction effect of advisor and student gender together in year 2, $F(1, 244) = 0.10$, $p = 0.75$, year 3, $F(1, 231) = 0.03$, $p = 0.86$, or year 4, $F(1, 230) = 0.06$, $p = 0.81$, of their doctoral program. The main effect of advisor gender across all years in the ANOVAs reflected the same finding as the *t*-test reported above. In other words, while all students reported higher-quality interactions when their advisor was a woman, having an advisor of the same or different gender as the student had no further bearing on how students reported the quality of their advisor interactions. Because these effects were not significant, post-hoc tests and simple effects were not evaluated.

6.2 | Research questions two and three: The indirect role of advisor gender

We found that, while advisor gender inconsistently predicted the outcome variables of interest, gender-outcome relationships were most often mediated by student–advisor interaction quality. For this reason, we report the results from the second research question (i.e., the role of advisor gender and student–advisor gender congruence) and third research question (i.e., the mediating role of advisor interaction quality) simultaneously. In no instances did advising quality moderate the relationship between advisor gender and the outcomes (i.e., the effect of advisor gender did not depend on advising quality). Effect sizes were compared to the recommended effect size interpretations reported in Ferguson (2009). When considering standardized β coefficients that are reported in the following results, the recommended minimum effect size of a practically significant effect is $\beta = 0.20$, a moderate effect is $\beta = 0.50$, and a strong effect is $\beta = 0.80$. Indirect effects are interpreted in relation to the strength of the combination of β effects.

6.2.1 | Sense of belonging

Preliminary analyses of sense of belonging revealed that model estimates were very similar for women and men; as such, we only report parameter estimates for the full sample. As shown in Figure 1, students who had women as advisors, regardless of their own gender, tended to report higher quality advising experiences in the second year of the program, $\beta = -0.18$, 95% CI [−0.33, 0.00], $p = 0.04$, though the effect was not practically significant according to Ferguson's (2009) guidelines. At the same time, advising quality in year 2 moderately and positively predicted sense of belonging in the same year, $\beta = 0.72$, 95% CI [0.50, 1.02]. However, advisor gender played no role in predicting the slope of advising quality, $\beta = 0.01$, 95% CI [−0.25, 0.24], though the slope of advising quality did moderately positively predict the slope of belonging scores, $\beta = 0.64$, 95% CI [0.30, 1.49]. In other words, students who reported increasingly positive advising quality over time also reported

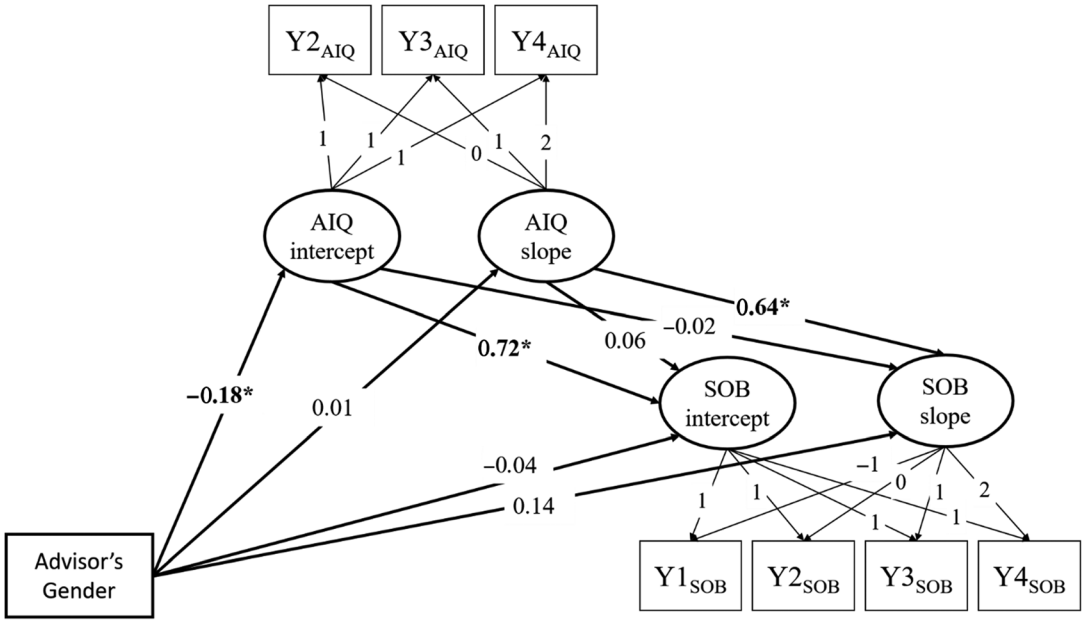

FIGURE 1 Structural equation model predicting sense of belonging. *Indicates that the confidence interval did not include 0; AIQ, advisor interaction quality; SOB, sense of belonging. The growth portion of the model shows non-standardized estimates while the structural (mediation) portion of the model shows standardized estimates. Rank was included as a covariate predicting both the mediator and outcome variables but is not included in the figure. All paths reflect direct effects, and indirect effects are noted in text. Model fit: $\chi^2 = 86.64$, df = 20, $p < 0.001$; RMSEA = 0.10; CFI = 0.84; SRMR = 0.05

increasing sense of belonging scores over time; this relationship was not statistically nor practically predicted by the gender of the advisor, $\beta = 0.14$, 95% CI [−0.20, 0.26]. Advisor gender indirectly predicted sense of belonging (intercept) through initial advisor interaction quality (intercept), indirect effect = −0.13, 95% CI [−0.33, −0.00]. That is, students who had women as advisors tended to have greater sense of belonging, but only because they also had higher-quality interactions with their advisors.

6.2.2 | Research self-efficacy

As was the case for the model predicting sense of belonging, the model estimates predicting self-efficacy were similar for men and women, so we report coefficients for the full sample (Figure 2). However, compared to the sense of belonging model, advisor gender did not seem to play a large role in predicting self-efficacy outcomes. While advisor gender showed a small predictive effect on advising quality intercepts—as it did in the sense of belonging model (i.e., advising quality in year 2 was higher when students had a woman advisor)—$\beta = -0.20$, 95% CI [−0.36, −0.02], advisor gender did not significantly predict self-efficacy scores directly or indirectly.

