

Time-to-Credit Gender Inequities of First-Year PhD Students in the Biological Sciences

David F. Feldon,^{*,*} James Peugh,[‡] Michelle A. Maher,[§] Josipa Roksa,^{||} and Colby Tofel-Grehl[¶]

[†]Department of Instructional Technology & Learning Sciences and [‡]School of Teacher Education & Leadership, Utah State University, Logan, UT 84322; [‡]Department of Pediatrics, Behavioral Medicine & Clinical Psychology Division, Cincinnati Children's Hospital Medical Center, Cincinnati, OH 45229; [§]School of Education, University of Missouri–Kansas City, Kansas City, MO 64110;

^{||}Department of Sociology and Curry School of Education, University of Virginia, Charlottesville, VA 22904

ABSTRACT

Equitable gender representation is an important aspect of scientific workforce development to secure a sufficient number of individuals and a diversity of perspectives. Biology is the most gender equitable of all scientific fields by the marker of degree attainment, with 52.5% of PhDs awarded to women. However, equitable rates of degree completion do not translate into equitable attainment of faculty or postdoctoral positions, suggesting continued existence of gender inequalities. In a national cohort of 336 first-year PhD students in the biological sciences (i.e., microbiology, cellular biology, molecular biology, developmental biology, and genetics) from 53 research institutions, female participants logged significantly more research hours than males and were significantly more likely than males to attribute their work hours to the demands of their assigned projects over the course of the academic year. Despite this, males were 15% more likely to be listed as authors on published journal articles, indicating inequality in the ratio of time to credit. Given the cumulative advantage that accrues for students who publish early in their graduate careers and the central role that scholarly productivity plays in academic hiring decisions, these findings collectively point to a major potential source of persisting underrepresentation of women on university faculties in these fields.

INTRODUCTION

Training the next generation of scientists is critical to the continued advancement of human knowledge and economic development (U.S. Department of Labor, 2007; Wendler *et al.*, 2010). An important and historically challenging component of growing the scientific workforce is ensuring equitable gender representation to secure a sufficient number of individuals and diversity of perspectives to meet projected workforce demands (National Science Foundation [NSF], 2000). Despite advances made over the past few decades, inequality in wages, promotion, evaluation, and recognition between women and men continues as a general trend in the United States.

These trends are mirrored in many fields, both overtly and subtly (Roos and Gatta, 2009). Scholars have repeatedly documented gender bias against women in academic science across key status markers, including the evaluation of research (Barres, 2006; Budden *et al.*, 2008; Knobloch-Westerwick *et al.*, 2013) and the distribution of scientific awards and honors (Lincoln *et al.*, 2012). Tenured female faculty are often expected to take on more mentorship and service, which is generally uncompensated and undervalued (Hirshfield, 2014). These disparities can in part be attributed to stereotypes and bias that are influential at the individual interaction and organizational

Kimberly Tanner, *Monitoring Editor*

Submitted August 3, 2016; Revised November 14, 2016; Accepted November 15, 2016

CBE Life Sci Educ March 1, 2017 16:ar4

DOI:10.1187/cbe.16-08-0237

*Address correspondence to: David F. Feldon (david.feldon@usu.edu).

© 2017 D. F. Feldon *et al.* CBE—Life Sciences Education © 2017 The American Society for Cell Biology. This article is distributed by The American Society for Cell Biology under license from the author(s). It is available to the public under an Attribution–Noncommercial–Share Alike 3.0 Unported Creative Commons License (<http://creativecommons.org/licenses/by-nc-sa/3.0>).

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

levels (Ridgeway and Correll, 2004; Weyer, 2007). Biases give rise to status beliefs regarding gender, wherein men are often viewed with greater confidence in their choices, abilities, and potential (Wagner and Berger, 1997; Ridgeway, 2001; Foschi, 2009). Even when objective criteria indicate equivalent performance, men are typically judged as being more competent or performing better on various tasks (Foschi, 2000). When men and women work equal hours, men are more readily perceived as being more dedicated to their work and more productive than women, receiving more positive performance evaluations (Heilman, 2001; Reid, 2015).

Hiring trends also reflect gender discrepancies, even in fields with equitable levels of PhD attainment across gender. In the biological sciences, women have accounted for more than 50% of all PhD recipients each year since 2008 (NSF, 2015), but according to current estimates, only 29–36% of tenure-line assistant professorships in the discipline are held by women (Nelson and Brammer, 2007; Sheltzer and Smith, 2014). Sheltzer and Smith suggest that the lower rate of women securing university faculty positions in the biological sciences is attributable to the disproportionate success of men in attaining postdoctoral research positions at top U.S. laboratories—especially those run by male principal investigators. In explaining their findings, the authors speculate that women’s propensity to underrate their own skills (e.g., Correll, 2001; Pallier, 2003; Steinmayr and Spinath, 2009) or male biases to undervalue women’s work contributions (Bowen *et al.*, 2000) may lead to decreased application and hiring rates for postdoctoral positions at elite laboratories. Irrespective of possible explanations, equitable rates of degree completion do not translate into equitable attainment of employment as university faculty or postdoctoral researchers, suggesting the continued existence of gender inequalities.

In the current study, we examine the potential contribution of graduate training experiences to these trends and further inform understanding of potential underlying causes. A number of previous investigations into doctoral education in the sciences also report patterns of gender disparity. Despite the centrality of doctoral mentoring as a key component of scientific training (Paglis *et al.*, 2006; Barnes and Austin, 2008), women report receiving less faculty guidance than their male peers in designing research (Nolan *et al.*, 2007), writing grant proposals (Fox, 2001), and collaborating on publications (Seagram *et al.*, 1998).

