The Effects of Doctoral Teaching Development on Early Career STEM Scholars' College-teaching Self-efficacy

WCER Working Paper No. 2015-1
March 2015

Mark R. Connolly, You-Geon Lee, and Julia Nelson Savoy
Wisconsin Center for Education Research
University of Wisconsin–Madison
mark.connolly@wisc.edu

WCER
Wisconsin Center for Education Research
School of Education • University of Wisconsin–Madison • http://www.wcer.wisc.edu/

Connolly, M.R.; Lee, Y.-G.; & Savoy, J. N. (2015). *The Effects of Doctoral Teaching Development on Early career STEM Scholars' College-teaching Self-efficacy* (WCER Working Paper No. 2015-1). Retrieved from University of Wisconsin–Madison, Wisconsin Center for Education Research website: http://www.wcer.wisc.edu/publications/workingPapers/papers.php

The Effects of Doctoral Teaching Development on Early Career STEM Scholars' College Teaching Self-efficacy

Mark R. Connolly, You-Geon Lee, Julia Nelson Savoy

With improving undergraduate STEM (science, technology, engineering, and mathematics) education now a national policy priority, retaining and graduating more students with STEM training requires improving undergraduate teaching in STEM fields. Seymour and Hewitt's (1997) landmark study of why students leave STEM majors found not only that poor teaching was a primary reason, but that poor teaching was the greatest concern for students who remained in STEM majors. Policymakers and funders are paying greater attention to the quality of undergraduate STEM instruction, especially to which teaching practices are proven effective by research and practice, and to how faculty learn to adopt promising instructional practices that engage more students (Kober, 2015; National Research Council, 2012). Not surprisingly, when the President's Council of Advisors on Science and Technology (2012) called for 1 million additional STEM graduates over the next decade, its top policy recommendation was "to train current and future faculty in evidence-based teaching practices" (p. iv). Similarly, when the Association of American Universities launched a 5-year project to reform undergraduate STEM education, four of its five policy goals focused on improving teaching in STEM courses (Association of American Universities, 2014). A 2014 sourcebook calling for systemic change in STEM higher education included "supporting faculty development" as one of seven overarching goals (Coalition for Reform of Undergraduate STEM Education, 2014, p. 5). The signs are clear: Preparing faculty more effectively as undergraduate educators is a priority in the national STEM agenda.

Although efforts to improve postsecondary instruction generally have focused on current faculty, more attention has been given to how *future faculty*[1] are being prepared to assume academic roles and responsibilities (e.g., Austin, 2010; Hershock, Groscurth, & Milkova, 2011; Hopwood & Stocks, 2008; Kalish et al., 2011; Palmer, 2011; Schönwetter & Ellis, 2010; Wulff & Austin, 2004; Wurgler, VanHeuvelen, Rohrman, Loehr, & Grace, 2013). Doctoral training in STEM fields has traditionally consisted of a doctoral student working closely with a faculty advisor to learn the research methods and content knowledge that are constitutive of the discipline (Anderson et al., 2011; Walker, Golde, Jones, Bueschel, & Hutchings, 2008). This apprenticeship model tends to emphasize doctoral students' formation as researchers and scholars and, as such, does not focus as intentionally on preparing academic aspirants to handle the full range of roles and responsibilities of 21st century academics, including teaching (Austin & McDaniels, 2006; Feldon et al., 2011). As a result, doctoral students often report feeling quite confident about their research skills but far less prepared for teaching and advising responsibilities (Golde & Dore, 2001). Because approximately one of every three STEM Ph.D.s is involved in some kind of college teaching within six years of completing a doctorate

[1] By "faculty," we mean people who hold teaching positions that are full- or part-time, and tenured or untenured, at postsecondary institutions.

(Connolly, 2012), professional development in teaching is needed for doctoral students, many of whom will teach, train, and mentor the next generation of STEM undergraduates.

Future Faculty Programs for STEM Doctoral Students

Since the early 1990s, professional development programs for future faculty have emerged to help doctoral students understand faculty roles and responsibilities. These future faculty programs range in scope from modest departmental programs to national, multi-institutional projects (Austin, Connolly, & Colbeck, 2008; DeNeef, 2002; Handelsman, Miller, & Pfund, 2006; Weisbuch, 2004). Although future faculty programs may address topics such as fostering diversity in the sciences, ethical conduct of research, and academic job hunting, a major component of most programs is *teaching development* (TD)—that is, helping doctoral students gain the knowledge, skills, and values needed to effectively teach undergraduates. How TD is provided to doctoral students varies, but institutions typically provide programs through some combination of their academic units (departments and colleges), graduate school, and center for teaching and learning. TD offerings also vary in format and duration, ranging from low-engagement events such as brown-bag discussions and one-off workshops to more-intensive semester-length pedagogy courses and certificate programs in college teaching (Connolly, Savoy, & Barger, 2010). In general, these programs seek three reform-oriented outcomes: (1) to improve the quality of undergraduate education by enhancing participants' pedagogical skills; (2) to provide training that better reflects the full range of faculty responsibilities; and (3) to change the culture and practice of graduate preparation such that undergraduate instruction is taken more seriously (Austin & Wulff, 2004; Gaff & Lambert, 1996; Goldsmith, Haviland, Dailey, & Wiley, 2004).

Although future faculty programs that emphasize TD are gaining momentum, we still know very little collectively about their effects, especially on their participants' preparation as college instructors (Austin, 2010). Findings from small-scale studies and program evaluations may not be useful to those concerned with improving undergraduate STEM education on a national scale. Administrators (including future faculty program coordinators), funders, future faculty, and their graduate advisors lack credible evidence that these programs enhance traditional doctoral training, improve participants' career options, or enhance their early career performance as academics. Lack of such evidence may affect not only doctoral students' interest in these programs but also the programs' sustainability. Thus, if developing better undergraduate teachers is to advance the national STEM agenda, then more and better information is needed about the effects of typical TD programs for STEM doctoral students.