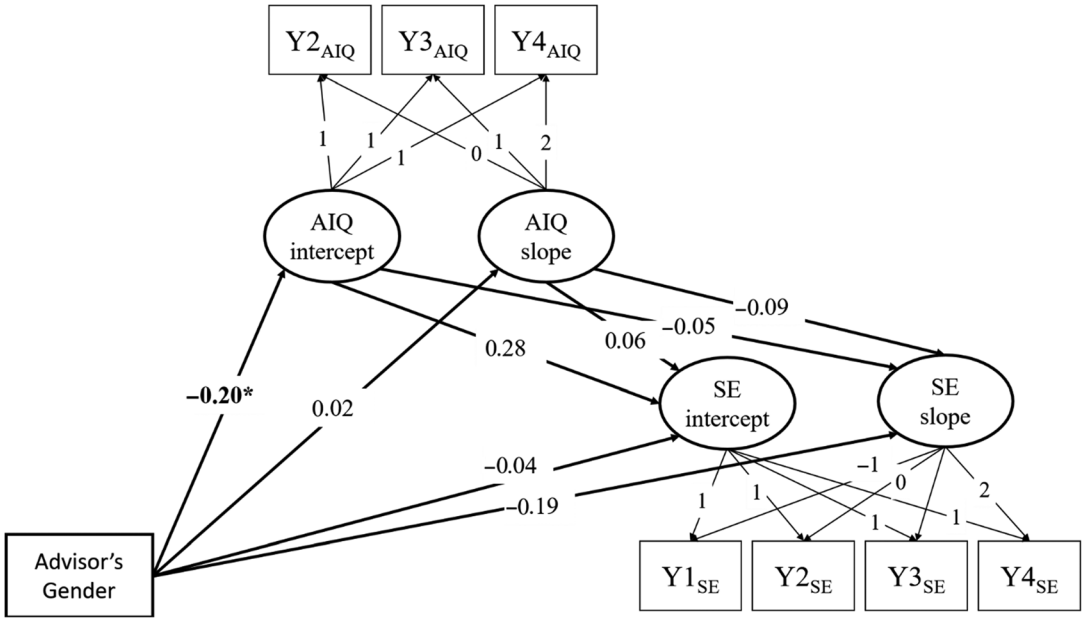

FIGURE 2 Structural equation model predicting self-efficacy. *Indicates that the confidence interval did not include 0; AIQ, advisor interaction quality; SE, self-efficacy. The growth portion of the model shows non-standardized estimates while the structural portion of the model shows standardized estimates. Rank was included as a covariate predicting both the mediator and outcome variables but is not included in the figure. All paths reflect direct effects, and indirect effects are noted in text. Model fit: $\chi^2 = 18.22$, df = 20, $p = 0.57$; RMSEA = 0.00; CFI = 1.00; SRMR = 0.05

6.2.3 | Research skills

Unlike the models predicting sense of belonging and research self-efficacy, the estimates predicting research skill scores did differ by student gender; thus, we present separate parameter estimates for women students and men students. As shown in Figure 3, both women and men had higher quality advising intercept scores when their advisors were women ($\beta = -0.23$, 95% CI [−0.49, −0.02] and $\beta = -0.20$, 95% CI [−0.38, −0.02], respectively), though these effects are rather small. However, the extent to which advisor gender and advisor interactions directly predicted research skills varied by student gender. Among men, the quality of advising experiences in year 2 (i.e., higher intercept scores) moderately positively predicted year 2 skills, $\beta = 0.57$, 95% CI [0.004, 1.12]. Advisor gender also directly predicted skills, such that men students who had women as advisors saw greater increases in their skills over time, relative to those who had men as advisors, $\beta = -0.55$, 95% CI [−0.98, −0.03], and this effect was moderate. No indirect effects were found. Notably, this was the only instance where advisor gender directly predicted the outcome variable in the present study. For women, advising quality in year 2 did not significantly predict increases in skill outcomes, $\beta = 0.20$, 95% CI [−0.18, 1.82]. Counterintuitively, women who had increasingly high-quality advising interactions over time tended to have decreasing skills over time (i.e., the slope of advising quality scores negatively predicted the slope of research skill scores), $\beta = -0.54$, 95% CI [−3.20, −0.18], and this effect was moderate. Results showed that advising quality did not mediate any relationship between advisor gender and research skill outcomes for women students.[6]

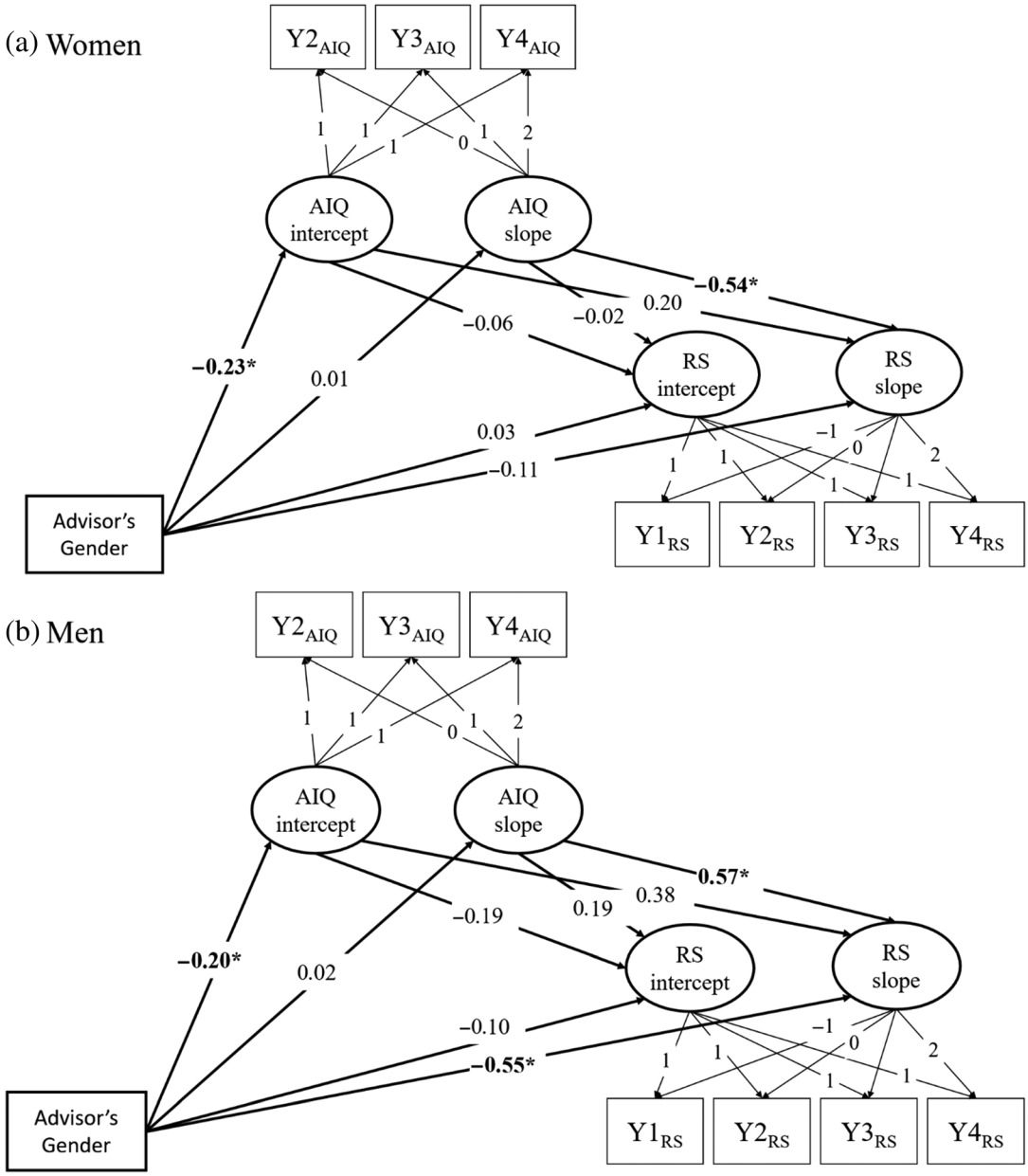

FIGURE 3 Structural equation model predicting research skills. Panel (a) presents the results for women and panel (b) presents the results for men. *Indicates that the confidence interval did not include 0; AIQ, advisor interaction quality; RS, research skills. The growth portion of the model shows non-standardized estimates while the structural portion of the model shows standardized estimates. Rank was included as a covariate predicting both the mediator and outcome variables but is not included in figures. All paths reflect direct effects, and indirect effects are noted in text. Model fit: $\chi^2 = 66.71$, df = 42, $p = 0.01$; RMSEA = 0.07; CFI = 0.89; SRMR = 0.08