A track record of scholarly productivity is essential for securing academic employment (Ehrenberg *et al.*, 2009), and, increasingly, it is expected that graduate students will have a strong track record of publishing before completing their degree programs (Nettles and Millett, 2006). Further, longitudinal studies indicate that the number of publications generated during graduate school significantly predicts subsequent productivity after degree completion (Kademan *et al.*, 2005; Paglis *et al.*, 2006), in keeping with the cumulative advantage of early publication for increased scholarly recognition observed among faculty (i.e., the “Matthew effect” [Merton, 1968, p. 56]). Therefore, female graduate students’ access to publishing opportunities may have direct impact on their future success in multiple phases of the career pipeline.

In this study, we compare the reported hours spent on research activity by participants and the rates of scholarly pro-

ductivity across gender for a national cohort of 336 first-year PhD students in laboratory-based biological research programs (i.e., microbiology, cellular and molecular biology, developmental biology, genetics) from 53 research institutions to assess the extent to which gender inequities may manifest at the earliest stages of research training. Thus, our research questions are as follows:

1. Do men and women report different amounts of time spent on supervised research?
2. Are there differential reported influences associated with research time spent for men and women?
3. Is there a differential publication yield for men and women per time spent on supervised research?

METHODS

In contrast to many previous studies of gender differences in academic science (e.g., Seagram *et al.*, 1998), the current study focuses on a single discipline with a constrained range of research practices (i.e., laboratory-based biological sciences, excluding field-based research) to avoid conflation of trends across distinct disciplinary subpopulations. Further, our analyses use multilevel modeling of individuals nested within institutions to appropriately account for normative cultural practices that may vary by university (i.e., nontrivial intraclass correlations), such as the programmatic use or exclusion of formal lab rotations, and to avoid inflated type I (false-positive) error rates (Musca *et al.*, 2011) that can occur when such nesting is not taken into account (e.g., Kaminski and Geisler, 2012).

Time spent on research tasks was reported biweekly as participants completed the first years of their academic programs. At the conclusion of the academic year, participants reported the number of journal articles, conference papers, and published abstracts for which they received authorship credit during that time. They also completed survey items to provide weighted attributions for the factors affecting time spent on research and levels of confidence they had in their abilities to perform specific research skills (i.e., self-efficacy).

Participant Recruitment and Characteristics

Participants were recruited for the study in two ways. First, program directors and department chairs for the 100 largest biological sciences doctoral programs in the United States were contacted by email to describe the study and request cooperation for informing incoming PhD students about the research project. Specifically, students entering “bench biology” programs, such as microbiology, cellular and molecular biology, genetics, and developmental biology, were targeted. Those who agreed either forwarded recruitment information on behalf of the study or provided students’ email addresses to project personnel for recruitment materials to be disseminated. In instances in which 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. Second, emails describing the student and eligibility criteria were forwarded to several listservs, including those of the American Society for Cell Biology and the CIRTL (Center for the Integration of Research, Teaching, and Learning) Network for broader dissemination.

TABLE 1. Participant demographic distribution^a

	Asian	Black	Hispanic/Latino	Caucasian
Females	53	14	14	117
Males	22	7	12	87
	Prior undergraduate research	Prior graduate research	Prior industry research	No prior research
Females	177	57	43	5
Males	116	29	33	1

^aDistribution of participants by gender, race/ethnicity, and prior research experience. Note: Missing data for gender $N = 4$; missing data for race/ethnicity $N = 6$. Participants with multiple types of prior research experience are counted multiple times.

Those individuals who responded to the recruitment emails or presentations were screened to ensure that they met the criteria for participation (i.e., beginning the first year of a PhD program in microbiology, cellular biology, molecular biology, developmental biology, or genetics in Fall 2014) and fully understood the expected scope of participation over the course of the funded project (4 years with possible renewal). It was further explained that all data collected would remain confidential, that all data would be scored blindly, and that no information disseminated regarding the study would individually identify them in any way. Participants signed consent forms per the requirements specified by the institutional review board for human subjects research. Participants who remained active in the study received a \$400 annual incentive, paid in semiannual increments.

Participants were informed that if they failed to provide two or more consecutive annual data items (i.e., annual surveys) or more than 50% of the biweekly surveys in a single academic year, they would be withdrawn from the study. In addition, any participants who took a leave of absence from their academic program greater than one semester would be withdrawn. All data points were checked and followed up by research assistants for timely completion and meaningful responses. Three participants were withdrawn during the time these data were collected (two due to low response rate; one due to taking leave from the degree program). Three participants left the study when they withdrew from their academic programs.

Overall sample size was $N = 336$ participants sampled from $C = 53$ institutions, with an average of 6.34 (336/53; $SD = 5.69$) participants per institution. Participant characteristics (i.e., distribution by gender, race/ethnicity, and prior research experience) are presented in Table 1. A large majority of participants (84.2%) reported rotating through multiple laboratories as part of their first years of doctoral training. The distribution of participants by specific program area (cell biology, developmental biology, etc.) can be found in Supplemental Table S1. Although not pertinent to the current analyses, participants also provided additional data on hours spent fulfilling teaching responsibilities, presented in Supplemental Table S2. The distribution of participants within institution by gender is presented in Table 2. The distribution of institutions across Carnegie research classifications is available in Supplemental Table S3.

Data Collection

Upon submitting informed consent paperwork, participants completed biweekly online surveys that focused on information specific to the preceding 2-week period. They also received additional surveys that were completed once per year. These

instruments are described under the following headings: biweekly surveys, annual survey 1, annual survey 2, and annual survey 3.

To address the first research question (reported time differing by gender), we drew data from the biweekly surveys and annual survey 1 (i.e., research self-efficacy). To address the second research question (gender-differential influences on time spent), we drew data from annual survey 2. To address the third research question (gender-differential publication yield), we drew data from annual survey 3.