To address this issue, we conducted a longitudinal study of the impact of TD during doctoral training on a panel of early career academics who in 2009 were late-stage doctoral students in STEM departments at three U.S. research universities. In this paper, we first explain why social cognitive career theory (SCCT) is useful for understanding the short- and long-term impact of future faculty programs. Next, we describe a 2011 survey of early career academics and our procedures for analyzing those data. Third, we present findings showing that participation in doctoral TD significantly influences early career academics' beliefs about their efficacy as college teachers, and that TD offerings requiring more time and engagement have the greatest influence on those efficacy beliefs. Finally, we discuss our findings' implications for the professional development of STEM early career academics against a backdrop of national efforts to improve undergraduate STEM teaching and learning.

Conceptual Framework for Understanding Doctoral Teaching Development

Although researchers frequently use socialization theory to understand the doctoral student experience (e.g., Austin & McDaniels, 2006; Gardner & Mendoza, 2010; Weidman, Twale, & Stein, 2001), the theory's strengths lie more in being descriptive of processes than predictive of outcomes. SCCT offers a different way to look at TD programs and their role in the formation of future faculty (Lent, Brown, & Hackett, 1994, 2000). SCCT is a career development model based on Bandura's (1986, 1997, 2005) general social cognitive theory, which posits that people learn by watching what others do and that the beliefs people hold about themselves are key to their personal agency. To explain how one's career goals, career expectations, and sense of personal efficacy collectively shape career choices, SCCT integrates four models: (1) how career interests are formed, (2) how key career-related choices are made, (3) what constitutes effective job performance, and (4) what constitutes satisfying work. SCCT also accounts for how personal characteristics, social contexts, and learning experiences influence career-related choices and outcomes.

Self-efficacy for college teaching. A central construct of social cognitive theory (and thus SCCT) is *self-efficacy*, which is a person's confidence in his or her ability to carry out a particular task or course of action (Bandura, 1986; Lent & Brown, 2006). By itself, self-efficacy has been studied widely, especially in educational contexts (Klassen & Usher, 2010; Pajares, 1996; Usher & Pajares, 2008). As applied to K–12 teachers, "a teacher's efficacy belief is a judgment of his or her capabilities to bring about desired outcomes of student engagement and learning, even among those students who may be difficult or unmotivated" (Tschannen-Moran & Woolfolk Hoy, 2001, p. 783). Teachers with a greater sense of teaching self-efficacy tend to be more enthusiastic, devote more time to planning and organization, show greater commitment, are more likely to experiment with new methods, and are more persistent under adverse circumstances (Woolfolk Hoy, 2004). In a review, Ross (2013) asserts, "The evidence is consistent: few teacher characteristics have as much impact on instructional practice and student outcomes as teacher efficacy" (p. 266).

Although the bulk of research on teaching self-efficacy is based on K–12 teachers, a few studies have applied Bandura's ideas about self-efficacy to postsecondary instructors (e.g., Major & Dolly, 2003; Prieto & Altmaier, 1994; Santiago & Einarson, 1998). Because one's self-efficacy tends to be most malleable when a skill is first being learned, teaching-efficacy beliefs of current and future faculty (hereafter called "college teaching self-efficacy") are significantly shaped while working as teaching assistants or participating in TD programs (Woolfolk Hoy, 2004). Given the robust body of evidence for the influence of self-efficacy on task performance (e.g., Pajares, 1996; Sitzmann & Ely, 2011), we should learn more about the college teaching self-efficacy of current and aspiring postsecondary faculty because these beliefs are very likely to influence how they perform as undergraduate instructors.

Thus, the primary advantage of using SCCT as a theoretical framework for the present study is that it not only examines self-efficacy related to a specific performance domain—college teaching—but also situates self-efficacy in a career development model. As such, SCCT is well suited to studying how college teaching self-efficacy of aspiring academics may affect, and be affected by, their career interests, their involvement in professional development activities, and their eventual career choices. Surprisingly, SCCT has never been applied to aspiring academics (i.e., late-stage doctoral students and recent doctorate recipients). Viewing the formation of postsecondary teachers through this lens is challenging because of the complexity of the SCCT model and the importance of constructing valid measures for each part of the model being investigated. If applied systematically and judiciously, however, SCCT has the potential to better explain the short-term effects of TD on doctoral student participants and the long-term effects on early career academics.

Because research shows self-efficacy beliefs may vary by gender and race/ethnicity, SCCT can also be used to examine the extent to which gender and race/ethnicity interact with the effects of TD programs on college teaching self-efficacy. Lindley's (2006) review, for example, examined critical differences in the strength of certain predictor variables for women and U.S. ethnic minorities. In most STEM domains, the self-efficacy beliefs of these underrepresented groups were often, but not always, lower than that of their majority peers. Huang's 2013 meta-analysis of 247 independent studies on gender differences in academic self-efficacy found a small difference favoring males. Lindley (2006) further noted that research on self-efficacy sometimes shows differences among racial/ethnic groups but not consistently.