7 | DISCUSSION

This study explores how students' experiences and outcomes descriptively differ as a product of their advisor's gender and the extent to which student–advisor gender alignment might predict key outcomes for doctoral students, particularly women. By examining these relationships across 4 years and employing mediation analysis, we provide new insights into the complex role of gender identity in doctoral student–advisor relationships. Our findings reveal differences in how students report interacting with their faculty advisors as a function of advisor gender, which is consistent with prior research documenting that women faculty tend to spend more time on graduate instruction than their men counterparts (Carrigan et al., 2011). The role of advisor gender also seems to change as students moved through their doctoral programs, with advisor gender potentially playing less of a role. Collectively, our findings underscore the importance of exploring the role of advisor gender in context and with a recognition that advising relationships and the salience of gender identity can change over time.

7.1 | Advisor gender functions similarly for women and men

Given the documented and theorized importance of shared identity characteristics in shaping student–faculty interactions (e.g., Solanki & Xi, 2018), particularly for students from minoritized groups (Burt, 2018), we expected that advisor gender would predict outcomes among women, even if the role of advisor gender was mediated by other aspects of student–advisor interactions. Somewhat in keeping with our hypotheses, women tended to report higher quality advising experiences when they had an advisor who shared their gender identity. Yet, men also appeared to benefit similarly from having an advisor who was a woman. In other words, while prior studies have concluded that women in doctoral programs may benefit from having an advisor who shares their gender identity, we found that *all* students reported higher quality experiences when their advisor was a woman. We conclude that advising experiences do not differ as a product of student–advisor gender identity congruence in the context of our study, though this certainly does not suggest that gendered dynamics are not present in doctoral training in the biological sciences. In fact, these findings suggest that gender inequities may persist in expectations and responsibilities of faculty (e.g., women advisors may be expected to take on more active mentoring than men advisors). This interpretation aligns with prior literature on faculty service and teaching expectations, documenting gender inequities such that women faculty are expected to spend more time on service and student-related activities (see Hanasono et al., 2019; O'Meara et al., 2017; Winslow, 2010).

Advisor interaction quality predicted the outcome variables differently. Although students who had women as advisors consistently reported more positive advising interactions, the quality of those interactions did not reliably predict outcomes across variables. Advisor interaction quality did predict more positive sense of belonging at specific time points, mediating the relationship between advisor gender and that outcome variable. Likewise, increases in perceived advising interaction quality predicted increases in sense of belonging (slope predicted slope). While the effects for self-efficacy and research skills are more complex, in general, our findings suggest that advising quality plays a stronger role in predicting outcomes than does advisor gender identity—and that this effect is similar for all students, regardless of their gender.

7.2 | The differential role of gender and advising quality of skill development

One of our more peculiar findings is that increasing interaction quality over time was positively associated with men's research skill growth but served as a negative predictor for women. While we cannot be sure of the reason for this finding, we believe this provides further evidence indicating the importance of longitudinal inquiry on advising relationships and their impacts on women in doctoral programs over time. At the same time, our findings demonstrating that men students had higher research skills when they had an advisor who was a woman also warrants further inquiry, especially given that this was the only direct effect of advisor gender in all of our analyses.

7.3 | Complicating existing theories of identity and graduate student development

Collectively, our findings present a complex picture of the ways in which advisor gender and interaction quality may impact various outcomes of interest for doctoral students. Typically, accounts of doctoral student socialization and development anticipate that high quality interactions with one's doctoral advisor would benefit the socialization process, yielding stronger sense of belonging, self-efficacy, and research skills (Austin & McDaniels, 2006; Weidman et al., 2001). Further, identity-based motivation theory (Solanki & Xi, 2018) posits that recognized similarities in identity between mentors and mentees should boost students' efforts and persistence with positive results. However, this study failed to consistently support these frameworks. In particular, having a woman advisor positively predicted research skill development for men students, playing no role in predicting skills for women students (direct or indirect). Put differently, we found no evidence that having a woman advisor *directly* predicts desirable outcomes for women graduate students.

While the focus of our study was on the role of student and advisor gender, we used socialization theory as a framework to examine these relationships. This guided our decision to control for advising quality in our analyses, as student–advisor interaction is a primary mechanism underlying socialization theory. Interestingly, advising quality did not consistently predict the outcomes explored in our study. In other words, our findings diverged from socialization theory, which is consistent with other recent longitudinal studies which have called socialization theory into question for its inability to explain doctoral student skill development and other key outcomes (see Feldon et al., 2019; Feldon, 2020). The present study provides further evidence suggesting that socialization theory does not adequately explain doctoral student development and outcomes, which has important implications for future research on doctoral training.

7.4 | Contextualizing advisor interactions

To further contextualize our findings, it is important to acknowledge recent scholarship on doctoral student mentorship. Recent literature documents the significance of "cascading mentorship" where senior peers (e.g., postdoctoral researchers, senior graduate students) often serve as the primary sources of active mentorship, which challenges traditional

notions of the one-on-one mentorship model between advisor and student (Golde et al., 2009). Like other emerging literature, our findings challenge the conventional cognitive apprenticeship framework, in which the quality of faculty mentorship received is asserted to drive graduate student development (e.g., Flores-Scott & Nerad, 2012; Gardner, 2010a; Littlefield et al., 2015). One recent study suggests that the likelihood of doctoral students improving their research skills over time is not affected by interactions with their supervising faculty members (Feldon et al., 2019). Instead, interactions with postdoctoral researchers in the laboratory environment substantially increase the odds of improvement, which aligns with a model of cascading, rather than one-to-one, mentorship. Thus, future research on gender identity congruence in doctoral training environments should also consider other mentors and socialization agents, including postdocs and peers.

7.5 | Disciplinary distinctions within STEM

When interpreting how our findings diverge from existing theories, one important difference between this study and previous work is that our participants were drawn from only one scientific discipline. This may be an important consideration for future work. While STEM is often invoked as an umbrella term to enhance the apparent generalizability of findings, it may obscure important cultural and contextual differences that impact the mechanisms of training and their effects on doctoral students. Given that the explicit emphasis of socialization theory is its ability to account for acculturation processes into a *specific* discipline (Austin & McDaniels, 2006) and that such disciplinary contexts yield notable differences in both process and outcomes (e.g., Gardner, 2010b, 2010c), future research should attend more deliberately to the potential impacts of discipline on findings and the suitability of general theories to specific contexts. Indeed, one reason for our divergent findings may be related to the unique context of the biological sciences, where women have reached parity in terms of graduate school enrollments but the professoriate continues to be dominated by men.