Biweekly Surveys. Biweekly surveys asked participants to report the number of hours spent teaching, engaging in supervised research, and writing for publication in collaboration with a faculty member or other senior researcher during the preceding 2-week period. Specifically, participants were provided with the prompt “Over the last two weeks, approximately how many hours have you spent engaged in supervised research activities (e.g., working in a lab)?” and a drop-down menu with integers from 0 to 150.

Although some methodological research on the collection of time data from work contexts indicates that time diaries are a more precise measure than surveys (Robinson and Bostrom, 1994), the level of intrusion into participants’ daily work processes rendered that approach impractical for the current study. The same research also raised tentative concerns that women’s responses might reflect an upward bias in reported hours relative to men, as measured by the discrepancy between survey-based and diary-based work hours. However, subsequent studies do not find discrepancies to be associated with gender or any other demographic variable (Jacobs, 1998).

Further, a critical reading of Robinson and Bostrom’s (1994) study raises questions about the conclusions and their applicability to the current sample. First, their data were drawn from workforce studies conducted in 1965, 1975, and 1985, and the authors note that data from each subsequent decade reflected increasing reporting discrepancies across all participants (i.e., male and female) due to increases in general cognitive “busyness” in work environments. Given that the proportion of women in the workforce increased substantially from 1965 (35% of the U.S. workforce) to 1985 (44% of the U.S. workforce) (U.S. Bureau of Labor Statistics, 2016), the reported bias likely reflects the collinearity of the increasing relative proportion of women in the workforce sample and the increasing hours bias across genders over time. That is, without additional data (e.g., comparisons of time discrepancy between genders within time periods), there is no way to determine that the increase in observed discrepancy between sources of reported

TABLE 2. Participant gender by institution^a

University	Male	Female	Total
1	0	1	1
2	1	0	1
3	1	3	4
4	1	2	3
5	0	1	1
6	2	0	2
7	0	2	2
8	5	4	9
9	13	9	24
10	2	2	4
11	1	1	2
12	0	6	6
13	8	10	19
14	0	1	1
15	5	2	7
16	3	1	4
17	2	5	7
18	5	4	9
19	2	5	8
20	3	0	3
21	4	7	12
22	4	13	17
23	7	9	16
24	0	7	8
25	2	3	5
26	2	3	5
27	2	1	3
28	4	3	7
29	2	2	4
30	1	2	3
31	0	2	2
32	2	1	3
33	6	6	13
34	0	1	1
35	0	4	4
36	3	1	4
37	4	8	12
38	3	2	5
39	0	1	2
40	1	0	1
41	3	8	11
42	1	6	7
43	6	7	13
44	0	1	1
45	0	2	2
46	1	0	1
47	0	1	1
48	7	16	23
49	1	6	7
50	3	5	8
51	2	0	2
52	0	4	4
53	3	5	8
Total	128	196	336

^aDistribution of participants within institution by gender. Note: Missing data for gender $N = 4$.

time is due to gender rather than uniform increases in cognitive busyness across genders, accompanied by coincidental but unrelated increases in the proportion of women in the workforce over time. Because Robinson and Bostrom theorize that the increase of busyness in the workplace accounts for increasing bias over time and do not have an articulated theoretical position that could establish a causal relationship between respondent gender and bias, we conclude that the noted correspondence between gender and bias is an artifact of the approach taken to the statistical analysis of their data rather than a durable trend that would skew the data collected for the current study.

Additional limitations on the applicability of Robinson and Bostrom's (1994) work to the current study include several aspects of the sample characteristics. First, the Robinson and Bostrom data were drawn from all sectors of the workforce, which differs from the graduate school environment substantially in terms of the population age, level of education, number of hours, and work setting of university research laboratories. Additionally, respondents reporting 30 hours of work or less per week had negligible discrepancies between diary and survey methods of data collection. Given that unadjusted mean weekly times reported by most students in our sample were ~20 hours (male = 20.99, SD = 9.90; female = 19.50, SD = 10.35), upward bias is even less likely on the basis of the 1994 analysis.

Annual Survey 1. During the Spring semester of 2015, participants received the Research Experience Self-Rating Survey (Kardash, 2000), which asked them to self-rate their abilities to perform each of 10 research-related tasks ("To what extent do you feel you can...?") on a Likert scale of 1–5 ("not at all," "less capable," "capable," "more capable," "a great deal"): "Understand contemporary concepts in your field," "Make use of the primary science literature in your field (e.g., journal articles)," "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 the hypothesis," "Understand the importance of controls' in research," "Observe and collect data," "Statistically analyze data," "Interpret data by relating results to the original hypothesis," and "Reformulate your original research hypothesis (as appropriate)."

Additional items included in this survey asked participants to report the number of months spent participating in research activities before entering their PhD programs. Specific categories included formal research in high school, undergraduate research, research during a previous graduate degree program, and research conducted in industry.

Annual Survey 2. Participants also received a survey asking them to respond to the prompt "What kinds of things affect your time spent on research on a weekly basis? Please categorize by percentage." Ten possible responses were provided, and the assigned percentages were required to sum to 100%. The response options were "Required hours," "Changes in workload based on project demands," "Comfort in lab," "Personal judgment/discretion," "Opportunity to contribute more to the research effort," "I'm not a good fit," "I'm not taken seriously," "Course work," "Familial responsibilities," and "Non-research obligations."