Research Questions

Using SCCT, we hypothesized that TD offerings include learning experiences that can directly influence college teaching self-efficacy, indirectly influence one's career interest and choice, and contribute to subsequent job performance and satisfaction. We focused only on the relationship between teaching-development experiences and college teaching self-efficacy of late-stage doctoral students and recent doctorate recipients—a group we call *early career academics*. The present study explored four research questions related to early career academics' college teaching self-efficacy:

(1) Does any kind of participation in teaching development activities affect college teaching self-efficacy?
(2) Does the degree of one's engagement in teaching development activities affect college teaching self-efficacy?
(3) Does the type of teaching development activity in which doctoral students participate affect college teaching self-efficacy?
(4) Do the effects of teaching development interact with gender and race/ethnicity?

Method

In 2008, the National Science Foundation (NSF) funded a 5-year longitudinal study of the effects of TD programs on future STEM scholars. The mixed-method study followed a panel of late-stage doctoral students using repeated surveys and interviews to (1) track their progress toward and after completing their doctorate, and (2) explore the short- and long-term effects of TD participation on their pedagogical preparation, career choices, and early career success. As part of that larger NSF-funded project, the present study examined the effects of TD on college teaching self-efficacy of study participants who were enrolled as doctoral students at the three participating institutions in 2008 and who responded to a survey in 2009.

Participant Characteristics and Sampling

Data were collected two years apart (Year 1 = 2009, Year 3 = 2011) using two survey questionnaires developed for this study. The sampling frame for the Year 1 instrument consisted of 3,060 late-stage doctoral students in STEM departments at Arizona State University, University of Washington–Seattle, and University of Wisconsin–Madison. To define STEM for this study, we used NSF's four broad disciplinary categories: life sciences (which includes agricultural and biological sciences and health-related fields); physical sciences (computer, physical, mathematical, and earth, atmospheric, and oceanic sciences); engineering; and psychology and social sciences (see Table 1). Contact information for enrolled doctoral students in STEM departments was obtained directly from the graduate schools of the three institutions following approval from their respective institutional review boards. The Year 1 questionnaire examined STEM doctoral students' participation in TD during their doctoral education. After its items were piloted and refined, the questionnaire was administered in paper- and web-based formats in summer 2009 by a major survey center. The response rate was 73% (n = 2,163).

The second instrument was designed to measure SCCT constructs and gather information about respondents' current employment (e.g., employment sector, job title, distribution of academic responsibilities). With the exception of using a web-only format, the method of administering the survey in 2011 was the same as in 2009. Of the 2,156 Year 1 respondents who could be reached, 1,445 responded (67%).[2] Of these Year 3 respondents, 977 (68%) had earned their doctorates, and 468 (32%) were still enrolled in Ph.D. programs. These late-stage doctoral

[2] Although this response rate is fairly good given the longitudinal nature of this survey, if any attrition at Year 3 is systemically related to respondents' characteristics, it may introduce attrition bias, which makes it hard to make valid inferences from these data. While we found that some characteristics, particularly race/ethnicity, doctorate-granting institution, TD, and teaching experience were associated with the likelihood of response in the Year 3 survey, it turned out that the disproportionality of the response for some groups was already explained by other covariates. Using inverse probability weights and propensity scores (Baulch & Quisumbing, 2011; Cuddeback, Wilson, Orme, & Combs-Orme, 2004; Fitzgerald, Gottschalk, & Moffitt, 1998; Foster & Fang, 2004; Miller & Hollist, 2007), we found that attrition bias did not substantially alter our conclusions, and it affected only a few estimates close to the margin of significance (p = 0.05). Although we found attrition from the 2011 survey (Year 3) did not substantially change our results, our data may still suffer from non-response bias if those who did not respond to the 2009 survey (Year 1) were systematically different from those who did.

students and doctorate recipients constitute our group of early career academics. Variables used in this study are described in Appendix A.

Table 1
Descriptive Statistics for Independent Variables

Variables	*n*	%
TD Participation	1,436	100.00
No	279	19.42
Yes	1,157	80.58
TD Engagement	1,058	100.00
Low engagement (<= 10 hours)	285	26.94
Low moderate engagement (> 10 hours, <=25 hours)	305	28.83
High moderate engagement (> 25 hours, <= 55 hours)	256	24.20
High engagement (> 55 hours)	212	20.04
TD Type	1,353	100.00
Non-participant	278	20.55
Non-intensive	368	27.20
Intensive	248	18.33
Formal courses	459	33.92
Amount of Teaching Experience	1418	100.00
Low teaching experience (<= 1 semester)	251	17.70
Low moderate teaching experience (> 1, <= 3 semesters)	486	34.27
High moderate teaching experience (> 3, <=6 semesters)	319	22.5
High teaching experience (> 6 semesters)	362	25.53
Gender	1,418	100.00
Male	757	53.39
Female	661	46.61
Race/ethnicity	1,400	100.00
White	1,039	74.21
Asian American	253	18.07
Other minority	108	7.71
Citizenship	1,422	100.00
U.S. citizen and permanent resident	1,129	79.40
Other	293	20.60
Year Doctoral Studies Began	1,420	100.00
~ 2000	16	1.13
2001–2003	436	30.70
2004–2006	915	64.44
2007– 2008	53	3.73
Primary Career Goal at Start of Doctoral Studies	1,421	100.00
Other	289	20.34
Faculty career	374	26.32
Research career	758	53.34
Interest in Teaching at Start of Doctoral Studies	1,430	100.00
Not at all interested	161	11.26
Slightly interested	329	23.01
Somewhat interested	441	30.84
Very interested	349	24.41
Extremely interested	150	10.49

Table 1, cont.