7.6 | Changing relationships over time

Finally, some of our findings may be explained by our longitudinal research design. Utilizing data on student experiences across 4 years highlighted changing relationships with advisor gender and advising over time. Most directly, students reported more positive interactions with advisors when their advisors were women in earlier years, but this difference was non-significant in year 4 of our study. As such, extrapolation of sustained effects based on fewer years of data could have led to inaccurate assumptions about these effects over time, failing to capture the complexity and dynamic nature of gender identity within advising relationships. These findings are not particularly surprising, as identity-based motivation theory (Oyserman, 2009) posits that identities are dynamic and identity salience can change over time, especially as environments evolve (e.g., students build deeper relationships with advisors over time, students identify other mentors who offer support, etc.).

7.7 | Limitations and suggestions for future research

When interpreting the above findings, it is important to consider the limitations of the present work. First, there may be myriad factors that shape the impact of advisor gender which were not captured in the present study, such as the departmental climate or the representation of women in the lab or doctoral program. Similarly, other sociodemographic factors may shape student–advisor interactions (e.g., Burt, 2018; Miles et al., 2020), including, race, class, sexual orientation, age, and other salient identities. Most importantly, future studies should consider how the intersectionality of race, ethnicity, and gender may further shape advising relationships and student outcomes, especially in light of other recent research demonstrating the impact of instructor racial diversity on whether or not Students of Color pursue and persist in STEM majors (Bottia et al., 2018). Within such inquiry, studies should consider larger contexts that shape advising relationships. Some recent research has documented ways in which Women of Color identify mentors within counterspaces in STEM programs, especially when they experience microaggressions from their formal advisors (Ong et al., 2018). Accordingly, it is important to study the intersection of gender and race as it relates to multiple mentoring relationships in STEM simultaneously. Indeed, the lack of significance of gender identity congruence in the present inquiry could be explained by other relationships students seek out within their STEM department, which we were not able to capture. Additionally, the composition of our sample limited our ability to examine gender identity in a non-binary way. We relied on a binary measure of both student and advisor gender; importantly, for non-binary or trans* identifying students, a lack of faculty representation and role models could mean that finding a mentor with a shared gender identity is particularly elusive, something that should be accounted for in future research.

Beyond considering additional study contexts and identities, future research should also consider additional advisor characteristics. More specifically, while we examine faculty rank as a covariate, future studies might consider other advisor characteristics, such as the number of students the faculty member advises or how long the advisor has been at the current institution. Additionally, our measure of advisor interaction quality relied on students' self-reports. Thus, student–faculty interaction quality may be confounded with gender differences in how students *rate* their advisors (see Feldon, 2020). Specifically, some research suggests that students may be more critical of women faculty in their evaluations of instructors (e.g., Superson, 2002). If this was true in the case of our study, our findings may understate how advisor interactions differ as a product of advisor gender. For these and other reasons, our reliance on self-reported survey data makes it difficult for us to draw causal inferences from the findings.

Additionally—while written research products serve as an authentic representation of research skills—our reliance on written research products to assess research skills may exclude certain salient skills, such as bench techniques. Further, the nature of writing products shared each year differed (e.g., students sometimes shared a journal article in one year and a dissertation proposal in another year); this may have contributed to additional variance in scores due to factors not directly related to research skills.

Ultimately, the ways in which gender dynamics play out in graduate training contexts are nuanced and multi-layered, as are the evolving relationships between faculty advisors and their students (Noy & Ray, 2012). Individually held expectations regarding appropriate and optimal roles in this context inherently color the actions and reactions of both students and faculty as they jointly navigate their mentoring relationships (Zhao et al., 2007). Thus, future research may benefit from more nuanced measures of attitudes and beliefs related to gender, faculty, and doctoral student roles to more clearly identify and control for differences that may impact

reported perceptions of advisement interaction quality. Likewise, further investigation of how gendered differences in expectations, conceptions, and execution of mentoring practices might manifest among faculty may shed further light on the current findings.

8 | CONCLUSION

While previous research documents that women in STEM doctoral programs tend to fare better when they have an advisor who shares their gender identity, the present study provides additional insights into the role of advisor gender for doctoral students within the biological sciences. In this study, advisor gender indirectly predicted some desirable outcomes for women students as a function of perceived advisor interaction quality. However, this relationship was also present for men, suggesting that all students may have more positive outcomes when their advisor is a woman; to be clear, the positive effects of having an advisor who is a woman were largely indirect (i.e., mediated by advising quality). Further, student outcomes are certainly shaped by many factors beyond gender identity. We conclude that doctoral programs should work to ensure that all faculty are providing a high-quality doctoral training experience. This will require faculty and administrators to consider *why* students might report lower quality experiences when their advisor is a man. As a first step, administrators and faculty could collaborate to develop systems for making faculty contributions to teaching and mentoring visible and transparent (see O'Meara et al., 2017), ensuring that they are holding all faculty to equitable standards for providing high-quality advising.

ACKNOWLEDGMENTS

The authors gratefully acknowledge the support of the National Science Foundation. This material is based on work supported under Awards 1431234, 1431290, and 1760894. The authors wish to thank Dr. Annie Wofford who reviewed an earlier version of this manuscript.

ORCID

Jennifer M. Blaney https://orcid.org/0000-0002-9146-6425
David F. Feldon https://orcid.org/0000-0003-3268-5764
Kaylee Litson https://orcid.org/0000-0003-1296-4811

ENDNOTES

[1] Among the original $n = 336$ students in our sample, 55 were removed because they left their doctoral program during the first four years of data collection, 40 of whom left in the very early stages of their program before identifying a "permanent" advisor following lab rotations. The demographic characteristics of students who left the study closely mirror those of the analytic sample.

[2] For each endogenous variable, we evaluated the dimensionality of the construct by performing parallel analyses (Horn, 1965), using the fa.parallel function in the R package, psych (Revelle, 2017). Analyses were conducted using 500 iterations, and results show that a one-component solution was generally supported for variables across years. Research skills in year one were the only analysis that resulted in a solution suggesting more than one component (two-components). Altogether, these analyses support using a single composite score for each endogenous variable.

[3] As a preliminary step, we used confirmatory factor analysis on the items developed by Barnes et al. (2011) which resulted in the 12-item composite variable used in the present study.

[4] Analyses were conducted separately for each endogenous variable for two reasons: first, the research questions do not ask about the relationships among the outcome variables. Second, latent growth mediation models are

power intensive, and adding more complexity reduces power to detect effects, which requires increasing the sample size needed to obtain interpretable effects.

[5] In response to a reviewer request, we conducted a power analysis using expected estimates rather than the obtained results. We adopted this strategy, because so-called post-hoc power analysis methods provide no useful information due to the robust relationship between statistical power and alpha (against which p is compared, resulting in statistical significance). Further, a post-hoc power analysis using obtained results is not recommended to evaluate power of known statistical estimates due to this functional relationship. We recognize our bias in knowing the results before conducting a power analysis and examined power of the current sample size, $n = 281$, including missing data that increased from 0.1 to 0.5 across time, to detect a- and b-paths ranging from 0.3 to 0.5 and indirect effects ranging from 0.09 to 0.25. These effects represent relatively small to medium effects. Statistical power to detect the indirect effects given these other known values averaged 0.81. Statistical power to detect specific a- and b-paths averaged 0.89.

[6] To ensure that these significant findings were not a result of research skills being confounded with writing ability, the models reported in this section were separately evaluated using writing skill scores (in place of research skill scores). Results showed that all effects between advisor interaction quality and writing quality for men and women included 0 in the 95% CI. Likewise, all direct and indirect effects between advisor gender and writing quality for men and women included 0 in the 95% CI. Accordingly, we conclude that significant effects observed for research skills could not be explained by writing quality.