TABLE 3. Effect sizes of gender hour differences by time point

Discrete time points	Males		Females		Intercept difference	Difference			
	Intercept	SD	Intercept	SD		95% CI	SE	<i>p</i> Value ^a	Cohen's <i>d</i> ^b
T9	22.74	25.25	53.07	28.25	30.33	(−8.07, 69.87)	20.27	0.07	1.18
T10	20.06	26.64	44.47	27.58	24.42	(−18.12, 68.17)	22.30	0.14	0.90
T11	21.95	25.72	52.53	28.48	30.58	(−8.80, 70.19)	20.47	0.07	1.12
T12	8.54	25.50	55.01	28.09	46.47	(10.59, 84.91)	18.69	0.01	1.72
T13	22.12	28.23	50.70	28.85	28.59	(−10.46, 69.83)	20.32	0.08	1.00
T14	24.20	27.12	43.62	29.14	19.41	(−17.30, 58.48)	18.76	0.15	0.68
T15	28.84	26.08	47.03	30.15	18.18	(−18.22, 55.34)	18.56	0.17	0.63

^aOne-tailed *p* values reported in accordance with hypothesis tested.

^bCohen's *d* effect sizes comparing males (*N* = 128) and females (*N* = 198) at each time point.

Annual Survey 3. At the conclusion of the Spring semester, participants received another survey that asked them to identify any journal articles, conference papers, or published abstracts for which they had received authorship credit during the academic year. Responses were validated through independent researcher verification of citation information provided in the surveys against conference proceedings and journal tables of contents. Respondents were contacted regarding any observed discrepancies, and finalized information was subsequently used for analysis.

Data Analysis

Data analyses are reported in the following three sections.

Analysis of Time Spent on Research (RQ1). The first goal of a longitudinal analysis is to quantify how the response variable changes over time, and polynomial trend components are the best way to capture when and how response variable changes occur (Raudenbush and Bryk, 2002; Singer and Willett, 2003). Hierarchical linear modeling (HLM) for longitudinal data was performed in two steps. In the first step, the most parsimonious longitudinal polynomial trend (linear, quadratic, cubic, etc.) that best modeled average changes in the response variable across participants during the study was selected. This was accomplished by adding a lower-order trend component to the model (e.g., adding a fixed linear slope to the model), followed immediately by a test of whether that trend component showed significant variation across participants (i.e., adding a linear slope random effect to the analysis model). The next polynomial trend component was then added to the model and tested in similar manner.

Before the examination of changes in time spent on supervised research activity over time, corrective measures (i.e., “Type = Complex” in Mplus) were taken to guard against type I inferential errors that could result from ignoring the nesting of participants within universities, and missing data were handled via the default (Maximum Likelihood Regression [MLR]) parameter estimation algorithm in Mplus (version 7.4). Specifically, the Mplus command determines the proportion of variance in the response variable that can be attributed to institutions due to clustering of individual participants within universities and applies a multiplier to inflate estimated SEs to prevent erroneous statistical significance attributable to the influences of clustering rather than the targeted independent variables. However, the notable

variation observed in hours spent on research over time coupled with the computational SE increases generated by “Type = Complex,” could have artificially inflated *p* values, resulting in type II inferential errors. Specifically, large observed variance statistics for the number of hours spent on supervised research activities, combined with average sample sizes and design effect–corrected SEs will result in wide 95% confidence intervals (CIs). Therefore, meaningful significance for interpretational purposes was made based on effect sizes (see Table 3).

Three guidelines were observed during this process. First, if a trend component fixed effect was nonsignificant, but the random effect for that trend component was significant, both were retained in the analysis model. Second, if a trend component fixed effect was significant, but the random effect for that trend component was not significant, only the fixed effect was retained in the analysis model. Third, this process continued until both the fixed and random effects of a given trend component were nonsignificant.

Following this strategy, polynomial functions of time (i.e., linear time, quadratic time [time²], cubic time [time³], etc.) were added to the level 1 linear analysis model as fixed effects (i.e., γ) to best capture and model average change in hours spent on research across participants over time. This process continued for a possible ($T = 13 - 1 = 12$) 12 fixed effects and ($T = 13 - 2 = 11$) 11 random effects possibly needed to adequately model changes in hours spent on research over time. Missing data for both males and females ranged between 1.2 and 17.4% across the 13 time points and were handled via the default longitudinal HLM parameter estimation algorithm (MLR).

Participants, on average, completed 12.44 biweekly time allocation surveys out of a possible 13 used for analysis. Participants as individuals accounted for 31.2% of variance in reported time spent on supervised research activities, and universities within which participants were nested accounted for 22.6% of variance. Data from biweekly periods 7 and 8 were excluded from analysis, because they coincided with the winter holidays, which introduced confounds related to the physical accessibility of university facilities, the personal preferences of supervising faculty, and atypical family obligations.

Next, a multivariate analysis of variance (MANOVA) was conducted to test for gender differences in the Kardash (2000) survey items (annual survey 2) assessing self-efficacy for

specific research skills described earlier. Missing data ranged from 1.5 to 16% across analysis variables and were handled via the default maximum-likelihood parameter estimation algorithm. In the analysis, 1000 bootstrap samples were requested to generate empirical rather than observed SEs.

Based on the outcomes of the MANOVA, two factors (i.e., self-efficacy scores for “Formulate a research hypothesis based on a specific question” and “Design an experiment or theoretical test of the hypothesis”) were added to the polynomial models for males and females as predictors of all significant trajectory components, and indicators for gender, ethnicity, and previous research experience were added to the model as control covariates (i.e., specifying the level-2 model). The final linear model used for both males and females is presented in the Supplemental Material.

Analysis of Reported Influences on Research Time (RQ2).