Variables	*n*	%
Principal Field of Study	1,425	100.00
Agricultural sciences	18	1.26
Biological sciences	416	29.19
Computer sciences	51	3.58
Earth, atmospheric, and oceanic sciences	62	4.35
Engineering	184	12.91
Health fields	69	4.84
Mathematical sciences	77	5.40
Other	61	4.28
Physical sciences	240	16.84
Psychology	58	4.07
Social sciences	189	13.26
Institutions	1,445	100.00
Arizona State University	188	13.01
University of Washington–Seattle	704	48.72
University of Wisconsin–Madison	553	38.27
Interest in Becoming a Faculty Member	1,419	100.00
No	257	18.11
Not sure	350	24.67
Yes	812	57.22
TD Participation is Required	1,445	100.00
No	657	45.47
Yes	788	54.53
Completed Doctorate	1,445	100.00
Currently enrolled in a Ph.D. program	468	32.39
Graduated from a Ph.D. program	977	67.61
Total	1,445	100.00

Outcome Measure

The outcome measure of interest was college teaching self-efficacy. Previous research has linked teachers' self-efficacy to teaching performance and preferred student outcomes (Ross, 1998; Tschannen-Moran & Woolfolk Hoy, 2001). Because self-efficacy beliefs are specific to a particular performance domain, measuring self-efficacy requires breaking a performance domain into meaningful and ostensibly independent components (Bandura, 2006; Betz & Hackett, 2006; Lent & Brown, 2006).

Drawing on work that attempts to identify the various components of college teaching (e.g., Chism, 2007; Hativa, 2014; Lowman, 1995; Theall, Mullinix, & Arreola, 2009) and a synthesis of research on college teaching in science and math (National Research Council, 2003), we subdivided the general domain of postsecondary teaching into the following six components: (1) planning courses, (2) refining teaching methods, (3) creating learning environments, (4) assessing student learning, (5) interacting with students, and (6) mastering subject knowledge. Each was originally measured by five items using five-point Likert-type scales ranging from 1 *(not at all confident)* to 5 *(extremely confident)*.

To address content validity, we asked 10 scholars and administrators with expertise in faculty development, doctoral education, and undergraduate STEM education to review our items, which we revised based on their feedback. The construct validity of college teaching self-efficacy was further addressed by exploratory factor analysis (EFA) and confirmatory factor analysis (CFA), which involved splitting the study's sample randomly into halves, using EFA on one sub-sample to identify the factor structure of college teaching self-efficacy, and then using CFA on the second sub-sample to test this factor structure. In EFA of the first sub-sample (n = 695), the maximum likelihood factor analysis with an oblique rotation was used to allow for correlations among all extracted factors. Items that did not load onto their corresponding factors (i.e., factor loadings lower than 0.5) and that loaded to a substantial degree on more than one factor were removed. EFA identified six factors from which seven items were discarded. Appendix A describes each factor, its original items, and items trimmed through EFA. The Cronbach's alpha coefficient of all factors ranged from 0.88–0.91, which indicated that internal consistency reliability was high. For the second part, CFA was run on the second sub-sample (n = 696) to assess model fit to the data. A six-factor model with the 23 items resulting from EFA was tested and provided an acceptable fit to the data (RMSEA = 0.058, RMSEA CI_{90} = 0.053–0.063; CFI = 0.958; TLI = 0.950; SRMR = 0.035). Therefore, EFA and CFA confirmed a six-dimensional structure of college teaching self-efficacy with 23 items.

In our analyses, we treated each of the six dimensions of college teaching self-efficacy as an outcome. Each dimension had three, four, or five items, measured as five-point Likert-type scales. The average item scores for each dimension were used as dependent variables in the analysis. Each average item score was standardized (mean = 0; standard deviation = 1).

Independent Variables of Interest

Our primary objective was to estimate the effects of TD during the doctoral program on college teaching self-efficacy of early career academics in STEM fields. We expected that (1) participation in TD would have a positive impact on college teaching self-efficacy, and (2) benefits from TD programs might differ according to participants' level of engagement and the type of participation. To examine these hypotheses, we constructed three independent variables of interest. The first was participation in TD, measured dichotomously as "yes" or "no." The second was the level of engagement in TD activities, measured at five levels: *none, low* (1–10 TD contact hours), *low–moderate* (11–25 hours), *high–moderate* (26–55 hours), and *high* (> 55 hours).[3] The third was the type of participation in TD activities, measured as *non-intensive, intensive,* and *formal courses*. *Non-intensive* participation includes *only* non-intensive TD offerings, including talks, presentations, and workshops, conferences, and other activities lasting less than a day. *Intensive* participation refers to intensive trainings, workshops, conferences, and symposia, *excludes* formal courses, and *includes* any participation in non-intensive type of TD. *Formal courses* refers to participation in formal pedagogy courses, *excludes* participation in intensive TD programs, and

[3] Because there is no known research into what constitutes a meaningful amount of engagement in TD for doctoral students, the cut-points for the level of engagement in this study are tentative and exploratory. Other cut-points were also tested and did not substantially change our conclusions. Of TD participants, each group represents approximately 27%, 29%, 24%, and 20%, respectively.

includes any participation in non-intensive type of TD. Intensive participation and formal courses were both defined as including non-intensive TD programs because people who participated in intensive programs and formal courses also participated in non-intensive programs.