REFERENCES

Austin, A. E., & McDaniels, M. (2006). Preparing the professoriate of the future: Graduate student socialization for faculty roles. In J. C. Smart (Ed.), *Higher education* (pp. 397–456). Springer.

Baird, L. L. (1995). Helping graduate students: A graduate adviser's view. *New Directions for Student Services*, *1995*(72), 25–32.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. Macmillan.

Barnes, B. J., & Austin, A. E. (2009). The role of doctoral advisors: A look at advising from the advisor's perspective. *Innovative Higher Education*, *33*(5), 297–315.

Barnes, B. J., Chard, L. A., Wolfe, E. W., Stassen, M. L., & Williams, E. A. (2011). An evaluation of the psychometric properties of the graduate advising survey for doctoral students. *International Journal of Doctoral Studies*, *6*, 1–17.

Bollen, K. A., & Hoyle, R. H. (1990). Perceived cohesion: A conceptual and empirical examination. *Social Forces*, *69*, 479–504.

Bottia, M. C., Mickelson, R. A., Giersch, J., Stearns, E., & Moller, S. (2018). The role of high school racial composition and opportunities to learn in students' STEM college participation. *Journal of Research in Science Teaching*, *55*(3), 446–476.

Burt, B. A. (2018). Toward a theory of engineering professorial intentions: The role of research group experiences. *American Educational Research Journal*, *56*(2), 289–332.

Carlone, H., Scott, C., & Lowder, C. (2014). Becoming (less) scientific: A longitudinal study of students' identity work from elementary to middle school science. *Journal of Research in Science Teaching*, *51*(7), 836–869.

Carrigan, C., Quinn, K., & Riskin, E. A. (2011). The gendered division of labor among STEM faculty and the effects of critical mass. *Journal of Diversity in Higher Education*, *4*(3), 131.

Cheong, J. (2011). Accuracy of estimates and statistical power for testing mediation in latent growth curve modeling. *Structural Equation Modeling*, *18*(2), 195–211.

Croom, N. N. (2017). Promotion beyond tenure: Unpacking racism and sexism in the experiences of Black womyn professors. *The Review of Higher Education*, *40*(4), 557–583.

Curtin, N., Malley, J., & Stewart, A. J. (2016). Mentoring the next generation of faculty: Supporting academic career aspirations among doctoral students. *Research in Higher Education*, *57*(6), 714–738.

Curtin, N., Stewart, A. J., & Ostrove, J. M. (2013). Fostering academic self-concept: Advisor support and sense of belonging among international and domestic graduate students. *American Educational Research Journal*, *50*(1), 108–137.

De Welde, K., & Laursen, S. (2011). The glass obstacle course: Informal and formal barriers for women Ph. D. students in STEM fields. *International Journal of Gender, Science, and Technology*, *3*(3), 571–595.

Enders, C. K. (2010). *Applied missing data analysis*. Guilford Publications.

Etzkowitz, H., Kemelgor, C., & Uzzi, B. (2000). *Athena unbound: The advancement of women in science and technology*. Cambridge University Press.

Feldon, D. F. (2020). Implications of measurement issues for advancing the socialization framework. In L. DeAngelo & J. C. Weidman (Eds.), *Socialization in higher education and the early career: Theory, research and application* (pp. 287–310). Springer International Publishing AG.

Feldon, D. F., Litson, K., Jeong, S., Blaney, J. M., Kang, J., Miller, C., Griffin, K., & Roksa, J. (2019). Postdocs' lab engagement predicts trajectories of PhD students' skill development. *Proceedings of the National Academy of Sciences*, *116*(42), 20910–20916.

Feldon, D. F., Maher, M. A., & Timmerman, B. E. (2010). Performance-based data in the study of STEM Ph.D. education. *Science*, *329*(5989), 282–283.

Feldon, D. F., Peugh, J., Timmerman, B. E., Maher, M. A., Hurst, M., Strickland, D., Gilmore, J. A., & Stiegelmeyer, C. (2011). Graduate students' teaching experiences improve their methodological research skills. *Science*, *333*(6045), 1037–1039.

Ferguson, C. J. (2009). An effect size primer: A guide for clinicians and researchers. *Professional Psychology: Research and Practice*, *40*(5), 532–538.

Flores-Scott, E. M., & Nerad, M. (2012). Peers in doctoral education: Unrecognized learning partners. *New Directions for Higher Education*, *2012*(157), 73–83.

Fox, M. F. (2006). Gender, hierarchy, and science. In J. S. Chafetz (Ed.), *Handbook of the sociology of gender* (pp. 441–457). Springer.

Freeman, T. M., Anderman, L. H., & Jensen, J. M. (2007). Sense of belonging in college freshmen at the classroom and campus levels. *The Journal of Experimental Education*, *75*(3), 203–220.

Gardner, S. K. (2010a). Keeping up with the Joneses: Socialization and culture in doctoral education at one striving institution. *The Journal of Higher Education*, *81*(6), 658–679.

Gardner, S. K. (2010b). Faculty perspectives on doctoral student socialization in five disciplines. *International Journal of Doctoral Studies*, *5*, 39–53.

Gardner, S. K. (2010c). Contrasting the socialization experiences of doctoral students in high- and low-completing departments: A qualitative analysis of disciplinary contexts at one institution. *The Journal of Higher Education*, *81*, 61–81.

Gardner, S. K., & Holley, K. A. (2011). "Those invisible barriers are real": The progression of first-generation students through doctoral education. *Equity & Excellence in Education*, *44*(1), 77–92.

Gastwirth, J. L., Gel, Y. R., Hui, W. L. W., Lyubchich, V., Miao, W., Noguchi, K., & Lyubchich, V. (2020). Lawstat: Tools for biostatistics, public policy, and law, v. 3.4. https://cran.r-project.org/package=lawstat

Gaule, P., & Piacentini, M. (2018). An advisor like me? Advisor gender and post-graduate careers in science. *Research Policy*, *47*(4), 805–813.

Golde, C. M., Conklin Bueschel, A., Jones, L., & Walker, G. E. (2009). Advocating apprenticeship and intellectual community: Lessons from the Carnegie initiative on the doctorate. In R. G. Ehrenberg & C. V. Kuh (Eds.), *Doctoral education and faculty of the future* (pp. 53–64). Cornell University Press.

Goodenow, C. (1993). Classroom belonging among early adolescent students: Relationships to motivation and achievement. *Journal of Early Adolescence*, *13*, 21–43.

Gopaul, B. (2016). Applying cultural capital and field to doctoral student socialization. *International Journal for Researcher Development*, *7*(1), 46–62.

Gumpertz, M., Durodoye, R., Griffith, E., & Wilson, A. (2017). Retention and promotion of women and underrepresented minority faculty in science and engineering at four large land grant institutions. *PLoS One*, *12*(11), e0187285.

Hackett, E., & Rhoten, D. (2009). The Snowbird charrette: Integrative interdisciplinary collaboration in environmental research design. *Minerva*, *47*, 407–440.