The next set of analyses examined potential explanatory factors that could account for observed differences in time spent on research by gender. Survey items asking participants to indicate perceived influences on the amount of time they spent in supervised research as percentages were analyzed using a MANOVA approach in Mplus that controlled for nesting of participants within institutions (i.e., “Type = Complex”). Because participants needed to make their cumulative responses sum to 100%, individual items were not independent, but the multivariate structure of the analysis permitted items to intercorrelate freely.

Analysis of Likelihood of Authorship on Scholarly Publications (RQ3).

The final analysis examined gender differentials in authorship during the first year of graduate study. Missing data were observed in four of the data analysis variables: the categorical self-efficacy in designing experiments and formulating research hypotheses both had 4.5% missing data, and the binary indicators for published articles and published abstracts showed 9.8 and 10.1% missing data, respectively. Missing data were handled via multiple imputation for categorical variables in Mplus (version 7.4) and $M = 100$ imputed data sets were used for all analyses. Before analyses, the variable “total hours spent on research” from T9 to T15 was both rescaled (i.e., a 1-unit increase reflected an additional 100 hours spent on research) and grand-mean centered to facilitate interpretation. Further, specific analysis commands in Mplus (i.e., “Type = Complex”) were used so that the nesting of participants within universities could be ignored without fear of type I inferential errors. Main effects for gender, total hours spent on research, and designing experiments and formulating research hypotheses self-efficacy scores as main effects, gender by total hours spent on research, gender by self-efficacy “designing experiments,” and gender by self-efficacy “formulating research hypotheses” interactions, were all entered into the model as predictor variables (independent variables). Finally, a multivariate binary logistic regression analysis was conducted in which both of the binary indicator response variables for article publication and abstract publication (dependent variables) were entered into the model and allowed to correlate, because it was possible that a participant could have published both an abstract and an article, making both correlated rather than independent.

RESULTS

Reported Time Spent on Supervised Research Differs by Gender

The data on time that doctoral students invested in research activities within laboratories were captured by having participants complete biweekly online surveys in which they were asked to report the number of hours spent teaching, engaging in supervised research, and writing for publication in collaboration with a faculty member or other senior researcher during the preceding 2-week period. In separate annual surveys, participants also reported the amount of their prior research experiences and their levels of confidence in performing each of 10 criterial research tasks (Kardash, 2000). Gender differences in response patterns were evident only for confidence in designing experiments and formulating research hypotheses, respectively, with men reporting significantly greater levels of confidence than women.

Accordingly, these values were entered into the level 2 (individual) model equations describing the relationship between gender and time spent on supervised research, described earlier, as predictors of significant intercept, linear slope, quadratic change, and cubic change variance while controlling for ethnicity and previous research experience. Across 13 time points, changes in time spent were modeled independently for men and women. All polynomial fixed-effects (level 1) coefficients were significant for women, as was variation on all but the quartic change term around each of the growth trajectory fixed effects ($p < 0.05$). In contrast, the model of men’s hours included a significant fixed-effects coefficient only for intercept ($p < 0.001$) in the polynomial model. However, these nonsignificant fixed effects were retained in the model due to significant intercept ($\tau_{00} = 840.86$; $p < 0.01$) a trend toward significant linear slope ($\tau_{11} = 79.22$; $p < 0.06$), significant quadratic change ($\tau_{22} = 1.44$; $p < 0.01$), and significant cubic change ($\tau_{33} = 0.25$; $p < 0.01$) for random effects that indicated significant variation around each of the growth trajectory fixed effects. Results for the level 2 model for females showed no significant effects. Results for males, however, showed that self-efficacy for designing experiments significantly predicted linear slope ($\tau_{17} = -4.24$; $p < 0.05$), quadratic change ($\tau_{27} = 0.58$; $p < 0.05$), and cubic change ($\tau_{37} = -0.22$; $p < 0.05$) variances. In short, trajectories of male and female time spent on research differed to the extent that different polynomial models were necessary to describe them at the group level.

To better illustrate these findings, Figure 1 shows the models of males and females, respectively. Estimated effect sizes by time point for T9–T15 ranged between Cohen’s $d = 0.63$ and $d = 1.72$, representing consistently large effects (Cohen, 1988). Interactions with self-efficacy are reflected in Figure 1 as separate trend lines for participants more than 1 SD above mean self-efficacy for experimental design, more than 1 SD below, and within 1 SD of the mean. Supplemental Figure S1 shows each model separately and includes SE estimates around each time point. Tabular representations of level 1 and level 2 models for males and females, respectively, are presented in Supplemental Tables S4–S7.

Reported Influences on Research Time Differ by Gender

To gain insight into the factors that participants perceived to influence the amount of time they spent on research, we

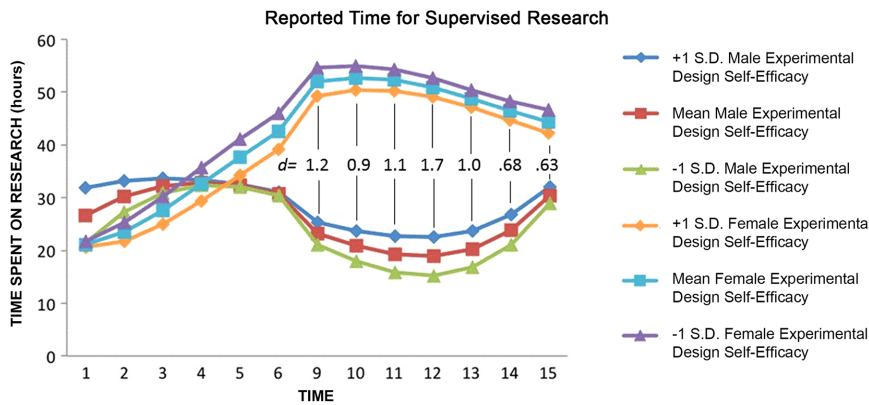


FIGURE 1. Differences in time spent on research by gender. Differences in male and female time spent on research per biweekly period, controlling for variance at the level of the institution. Lines represent males' and females' levels of confidence in designing experiments, as indicated. Interactions with this variable were significant only for males. Estimated effect sizes between males and females by time point for T9–T15 ranged between Cohen's $d = 0.63$ and $d = 1.7$.