Covariates

To estimate the effect of the TD program, we controlled for not only respondents' demographic and academic backgrounds but also their teaching experience and initial career interest, which may affect respondents' college teaching self-efficacy, TD participation, TD engagement, and TD type.[4] Covariates included (1) gender, (2) race/ethnicity, (3) citizenship, (4) year that doctoral study began, (5) amount of teaching experience, (6) primary career goal, (7) level of interest in teaching when starting the doctoral program, (8) principal field of study, (9) participant's doctorate-granting institution, (10) level of interest in becoming faculty members, (11) whether TD participation was required, and (12) whether they completed their doctorates.[5]

According to social cognitive theory, a person's self-efficacy beliefs are shaped by four sources of information: mastery experiences, vicarious experiences, verbal persuasion from others, and one's own emotional and physiological states (Bandura, 1997; Usher & Pajares, 2008). Of these four sources, mastery experiences are the most influential in shaping self-efficacy "because [mastery experiences] provide the most authentic evidence of whether one can muster whatever it takes to succeed" (Bandura, 1997, p. 80). Thus, we hypothesized that doctoral students would be exposed to sources of self-efficacy information from not only participation in TD but also from authentic teaching experiences (e.g., being a teaching assistant, guest lecturer, or instructor of record). Furthermore, we assumed that doctoral students with teaching experience were also likely to participate in TD programs. Thus, to test a relationship between the TD programs and college teaching self-efficacy, it was necessary to partial out the effect of teaching experience. Because 94% of our survey respondents had actual teaching experience (not reported in tables) and the relationship between college teaching self-efficacy and the amount of college teaching experience was not linear, we created four dummy variables indicating the amount of college teaching experience based on total semesters of diverse teaching activities (range: 0–22 semesters; mean: 4.5 semesters; standard deviation: 4.1 semesters): *low* (≤ 1 semester), *low-moderate* (>1, ≤ 3 semesters), *high-moderate* (> 3, ≤ 6 semesters), and *high* (> 6 semesters). According to SCCT (Lent et al., 1994), students' initial career interests are also likely to affect their participation in certain types of learning experiences (e.g., TD programs, departmental training for teaching assistants). To adjust for their initial career aspirations, we controlled for their primary career goal, the level of interest in becoming faculty members, and

---

[4] Admittedly, we recognize that self-confident doctoral students are more likely to participate or be engaged in TD activities, which could lead a self-selection bias for our estimates on the relationship between respondents' college-teaching self-efficacy and TD experiences. To reduce this self-selection bias, we controlled for not only respondents' teaching experience, but also their career goal and interest in teaching at the start of doctoral studies, which are important factors affecting respondents' previous self-efficacy on college teaching. However, the extent to which unobserved factors affect both TD participation and college-teaching self-efficacy, our estimates could be biased same as traditional regression models.

[5] Race/ethnicity and citizenship status were collected from separate questionnaire items. Therefore, racial categories such as "Asians" include both U.S. citizens and non-citizens.

the level of interest in teaching when starting their doctoral programs.

Even though we followed a panel of late-stage doctoral students since 2009, there could be significant variation in time to begin and complete their doctorate. To account for possible cohort effects, we controlled for the year that doctoral study began and current academic status (i.e., whether they completed their doctorate). Because academic fields have structural differences (Becher & Trowler, 2001; Braxton & Hargens, 1996), we accounted in our analysis for differences among academic fields by including a series of dummy variables indicating respondents' principal fields of study. In addition, in some departments, participation in TD programs was required during doctoral training, which could be a source of self-selection bias in estimating the effect of the TD program. Thus, we also controlled in our analysis for whether respondents' departments required TD participation (see Appendix A for definitions of covariates and Table 1 for descriptive statistics of the independent variables).

Analysis

We used ordinary least squares (OLS) regression to estimate the effect of the TD program on STEM doctoral students' college teaching self-efficacy. To address our four research questions, we examined the multivariate relationship after accounting for appropriate covariates. Then, we tested the interaction effects of race/ethnicity and gender with the TD program by adding the interaction terms of the TD experience with gender and race/ethnicity. The effects of participation in TD, of the level of engagement in TD, and of the type of TD were estimated separately.

Results

Participation in Teaching Development

Table 2 shows results from the OLS regression of college teaching self-efficacy on participation in TD. All bivariate relationships were statistically significant and positive (not reported). After controlling for relevant covariates, participation in TD activities still had a statistically positive impact on STEM early career academics' college teaching self-efficacy. The gap between TD participants and non-participants ranged from 0.17 to 0.25 standard deviation (*SD*). Specifically, TD participants were more confident than non-participants in planning courses (0.185, $p < 0.05$ in M1), using teaching methods (0.251, $p < 0.01$ in M2), and mastering subject knowledge (0.167, $p < 0.05$ in M6). Differences between participants and non-participants on the factors for creating learning environments, assessing student learning, and interacting with students were not significant.

It is worth noting the effects of other covariates on college teaching self-efficacy. Table 2 clearly shows that, as expected, teaching experience played an important role in improving STEM early career academics' college teaching self-efficacy. Even when accounting for other covariates, STEM early career academics with the higher level of teaching experience showed a higher level of confidence in all dimensions of college teaching than those with the low and low-moderate levels of teaching experience. While the effect of teaching experience was relatively weak on mastering subject knowledge (see M6 in Table 2), early career academics with high-moderate or high levels of teaching experience showed considerably higher levels of confidence in other dimensions of college teaching than those with low teaching experience. Their gaps range from approximately 0.38 to 0.70 of the standard deviation (see M1 – M5 in Table 2). These findings indicate that actual teaching experience positively contributes to improving STEM early career academics' college teaching self-efficacy in addition to the TD activities, and each has its own contribution.

Accounting for all covariates, women STEM early career academics were significantly less confident in their abilities than men in five of six college teaching domains: planning courses (-0.245, $p < 0.001$ in M1), using teaching methods (-0.107, $p < 0.05$ in M2), assessing student learning (-0.109, $p < 0.05$ in M4), interacting with students (-0.228, $p < 0.001$ in M5), and mastering subject knowledge (-0.237, $p < 0.001$ in M6); only in creating learning environments were men and women similarly confident. Table 2 also shows racial disparities in some dimensions of college teaching self-efficacy. Compared with other non-Asian minorities, White students were less confident in planning courses and interacting with students, Asian students were less confident in mastering subject knowledge, and White and Asian students were less confident than non-Asian minorities in creating learning environments.