Hanasono, L. K., Broido, E. M., Yacobucci, M. M., Root, K. V., Peña, S., & O'Neil, D. A. (2019). Secret service: Revealing gender biases in the visibility and value of faculty service. *Journal of Diversity in Higher Education*, *12*(1), 85.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*, 179–185.

Jackson, P., & Seiler, G. (2013). Science identity trajectories of latecomers to science in college. *Journal of Research in Science Teaching*, *50*(7), 826–857.

Kachchaf, R., Ko, L., Hodari, A., & Ong, M. (2015). Career–life balance for women of color: Experiences in science and engineering academia. *Journal of Diversity in Higher Education*, *8*(3), 175.

Kardash, C. (2000). Evaluation of undergraduate research experience: Perceptions of undergraduate interns and the faculty mentors. *Journal of Educational Psychology*, *92*, 191–201.

Laird, N. M. (1988). Missing data in longitudinal studies. *Statistics in Medicine*, *7*(1–2), 305–315. https://doi.org/10.1002/sim.4780070131

Lawson, A. E. (2008). *Biology: An inquiry approach* (2nd ed.). Kendall/Hunt Publishing Company.

Le, A. T., LaCost, B. Y., & Wismer, M. (2016). International female graduate students' experience at a midwestern university: Sense of belonging and identity development. *Journal of International Students*, *6*(1), 128–152.

Lechuga, V. M. (2011). Faculty-graduate student mentoring relationships: Mentors' perceived roles and responsibilities. *Higher Education*, *62*, 757–771.

Lent, R. W., Brown, S. D., & Larkin, K. C. (1986). Self-efficacy in the prediction of academic performance and perceived career options. *Journal of Counseling Psychology*, *33*(3), 265.

Leslie, S., Cimpian, A., Meyer, M., & Freeland, E. (2015). Expectations of brilliance underlie gender distributions across academic disciplines. *Science*, *347*(6219), 262–265.

Little, R. (1988). A test of missing completely at random for multivariate data with missing values. *Journal of the American Statistical Association*, *83*(404), 1198–1202.

Litson, K., Blaney, J. M., & Feldon, D. F. (2021). Understanding the transient nature of STEM doctoral students' research self-efficacy across time: Considering the role of gender, race, and first-generation college status. *Frontiers in Psychology*, *12*, 28.

Littlefield, C. M., Taddei, L. M., & Radosh, M. E. (2015). Organic collaborative teams: The role of collaboration and peer to peer support for part-time doctoral completion. *International Journal of Doctoral Studies*, *10*, 129–142.

Lockhart, J. W. (2021). Paradigms of sex research and women in STEM. *Gender & Society*, *35*(3), 449–475.

Maher, M., Wofford, A., Roksa, J., & Feldon, D. F. (2018). Exploring early exits: Doctoral attrition in the biomedical sciences. *Journal of College Student Retention: Research, Theory, & Practice*, *22*, 152102511773687.

Maher, M., Wofford, A., Roksa, J., & Feldon, D. F. (2020). Finding a fit: Biological science doctoral students' selection of a principal investigator and research laboratory. *CBE—Life Sciences Education*, *19*(3), ar31.

Maher, M. A., Timmerman, B. E., Feldon, D. F., & Strickland, D. (2013). Factors affecting the occurrence of faculty-doctoral student coauthorship. *The Journal of Higher Education*, *84*, 121–143.

Main, J. B. (2014). Gender homophily, Ph.D. completion and time to degree in the humanities and humanistic social sciences. *The Review of Higher Education*, *37*(3), 349–375.

Main, J. B. (2018). Kanter's theory of proportions: Organizational demography and PhD completion in science and engineering departments. *Research in Higher Education*, *59*(8), 1059–1073.

McPherson, M., Smith-Lovin, L., & Cook, J. M. (2001). Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, *27*(1), 415–444.

Miles, M. L., Brockman, A. J., & Naphan-Kingery, D. E. (2020). Invalidated identities: The disconfirming effects of racial microaggressions on Black doctoral students in STEM. *Journal of Research in Science Teaching*, *57*(10), 1608–1631.

Moss-Racusin, C. A., Dovidio, J. F., Brescoll, V. L., Graham, M. J., & Handelsman, J. (2012). Science faculty's subtle gender biases favor male students. *Proceedings of the National Academy of Sciences*, *109*, 16474–16479.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology*, *38*(1), 30.

Muthén, B., & Satorra, A. (1995). Complex sample data in structural equation modeling. *Sociological Methodology*, *25*, 267–316.

Muthén, L. K., & Muthén, B. O. (1998–2018). *Mplus user's guide* (8th ed.). Muthén & Muthén.

National Center for Science and Engineering Statistics. (2016). Survey of Graduate Students and Postdoctorates in Science and Engineering, Fall 2014. http://ncsesdata.nsf.gov/gradpostdoc/

National Science Foundation. (2018). Doctoral sciences and engineers in academia. https://www.nsf.gov/statistics/2018/nsb20181/report/sections/academic-research-and-development/doctoral-scientists-and-engineers-in-academia

National Science Foundation, Division of Science Resources Statistics. (2018). Graduate education, enrollment, and degrees in the United States. https://www.nsf.gov/statistics/2018/nsb20181/report/sections/higher-education-in-science-and-engineering/graduate-education-enrollment-and-degrees-in-the-united-states

Nelson, D. J., Brammer, S. C. N., & Rhoads, H. (2007). A national analysis of minorities in science and engineering faculties at research universities. Report. http://w.astro.berkeley.edu/∼kalas/ethics/documents/diversity/nelson07.pdf

Nettles, M. T., & Millett, C. M. (2006). *Three magic letters: Getting to PhD*. The Johns Hopkins University Press.

Noy, S., & Ray, R. (2012). Graduate students' perceptions of their advisors: Is there systematic disadvantage in mentorship? *The Journal of Higher Education*, *83*, 876–914.

O'Meara, K., Kuvaeva, A., Nyunt, G., Waugaman, C., & Jackson, R. (2017). Asked more often: Gender differences in faculty workload in research universities and the work interactions that shape them. *American Educational Research Journal*, *54*(6), 1154–1186.

Ong, M., Smith, J. M., & Ko, L. T. (2018). Counterspaces for women of color in STEM higher education: Marginal and central spaces for persistence and success. *Journal of Research in Science Teaching*, *55*(2), 206–245.

Ostrove, J. M., Stewart, A. J., & Curtin, N. L. (2011). Social class and belonging: Implications for graduate students' career aspirations. *The Journal of Higher Education*, *82*(6), 748–774.

Overall, N. C., Deane, K. L., & Peterson, E. R. (2011). Promoting doctoral students' research self-efficacy: Combining academic guidance with autonomy support. *Higher Education Research & Development*, *30*(6), 791–805.

Oyserman, D. (2009). Identity-based motivation: Implications for action-readiness, procedural-readiness, and consumer behavior. *Journal of Consumer Psychology*, *19*(3), 250–260.

Oyserman, D. (2014). Identity-based motivation: Core processes and intervention examples. In *Motivational interventions*. Emerald Group Publishing Limited.

Paglis, L. L., Green, S. G., & Bauer, T. N. (2006). Does advisor mentoring add value? A longitudinal study of mentoring and doctoral student outcomes. *Research in Higher Education*, *47*(4), 451–476.