administered a survey during the Spring semester, asking participants to respond to the prompt “What kinds of things affect your time spent on research on a weekly basis? Please categorize by percentage.” Ten possible responses were provided, and the assigned percentages were required to sum to 100%. A MANOVA detected a significant difference between male and female responses on only one response item: women showed a significantly higher score (mean = 27.85) than men (mean = 21.49) for the response option “demands required for the task determining the amount of time spent on research” ($\Delta\bar{X} = 6.36$, $p < 0.001$; $d = 0.28$), representing a small but significant effect (see Figure 2).

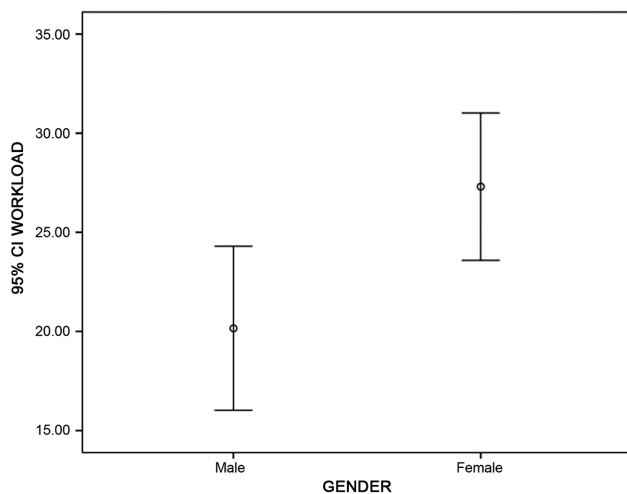


FIGURE 2. Difference of estimated influence of task demands on research time by gender. Participants provided survey responses in which they weighted ten items reflecting influences on the amount of time they spent conducting research, summing to 100%. MANOVA computations indicated that males and females differed only in the reported influence of changes in workload based on project demands, with females reporting significantly greater influence (mean difference = 6.36, $p = 0.003$, Cohen's $d = 0.27$).

Men More Likely to Receive Authorship Credit per Hours Worked

At the conclusion of the academic year, participants received another survey that asked them to identify any journal articles or published abstracts for which they had received authorship credit during the academic year. Of 303 responding participants, 68 reported authorship on a published journal article (22.4%), and 40 reported authorship on a published abstract (13.2%). Logistic regression analysis evaluated the likelihood of authorship (dependent variable) by gender, total hours spent on research, self-efficacy for experimental design and framing hypotheses respectively, and gender interactions with research hours and the two self-efficacy variables (independent variables). No significant results were observed for predicting abstract publication. However, a significant gender by total hours spent on

research interaction effect ($b = 0.144$; $p < 0.05$; 95% CI: [0.027, 0.262]) was found for journal articles, indicating that for every 100 h spent on research and compared with females, males were 15% more likely (odds ratio = $\exp[0.144] = 1.15$) to receive authorship credit for a journal article (see Figure 3). Neither the inclusion of variables in the logistic regression model reflecting confidence in designing experiments and formulating research hypotheses nor their interactions with gender in the logistic regression model yielded significant

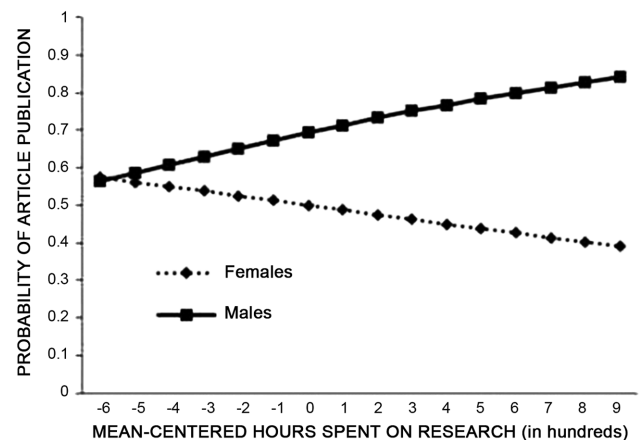


FIGURE 3. Logistic regression of publication on research time by gender. Participants provided survey responses in which they indicated having received authorship credit on journal articles, conference papers, and/or published abstracts. Logistic regression analyses for authorship on each type of publication, respectively, included gender, research time spent in the second semester, self-efficacy for experimental design and hypothesis framing skills, and gender interactions with each as predictors. The only significant predictor of journal article authorship was the gender by research time interaction ($b = 0.144$; $p = 0.016$; $\exp[0.144] = 1.15$), indicating that males were 15% more likely to receive authorship credit than females per 100 hours of reported research time.

coefficients ($p > 0.05$), indicating that participant confidence in research skills did not influence the likelihood of authorship for either men or women.

DISCUSSION

Our results show that, after controlling for variance at the institutional level, men spend significantly less time engaging in supervised research, are less likely to attribute their time allocation to the demands of assigned tasks, and are 15% more likely to author published journal articles than their female counterparts per 100 hours of research time. Collectively, these findings suggest that gender inequality manifests in the form of differential time-to-credit payoff as early as the first year of doctoral training. The men in our sample were better able to procure or were provided with better opportunities to capitalize on publishing prospects as a function of time spent on research than their female counterparts despite the reverse trend for time spent on research. These results provide convergent evidence for the conclusions of Smith *et al.* (2013), who found that female graduate students perceive a greater investment of effort to be necessary for success in their academic programs compared with their male counterparts. Although perceived effort and time invested are not identical constructs, it is possible that experiences of discrepant time-to-publication ratios may contribute to such beliefs.