Table 2
OLS Regression of Self-Efficacy on TD Program Participation

Variable	Planning Courses (b/se) M1	Using Teaching Methods (b/se) M2	Creating Learning Environ-ments (b/se) M3	Assessing Student Learning (b/se) M4	Interacting with Students (b/se) M5	Mastering Subject Knowledge (b/se) M6
TD Participation	0.185* (0.076)	0.251** (0.078)	0.134 (0.082)	0.133 (0.080)	0.094 (0.082)	0.167* (0.083)
Amount of Teaching Experience (ref: Low teaching experience)						
Low Moderate (> 1, <= 3 semesters)	0.150* (0.072)	0.258*** (0.075)	0.236** (0.078)	0.291*** (0.076)	0.134 (0.078)	0.078 (0.079)
High Moderate (> 3, <= 6 semesters)	0.503*** (0.081)	0.529*** (0.083)	0.455*** (0.086)	0.550*** (0.084)	0.376*** (0.087)	0.285** (0.088)
High (> 6 semesters)	0.655*** (0.084)	0.659*** (0.087)	0.582*** (0.090)	0.699*** (0.088)	0.479*** (0.091)	0.270** (0.091)
Female	-0.245*** (0.051)	-0.107* (0.053)	0.009 (0.055)	-0.109* (0.053)	-0.228*** (0.055)	-0.237*** (0.055)
Race/ethnicity (ref: minority)						
White	-0.225* (0.095)	-0.172 (0.098)	-0.275** (0.102)	-0.088 (0.100)	-0.226* (0.103)	-0.170 (0.104)
Asian	-0.108 (0.110)	-0.042 (0.113)	-0.252* (0.118)	-0.047 (0.116)	-0.154 (0.119)	-0.265* (0.119)
N	1346	1347	1345	1338	1338	1328
Adjusted R^2	0.223	0.181	0.100	0.152	0.112	0.096

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Covariates included in the analysis but not reported in this table include citizenship, year doctoral studies began, primary career goal at start of doctoral studies, interest in teaching at start of doctoral studies, institution, principal field of study, interest in becoming a faculty member, required teaching development, and completed doctorate.

Degree of Engagement in Teaching Development

Table 3 shows results from the OLS regression of college teaching self-efficacy on level of TD engagement (low, low-moderate, high-moderate, and high). The reference group is non-participants. Models without covariates consistently showed that greater engagement in TD led to greater benefits for STEM early career academics (not reported). After accounting for relevant covariates, STEM early career academics in the high-moderate engagement group were significantly more confident in planning courses (0.230, $p < 0.05$ in M1) and using teaching methods (0.291, $p < 0.01$ in M2) than non-participants. Although differences between low-moderate and high-moderate engagement groups were not significant, high-moderate engagement groups were significantly more confident than low-engagement groups in M1 and M2 ($p < 0.05$).

Table 3
OLS Regression of Self-Efficacy on the TD Program Engagement

	Planning Courses (b/se)	Using Teaching Methods (b/se)	Creating Learning Environments (b/se)	Assessing Student Learning (b/se)	Interacting with Students (b/se)	Mastering Subject Knowledge (b/se)
	M1	M2	M3	M4	M5	M6
TD Engagement (ref: non-participants)						
Low (<10 hours)	0.054	0.107	0.079	0.114	0.045	0.165
	(0.091)	(0.094)	(0.098)	(0.096)	(0.100)	(0.100)
Low Moderate (>10, <=25 hours)	0.114	0.172	0.082	0.126	-0.001	0.194
	(0.091)	(0.094)	(0.098)	(0.096)	(0.100)	(0.099)
High Moderate (>25, <=55 hours)	0.230*	0.291**	0.073	0.120	0.053	0.184
	(0.094)	(0.097)	(0.101)	(0.099)	(0.103)	(0.103)
High (>55 hours)	0.389***	0.415***	0.189	0.246*	0.164	0.141
	(0.094)	(0.097)	(0.101)	(0.099)	(0.103)	(0.103)
Amount of Teaching Experience (ref: low teaching experience)						
Low Moderate (> 1, <= 3 semesters)	0.130	0.245**	0.226**	0.261***	0.104	0.069
	(0.075)	(0.077)	(0.081)	(0.078)	(0.082)	(0.082)
High Moderate (> 3, <= 6 semesters)	0.449***	0.514***	0.447***	0.520***	0.354***	0.290**
	(0.083)	(0.086)	(0.090)	(0.087)	(0.091)	(0.091)
High (> 6 semesters)	0.596***	0.605***	0.568***	0.646***	0.434***	0.244*
	(0.087)	(0.089)	(0.093)	(0.091)	(0.095)	(0.095)
Female	-0.251***	-0.109*	0.018	-0.105	-0.226***	-0.229***
	(0.052)	(0.054)	(0.056)	(0.055)	(0.057)	(0.057)
Race/ethnicity (ref: minority)						
White	-0.209*	-0.153	-0.277**	-0.077	-0.206	-0.163
	(0.097)	(0.101)	(0.105)	(0.103)	(0.107)	(0.107)
Asian	-0.111	-0.071	-0.282*	-0.039	-0.171	-0.265*
	(0.113)	(0.116)	(0.121)	(0.119)	(0.124)	(0.123)
N	1265	1266	1264	1257	1257	1247
Adjusted R^2	0.228	0.185	0.097	0.145	0.103	0.093

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$

Covariates included in the analysis but not reported in this table include citizenship, year doctoral studies began, primary career goal at start of doctoral studies, interest in teaching at start of doctoral studies, institution, principal field of study, interest in becoming a faculty member, required teaching development, and completed doctorate.