Pezzoni, M., Mairesse, J., Stephan, P., & Lane, J. (2016). Gender and the publication output of graduate students: A case study. *PLoS One*, *11*(1), e0145146.

Pittman, L. D., & Richmond, A. (2008). University belonging, friendship quality, and psychological adjustment during the transition to college. *The Journal of Experimental Education*, *76*(4), 343–362.

Pugh, K. J., Paek, S. H., Phillips, M. M., Sexton, J. M., Bergstrom, C. M., Flores, S. D., & Riggs, E. M. (2021). Predicting academic and career choice: The role of transformative experience, connection to instructor, and gender accounting for interest/identity and contextual factors. *Journal of Research in Science Teaching*, *58*, 822–851.

Revelle, W. R. (2017). psych: Procedures for Psychological, Psychometric, and Personality Research. *R package version 2.1.9*. https://cran.r-project.org/web/packages/psych/index.html

Robnett, R. D., Chemers, M. M., & Zurbriggen, E. L. (2015). Longitudinal associations among undergraduates' research experience, self-efficacy, and identity. *Journal of Research in Science Teaching*, *52*(6), 847–867.

Rosser, S. V. (2018). Breaking into the lab: Engineering progress for women in science and technology. *International Journal of Gender, Science and Technology*, *10*(2), 213–232.

Sallee, M. W. (2011). Performing masculinity: Considering gender in doctoral student socialization. *The Journal of Higher Education*, *82*(2), 187–216.

Sax, L. J. (2008). *The gender gap in college: Maximizing the developmental potential of women and men*. Jossey-Bass.

Sheltzer, J. M., & Smith, J. C. (2014). Elite male faculty in the life sciences employ fewer women. *Proceedings of the National Academy of Sciences*, *111*(28), 10107–10112.

Solanki, S. M., & Xi, D. (2018). Looking beyond academic performance: The influence of instructor gender on student motivation in STEM fields. *American Educational Research Journal*, *55*(4), 801–835.

Strayhorn, T. L. (2012). *College students' sense of belonging: A key to educational success for all students*. Routledge.

Superson, A. M. (2002). Sexism in the classroom: The role of gender stereotypes in the evaluation of female faculty. In A. M. Superson & A. E. Cudd (Eds.), *Theorizing backlash: Philosophical reflections on the resistance to feminism*. Rowman & Littlefield.

Tenenbaum, H. R., Crosby, F. J., & Gliner, M. D. (2001). Mentoring relationships in graduate school. *Journal of Vocational Behavior*, *59*(3), 326–341.

Timmerman, B. C., Feldon, D., Maher, M., Strickland, D., & Gilmore, J. (2013). Performance-based assessment of graduate student research skills: Timing, trajectory, and potential thresholds. *Studies in Higher Education*, *38*(5), 693–710.

Timmerman, B. C., Strickland, D., Johnson, R., & Payne, J. (2011). Development of a 'universal' rubric for assessing undergraduates' scientific reasoning skills using scientific writing. *Assessment & Evaluation in Higher Education*, *36*(5), 509–547.

Weidman, J. C., & DeAngelo, L. (2020). Toward a 21st century socialization model of higher education's impact on students. In J. C. Weidman & L. DeAngelo (Eds.), *Socialization in higher education and the early career: Theory, research, and application* (pp. 311–323). Springer.

Weidman, J. C., & Stein, E. L. (2003). Socialization of doctoral students to academic norms. *Research in Higher Education*, *44*(6), 641–656.

Weidman, J. C., Twale, D. J., & Stein, E. L. (2001). *Socialization of graduate and professional students in higher education: A perilous passage? Report*. Jossey-Bass.

West, C., & Zimmerman, D. H. (1987). Doing gender. *Gender and Society*, *1*(2), 125–151.

White, J., & Nonnamaker, J. (2008). Belonging and mattering: How doctoral students experience community. *NASPA Journal*, *45*(3), 350–372.

Williams, J., & MacKinnon, D. P. (2008). Resampling and distribution of the product methods for testing indirect effects in complex models. *Structural Equation Modeling*, *15*(1), 23–51.

Winslow, S. (2010). Gender inequality and time allocations among academic faculty. *Gender & Society*, *24*(6), 769–793.

Wofford, A. M., & Blaney, J. M. (2021). (Re)shaping the socialization of scientific labs: Understanding women's doctoral experiences in STEM lab rotations. *The Review of Higher Education*, *44*(3), 357–386.

Zhao, C. M., Golde, C. M., & McCormick, A. C. (2007). More than a signature: How advisor choice and advisor behaviour affect doctoral student satisfaction. *Journal of Further and Higher Education*, *31*(3), 263–281.

Zimmerman, B. J., Bandura, A., & Martinez-Pons, M. (1992). Self-motivation for academic attainment: The role of self-efficacy beliefs and personal goal setting. *American Educational Research Journal*, *29*(3), 663–676.

How to cite this article: Blaney, J. M., Feldon, D. F., & Litson, K. (2022). Student and advisor gender identity in STEM doctoral programs: Examining longitudinal and mediating effects with latent growth models. *Journal of Research in Science Teaching*, *59*(8), 1416–1446. https://doi.org/10.1002/tea.21761

APPENDIX A

MPLUS MODEL FOR INDEPENDENT SAMPLES *T*-TEST (YEAR 2)

```
VARIABLE:
      names are ID PartID univ p_gender s_gender Rank
      Y1_FI_O Y2_FI_O Y3_FI_O Y4_FI_O
      Y1_FI_Q Y2_FI_Q Y3_FI_Q Y4_FI_Q
      Y1LTSRT Y2LTSRT Y3LTSRT Y4LTSRT
      Y1SE Y2SE Y3SE Y4SE
      Y1Lab Y2Lab Y3Lab Y4Lab
      Y1PubNum Y2PubNum Y3PubNum Y4PubNum
      Y4Pubtot PI_Gender_Match Mean_FI_Q Mean_FI_O
      Y2_AI_Q Y3_AI_Q Y4_AI_Q
      Y1Skill Y2Skill Y3Skill Y4Skill;

      usevar = p_gender Y2_AI_Q weight;
      cluster = univ;
      weight = weight;
      missing are all (-99);
      grouping = p_gender (1 = Man_Advisor 0 = Woman_Advisor)

DEFINE:
      weight = 1;

ANALYSIS:
      TYPE=COMPLEX MGROUP;
      CONVERGENCE = .000000000001;

MODEL:
      Y2_AI_Q (1);
      [Y2_AI_Q];

      Model Man_Advisor: [Y2_AI_Q] (A1);
      Model Woman_Advisor: [Y2_AI_Q] (A2);

MODEL CONSTRAINT:
      new(diff);
      diff = A1-A2;

OUTPUT:
      stdyx cinterval(bcbootstrap);
```