The finding of significance for journal articles is notable, because these publications are typically the most highly valued as indicators of scholarly productivity for professional evaluation in academe (McGrail *et al.*, 2006; Ehrenberg *et al.*, 2009). Given the importance of scholarly productivity in the evaluation of candidates for academic positions and the cumulative advantage that early publications provide over time (Merton, 1968, 1988; Kademani *et al.*, 2005), these findings may account—at least in part—for the inequitable hiring rates for postdoctoral research positions reported in prior research (Sheltzer and Smith, 2014). Such cumulative advantage has been documented with graduate student populations across STEM (science, technology, engineering, mathematics) disciplines, in which both skills (Feldon *et al.*, 2016) and faculty recognition of students' ability (Gopaul, 2016) increase geometrically from small initial advantages.

The failure of confidence in research skills (i.e., self-efficacy) to explain any significant variance in either the amount of time spent on research by women or the likelihood of publishing by women or men is also of interest. These patterns in the first year of doctoral study indicate that, in contrast to suggestions in previous studies (e.g., Correll, 2001; Sheltzer and Smith, 2014), there is no evidence that lower self-efficacy prompts women to self-select out of professional opportunities in the first years of their doctoral studies. While it is possible that this pattern changes over the course of PhD attainment, caution should clearly be used in applying this explanation to underrepresentation of women in professional academic science.

In contrast, our finding that confidence in research skills affected only men's time investment in research has two possible implications. First, the relevance of confidence in experimental design skills to time spent on research may point to a greater relevance of those skills in the tasks assigned to male graduate students within the laboratory environment. If men are more likely than women to engage in methodological

decision-making tasks, it could explain the observed difference in publication rates. It would also better position men to discuss their contributions to laboratory research when applying for postdoctoral positions, increasing their competitiveness for those positions, above and beyond possible differential rates of publication. Second, the significant gender difference on this specific aspect of research and the lack of observed differences on confidence related to other aspects suggest that the ability to engage successfully in laboratory experimental design efforts may be differentially important in the training of graduate students for the purposes of setting career trajectories. Future research may inform the extent to which the nature of assigned research tasks differ and expand the scope of the current findings.

With peer-reviewed publications serving as the proverbial “coin of the realm” (Wilcox, 1998, p. 216) for assessing research prowess, the ability of early-career researchers to convert time spent into publications leads to an increased likelihood of career success (Merton, 1968, 1988; Kademani *et al.*, 2005). Because the results of this study reflect gender inequality with long-term ramifications in a scientific field that awards more doctorates to women than men, attention to degree completion rates reflects a necessary, but not sufficient, metric by which to evaluate gender equity in graduate training for the biological sciences. To best improve the equitable access of men and women to professional academic success, understanding the ways in which research training tasks differently enculturate men and women is essential. Increasing professional awareness of gender disparities is an important first step toward eliminating the effects of gender bias in the field. It may be further valuable for faculty who supervise graduate students to increase vigilance in their management of publications coming from their laboratories to ensure that both opportunities for authorship and recognition of invested effort toward publishable findings are allocated appropriately and equitably.

ACKNOWLEDGMENTS

The work reported in this article was supported by a grant from the NSF (NSF-1431234). The views in this paper are those of the authors and do not necessarily represent the views of the supporting funding agency.

REFERENCES

- Barnes BJ, Austin AE (2008). The role of doctoral advisors: a look at advising from the advisor's perspective. *Innov High Educ* 33, 297–315.
- Barres B (2006). Does gender matter? *Nature* 442, 133–136.
- Bowen C, Swim JK, Jacobs RR (2000). Evaluating gender biases on actual job performance of real people: a meta-analysis. *J Applied Soc Psych* 30, 2194–2215.
- Budden A, Tregenza T, Aarssen L, Koricheva J, Leimu R, Lortie C (2008). Double-blind review favours increased representation of female authors. *Trends Ecol Evol* 23, 4–6.
- Cohen J (1988). *Statistical Power Analysis for the Behavioral Sciences*, 2nd ed., Hillsdale, NJ: Erlbaum.
- Correll S (2001). Gender and the career choice process: the role of biased self-assessments. *Am J Soc* 106, 1691–1730.
- Ehrenberg R, Zuckerman H, Groen J, Brucker S (2009). Changing the education of scholars. In: *Doctoral Education and the Faculty of the Future*, 1st ed., ed. RG Ehrenberg and CV Kuh, Ithaca, NY: Cornell University Press, 15–34.