Compared with non-participants, respondents in the high-engagement group substantially benefited from TD activities, especially in planning courses (0.389, $p < 0.001$ in M1) and using teaching methods (0.415, $p < 0.001$ in M2). Although effects were relatively small for assessing student learning (0.246, $p < 0.05$ in M4), they were still statistically significant and not trivial. Similarly, while differences between high-moderate and high engagement groups were not significant, high engagement groups were significantly more confident than low-moderate engagement groups in M1 and M2 ($p < 0.05$). Not all dimensions were associated with TD engagement, however: There was no significant relationship at any level of engagement with mastering subject knowledge, interacting with students, or creating a learning environment.

Type of TD Program

Table 4 shows results from the OLS regression of college teaching self-efficacy on the type of TD program (non-intensive, intensive, and formal courses). Participation in formal courses had significant impact on confidence in planning courses (0.340, $p < 0.001$ in M1), using teaching methods (0.393, $p < 0.001$ in M2), and assessing student learning (0.202, $p < 0.05$ in M4), but the effects of formal courses on confidence in creating learning environments, interacting with students, and mastering subject knowledge were not statistically significant. When compared with non-participants, early career academics who participated in intensive TD activities gained confidence in using teaching methods (0.276, $p < 0.01$ in M2). Differences between formal courses and intensive TD activities were significant only in planning courses, whereas differences between formal courses and non-intensive TD activities were significant in planning courses and using teaching methods ($p < 0.05$). These findings suggest that participation in intensive TD activities or formal courses may make a significant contribution to improving STEM early career academics' college teaching self-efficacy in a way that participation in non-intensive TD courses does not.

Table 4
OLS Regression of Self-Efficacy on the TD Program Type

	Planning Courses (b/se)	Using Teaching Methods (b/se)	Creating Learning Environ-ments (b/se)	Assessing Student Learning (b/se)	Interactin g with Students (b/se)	Mastering Subject Knowledge (b/se)
	M1	M2	M3	M4	M5	M6
TD Type (ref: non-participants)						
Non-Intensive	0.071	0.087	0.080	0.107	0.073	0.159
	(0.086)	(0.088)	(0.092)	(0.090)	(0.093)	(0.094)
Intensive	0.145	0.276**	0.152	0.137	0.014	0.206
	(0.098)	(0.100)	(0.105)	(0.102)	(0.107)	(0.107)
Formal Course	0.340***	0.393***	0.125	0.202*	0.092	0.163
	(0.085)	(0.087)	(0.091)	(0.089)	(0.093)	(0.093)
Amount of Teaching Experience (ref: low teaching experience)						
Low Moderate (> 1, <= 3 semesters)	0.123	0.233**	0.213**	0.247**	0.092	0.069
	(0.075)	(0.077)	(0.080)	(0.078)	(0.081)	(0.082)
High Moderate (> 3, <= 6 semesters)	0.454***	0.504***	0.440***	0.509***	0.355***	0.288**
	(0.083)	(0.085)	(0.089)	(0.087)	(0.090)	(0.091)
High (> 6 semesters)	0.610***	0.612***	0.575***	0.651***	0.448***	0.253**
	(0.086)	(0.088)	(0.092)	(0.090)	(0.094)	(0.094)
Female	-0.255***	-0.120*	0.013	-0.113*	-0.225***	-0.231***
	(0.052)	(0.054)	(0.056)	(0.055)	(0.057)	(0.057)
Race/ethnicity (ref: minority)						
White	-0.217*	-0.171	-0.271**	-0.073	-0.208	-0.161
	(0.097)	(0.099)	(0.104)	(0.102)	(0.106)	(0.106)
Asian	-0.100	-0.061	-0.263*	-0.020	-0.160	-0.254*
	(0.112)	(0.115)	(0.120)	(0.118)	(0.123)	(0.123)
N	1273	1274	1272	1265	1265	1256
Adjusted R^2	0.231	0.191	0.097	0.145	0.105	0.092

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$
Covariates included in the analysis but not reported in this table include citizenship, year doctoral studies began, primary career goal at start of doctoral studies, interest in teaching at start of doctoral studies, institution, principal field of study, interest in becoming a faculty member, required teaching development, and completed doctorate.

The level of engagement and the type of TD program correlate. That is, participants in formal courses or intensive TD activities are likely to spend more time in TD programs than participants in non-intensive TD activities because these TD programs require intensive time commitment. To partial out this relationship, we estimated the effect of TD engagement and the effect of TD type simultaneously in Table 5. Even after accounting for the type of TD activities and relevant covariates, high engagement still had a significant impact on confidence in planning courses (0.265, $p < 0.05$ in M1) and using teaching methods (0.235, $p < 0.05$ in M2). Participation in intensive TD activities or formal courses also had a significant impact on confidence in using teaching methods (0.171, $p < 0.05$ and 0.217, $p < 0.01$ in M2), even after accounting for level of engagement in TD and relevant covariates. This finding suggests that TD engagement and TD type separately contribute to improving STEM early career academics' college teaching self-efficacy.