APPENDIX B

MPLUS MODEL FOR TWO-WAY ANOVA (YEAR 2)

```
VARIABLE:
      names are ID PartID univ p_gender s_gender pXsgend Rank
      Y1_FI_O Y2_FI_O Y3_FI_O Y4_FI_O
      Y1_FI_Q Y2_FI_Q Y3_FI_Q Y4_FI_Q
      Y1LTSRT Y2LTSRT Y3LTSRT Y4LTSRT
      Y1SE Y2SE Y3SE Y4SE
      Y1Lab Y2Lab Y3Lab Y4Lab
      Y1PubNum Y2PubNum Y3PubNum Y4PubNum
      Y4Pubtot PI_Gender_Match Mean_FI_Q Mean_FI_O
      Y2_AI_Q Y3_AI_Q Y4_AI_Q
      Y1Skill Y2Skill Y3Skill Y4Skill;

      usevar = pXsgend Y2_AI_Q weight;
      cluster = univ;
      weight = weight;
      missing are all (-99);
      GROUPING = pXsgend (1 = WS_WA 2 = WS_MA 3 = MS_WA 4 = MS_MA);

DEFINE:
      weight = 1;

ANALYSIS:
      TYPE= COMPLEX MGROUP;
      CONVERGENCE = .00000000000000001;

Model:
      Y2_AI_Q* (a);
      [Y2_AI_Q*];
Model WS_WA:
      [Y2_AI_Q*] (DV_k11);
      Y2_AI_Q (a);
Model WS_MA:
      [Y2_AI_Q*] (DV_k12);
      Y2_AI_Q (a);
Model MS_WA:
      [Y2_AI_Q*] (DV_k21);
      Y2_AI_Q (a);
Model MS_MA:
      [Y2_AI_Q*] (DV_k22);
      Y2_AI_Q (a);

MODEL CONSTRAINT:
!Compute group & grand means
new(DV_kX1 DV_kX2);
  DV_kX1=(DV_k11+DV_k21)/2; !women advisor group mean
  DV_kX2=(DV_k12+DV_k22)/2; !men advisor group mean

new(DV_k1X DV_k2X);
  DV_k1X=(DV_k11+DV_k12)/2; !women student group mean
  DV_k2X=(DV_k21+DV_k22)/2; !men student group mean

new(DV_kXX);
  DV_kXX=(DV_k1X+DV_k2X)/2; !grand mean

!SS for F statistics
new(ss_stuG ss_advG ss_int ss_tot ss_res);
  ss_stuG=(154*(DV_k1X-DV_kXX)**2+94*(DV_k2X-DV_kXX)**2); !154= nWS, 94= nMS
  ss_advG=(66*(DV_kX1-DV_kXX)**2+182*(DV_kX2-DV_kXX)**2); !66= nWA, 182= nMA
  ss_int=
  45*(DV_k11+DV_kXX-DV_k1X-DV_kX1)**2+ !45 = n of women stu, women adv
  109*(DV_k12+DV_kXX-DV_k1X-DV_kX2)**2+ !109 = n of women stu, men adv
  21*(DV_k21+DV_kXX-DV_k2X-DV_kX1)**2+ !21 = n of men stu, women adv
  73*(DV_k22+DV_kXX-DV_k2X-DV_kX2)**2; !73 = n of men stu, men adv

  ss_tot=a*248; !total sum of squares
  ss_res = ss_tot - ss_stuG - ss_advG - ss_int; !residual sum of squares

!DF for F statistics
new(df_stuG df_advG df_int df_tot df_res);
  df_stuG = 2 - 1; ! ncategories_stuG - 1
  df_advG = 2 - 1; ! ncategories_advG - 1
  df_int = (2 - 1)*(2 - 1);
  df_tot = 248 - 1;
  df_res = df_tot - df_stuG - df_advG - df_int;

!MS = SS / DF
new(ms_stuG ms_advG ms_int ms_res ms_tot);
  ms_stuG = ss_stuG/df_stuG;
  ms_advG = ss_advG/df_advG;
  ms_int = ss_int/df_int;
  ms_tot = ss_tot/df_tot;
  ms_res = ss_res/df_res;

!F statistics
new(W_stuG W_advG W_int);
  W_stuG = ms_stuG / ms_res;
  W_advG = ms_advG / ms_res;
  W_int = ms_int / ms_res;

!partial eta-squared
new(eta_stuG eta_advG eta_int eta_tot);
  eta_stuG = ss_stuG / ss_tot;
  eta_advG = ss_advG / ss_tot;
  eta_int = ss_int / ss_tot;
  eta_tot = eta_stuG + eta_advG + eta_int;

!tests of Null Hypotheses simple effects
new(N_11v21 N_12v22 N_11v12 N_21v22);
  N_11v21 = DV_k11 - DV_k21;
  N_12v22 = DV_k12 - DV_k22;
  N_11v12 = DV_k11 - DV_k12;
  N_21v22 = DV_k21 - DV_k22;

OUTPUT: stdyx;
```

APPENDIX C

ADDITIONAL DESCRIPTIVE STATISTICS

TABLE C1 Simple correlations and descriptive statistics

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1. SOB Y1																		
2. SOB Y2	0.43																	
3. SOB Y3	0.31	0.41																
4. SOB Y4	0.23	0.44	0.49															
5. SE Y1	0.20	0.08	0.07	0.10														
6. SE Y2	0.21	0.17	0.17	0.21	0.60													
7. SE Y3	0.25	0.23	0.20	0.21	0.51	0.63												
8. SE Y4	0.18	0.18	0.20	0.24	0.33	0.53	0.6											
9. SK Y1	−0.03	−0.08	0.00	−0.15	−0.07	0.06	−0.05	−0.12										
10. SK Y2	0.00	−0.11	0.01	−0.04	−0.01	−0.04	−0.07	0.01	0.00									
11. SK Y3	−0.01	−0.15	−0.08	−0.10	−0.05	0.06	0.04	0.08	0.11	0.14								
12. SK Y4	0.18	0.06	0.06	−0.09	0.01	0.01	0.08	0.1	−0.02	−0.03	0.34							
13. AIQ Y2	0.33	0.48	0.29	0.25	0.13	0.2	0.13	0.11	−0.06	−0.06	0.01	0.15						
14. AIQ Y3	0.29	0.29	0.46	0.22	0.17	0.26	0.17	0.14	0.00	−0.09	−0.03	0.11	0.58					
15. AIQ Y4	0.21	0.27	0.31	0.45	0.13	0.20	0.08	0.13	−0.02	−0.03	0.01	0.00	0.44	0.62				
16. Advisor gender	−0.15	−0.14	−0.09	−0.05	0.00	−0.07	−0.07	−0.16	0.03	0.04	−0.03	−0.21	−0.16	−0.23	−0.11			
17. Advisor rank	−0.06	−0.07	−0.02	−0.08	0.02	0.00	0.00	−0.03	−0.09	0.02	−0.08	−0.03	0.02	0.01	−0.04	0.12		
18. Student gender	0.02	−0.03	0.05	0.00	0.16	0.09	0.19	0.07	−0.06	−0.02	0.09	−0.04	−0.02	0.02	0.06	0.06	0.08	

(Continues)

TABLE C1 (Continued)

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
Sample size	264	256	257	240	266	247	262	236	206	217	193	146	259	248	244	263	281	281
Mean	25.27	27.28	26.79	26.73	34.30	35.76	37.58	39.24	1.17	1.50	1.43	1.50	31.90	31.33	30.66	0.73	4.23	0.38
SD	5.66	5.74	5.73	5.96	6.07	6.04	5.90	6.28	0.61	0.66	0.71	0.73	4.00	4.63	5.10	0.45	2.31	0.49

Abbreviations: AIQ, advisor interaction quality; SE, research self-efficacy; SK, research skills; SOB, sense of belonging.