- Feldon DF, Maher M, Roksa J, Peugh J (2016). Cumulative advantage in the skill development of STEM graduate students: a mixed methods study. *Am Educ Res J* 53, 132–161.
- Foschi M (2000). Double standards for competence: theory and research. *Annu Rev of Soc* 26, 21–42.
- Foschi M (2009). Gender, performance level, and competence standards in task groups. *Soc Sci Res* 38, 447–457.
- Fox MF (2001). Women, science, and academia: graduate education and careers. *Gender Soc* 15, 654–666.
- Gopaul B (2016). Applying cultural capital and field to doctoral student socialization. *Intl J Res Dev* 7, 46–62.
- Heilman ME (2001). Description and prescription: how gender stereotypes prevent women's ascent up the organizational ladder. *J Soc Iss* 57, 657–674.
- Hirshfield LE (2014). "She's not good with crying": the effect of gender expectations on graduate students' assessments of their principal investigators. *Gender Educ* 26, 601–617.
- Jacobs JA (1998). Measuring time at work: are self-reports accurate? *Mon Labor Rev* 121, 42–53.
- Kademani BS, Kalyane VL, Kumar V, Mohan L (2005). Nobel laureates: their publication productivity, collaboration and authorship status. *Scientometrics* 62, 261–268.
- Kaminski D, Geisler C (2012). Survival analysis of faculty retention in science and engineering by gender. *Science* 335, 864–866.
- Kardash CM (2000). Evaluation of undergraduate research experience: perceptions of undergraduate interns and their faculty mentors. *J Educ Psychol* 92, 191–201.
- Knobloch-Westerwick S, Glynn CJ, Hugué M (2013). The Matilda effect in science communication: an experiment on gender bias in publication quality perceptions and collaboration interest. *Sci Comm* 35, 603–625.
- Lincoln AE, Pincus S, Koster JB, Leboy PS (2012). The Matilda effect in science: awards and prizes in the US, 1990s and 2000s. *Soc Stud Sci* 42, 307–320.
- Mcgrail MR, Rickard CM, Jones R (2006). Publish or perish: a systematic review of interventions to increase academic publication rates. *High Educ Res Dev* 25, 19–35.
- Merton RK (1968). The Matthew effect in science: the reward and communication systems of science are considered. *Science* 159, 56–63.
- Merton RK (1988). The Matthew effect in science, II: cumulative advantage and the symbolism of intellectual property. *Isis* 79, 606–623.
- Musca SC, Kamiejski R, Nugier A, Méot A, Er-Rafiy A, Brauer M (2011). Data with hierarchical structure: impact of intraclass correlation and sample size on type-I error. *Front Psychol* 2, 74.
- National Science Foundation (NSF) (2000). Land of Plenty: Diversity as America's Competitive Edge in Science, Engineering, and Technology, Report of the Congressional Commission on the Advancement of Women and Minorities in Science, Engineering, and Technology Development, Washington, DC.
- NSF (2015). Science and Engineering Degrees: 1966–2012. www.nsf.gov/statistics/2015/nsf15326/ (accessed 1 October 2016).
- Nelson DJ, Brammer CN (2007). A National Analysis of Minorities in Science and Engineering Faculties at Research Universities, Norman: University of Oklahoma.
- Nettles MT, Millett CM (2006). Three Magic Letters: Getting to Ph.D. Baltimore, MD: Johns Hopkins University Press.
- Nolan SA, Buckner JP, Marzabadi CH, Kuck VJ (2007). Training and mentoring of chemists: a study of gender disparity. *Sex Roles* 58, 235–250.
- Paglis LL, Green SG, Bauer TN (2006). Does adviser mentoring add value? A longitudinal study of mentoring and doctoral student outcomes. *Res High Educ* 47, 451–476.
- Pallier G (2003). Gender differences in the self-assessment of accuracy on cognitive tasks. *Sex Roles* 48, 265–276.
- Raudenbush SW, Bryk AS (2002). Hierarchical Linear Models: Applications and Data Analysis Methods, 2nd ed., Thousand Oaks, CA: Sage.
- Reid E (2015). Embracing, passing, revealing, and the ideal worker image: how people navigate expected and experienced professional identities. *Org Sci* 26, 997–1017.
- Ridgeway CL (2001). Gender, status, and leadership. *J Soc Iss* 57, 637–655.
- Ridgeway CL, Correll SJ (2004). Unpacking the gender system: a theoretical perspective on gender beliefs and social relations. *Gender Soc* 18, 510–531.
- Robinson JP, Bostrum A (1994). The overestimated workweek? What time diary measures suggest. *Mon Labor Rev* 117(8), 11–23.
- Roos PA, Gatta ML (2009). Gender (in)equity in the academy: subtle mechanisms and the production of inequality. *Res Soc Stratif Mobil* 27, 177–200.
- Seagram BC, Gould J, Pyke SW (1998). An investigation of gender and other variables on time to completion of doctoral degrees. *Res High Educ* 39, 319–335.
- Sheltzer JM, Smith JC (2014). Elite male faculty in the life sciences employ fewer women. *Proc Natl Acad Sci USA* 111, 10107–10112.
- Singer JD, Willett JB (2003). Applied Longitudinal Data Analysis: Modeling Change and Event Occurrence, New York: Oxford University Press.
- Smith J, Lewis K, Hawthorne L, Hodges S (2013). When trying hard isn't natural: women's belonging with and motivation for male-dominated STEM fields as a function of effort expenditure concerns. *Pers Soc Psych Bull* 39, 131–143.
- Steinmayr R, Spinath B (2009). What explains boys' stronger confidence in their intelligence? *Sex Roles* 61, 736–749.
- U.S. Bureau of Labor Statistics (2016). Facts over time. www.dol.gov/wb/stats/facts_over_time.htm#wilf (accessed 22 September 2016).
- U.S. Department of Labor (2007). The STEM Workforce Challenge: The Role of the Public Workforce System in a National Solution for a Competitive Science, Technology, Engineering, and Mathematics (STEM) Workforce, Washington, DC.
- Wagner DG, Berger J (1997). Gender and interpersonal task behaviors: status expectation accounts. *Soc Perspect* 40, 1–32.
- Wendler C, Bridgeman B, Cline F, Millett C, Rock J, Bell N, McAllister P (2010). The Path Forward: The Future of Graduate Education in the United States, Princeton, NJ: Educational Testing Service.
- Weyer B (2007). Twenty years later: explaining the persistence of the glass ceiling for women leaders. *Women Manage Rev* 22, 482–496.
- Wilcox LJ (1998). The coin of the realm, the source of complaints. *J Am Med Assoc* 280, 216–217.