Table 5
OLS Regression of Self-Efficacy on the TD Program Engagement and Type

	Planning Courses (b/se)	Using Teaching Methods (b/se)	Creating Learning Environ-ments (b/se)	Assessing Student Learning (b/se)	Interacting with Students (b/se)	Mastering Subject Knowledge (b/se)
	M1	M2	M3	M4	M5	M6
TD Engagement (ref: non-participants)						
Low (<10 hours)	0.030	0.063	0.075	0.106	0.061	0.161
	(0.092)	(0.095)	(0.100)	(0.097)	(0.101)	(0.101)
Low Moderate (>10, <=25 hours)	0.067	0.070	0.062	0.106	0.035	0.181
	(0.099)	(0.101)	(0.106)	(0.103)	(0.107)	(0.108)
High Moderate (>25, <=55 hours)	0.140	0.145	0.083	0.097	0.113	0.179
	(0.109)	(0.112)	(0.118)	(0.114)	(0.119)	(0.119)
High (>55 hours)	0.265*	0.235*	0.198	0.206	0.222	0.139
	(0.115)	(0.119)	(0.124)	(0.120)	(0.126)	(0.126)
TD Type (ref: non-participants and non-intensive)						
Intensive and Others	0.046	0.171*	0.071	0.029	-0.069	0.041
	(0.081)	(0.084)	(0.087)	(0.085)	(0.088)	(0.088)
Formal Course and Others	0.154	0.217**	-0.009	0.054	-0.063	0.001
	(0.080)	(0.082)	(0.086)	(0.083)	(0.087)	(0.087)
Amount of Teaching Experience (ref: low teaching experience)						
Low Moderate (> 1, <= 3 semesters)	0.119	0.233**	0.209**	0.246**	0.091	0.068
	(0.075)	(0.077)	(0.081)	(0.079)	(0.082)	(0.082)
High Moderate (> 3, <= 6 semesters)	0.432***	0.495***	0.433***	0.503***	0.342***	0.290**
	(0.083)	(0.086)	(0.090)	(0.087)	(0.091)	(0.091)
High (> 6 semesters)	0.583***	0.588***	0.550***	0.630***	0.421***	0.243*
	(0.087)	(0.089)	(0.093)	(0.091)	(0.095)	(0.095)
Female	-0.262***	-0.123*	0.014	-0.112*	-0.226***	-0.229***
	(0.052)	(0.054)	(0.056)	(0.055)	(0.057)	(0.057)
Race/ethnicity (ref: minority)						
White	-0.216*	-0.163	-0.272**	-0.077	-0.199	-0.163
	(0.097)	(0.100)	(0.105)	(0.103)	(0.107)	(0.107)
Asian	-0.112	-0.070	-0.280*	-0.039	-0.171	-0.264*
	(0.113)	(0.116)	(0.121)	(0.119)	(0.124)	(0.123)
N	1264	1265	1263	1256	1256	1247
Adjusted R^2	0.229	0.188	0.095	0.142	0.102	0.092

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$
Covariates included in the analysis but not reported in this table include citizenship, year doctoral studies began, primary career goal at start of doctoral studies, interest in teaching at start of doctoral studies, institution, principal field of study, interest in becoming a faculty member, required teaching development, and completed doctorate.

Interaction with Gender and Race/ethnicity

Tables 6, 7, and 8 show interaction results from the OLS regression of self-efficacy on TD participation, engagement, and type (only main effects and interaction effects are presented). In Table 6, interaction terms of TD participation with gender and race/ethnicity were added and estimated in each model with relevant covariates. Although there was no significant interaction effect of TD participation with race/ethnicity (not reported in Table 6), interaction effects of being a woman with TD participation were statistically significant and positive in predicting confidence in planning courses (0.377, $p < 0.01$ in M1), using teaching methods (0.316, $p < 0.05$ in M2), assessing student learning (0.274, $p < 0.05$ in M4), and interacting with students (0.343, $p < 0.05$ in M5).

Table 6
OLS Regression of Self-Efficacy on the Interaction of Female with TD Program Participation

	Planning Courses (b/se)	Using Teaching Methods (b/se)	Creating Learning Environments (b/se)	Assessing Student Learning (b/se)	Interacting with Students (b/se)	Mastering Subject Knowledge (b/se)
	M1	M2	M3	M4	M5	M6
Female	-0.555***	-0.367**	-0.103	-0.334**	-0.511***	-0.352**
	(0.118)	(0.122)	(0.127)	(0.124)	(0.127)	(0.129)
TD Participation	-0.182	-0.046	0.054	-0.194	-0.293	-0.108
	(0.246)	(0.254)	(0.265)	(0.258)	(0.265)	(0.267)
Interaction						
Female X TD Participation	0.377**	0.316*	0.135	0.274*	0.343*	0.140
	(0.129)	(0.133)	(0.138)	(0.135)	(0.138)	(0.140)
N	1346	1347	1345	1338	1338	1328
Adjusted R^2	0.227	0.183	0.098	0.154	0.115	0.096

* p < 0.05; ** p < 0.01; *** p < 0.001

Note. Only main and interaction effects are presented. Interaction effects of being female with race are not presented. All covariates are included in each model. Covariates included in the analysis but not reported in this table include citizenship, year doctoral studies began, primary career goal at start of doctoral studies, interest in teaching at start of doctoral studies, institution, principal field of study, interest in becoming a faculty member, required teaching development, and completed doctorate.

In Table 7, we include the level of TD engagement and its interaction terms with gender and race/ethnicity instead of TD participation and its interaction terms. As in Table 6, there was no significant interaction effect of race/ethnicity with any level of TD engagement (not reported in Table 7). However, we found consistently positive interaction effects of TD engagement with gender on self-efficacy, especially for women in the high-engagement group. An exception was the interaction effect on confidence in creating learning environments (0.009, $p > 0.05$ in M3). When compared with non-participants, gender interaction effects with the high-engagement group were substantial across all models except M3 (0.674, $p < 0.001$ in M1; 0.595, $p < 0.001$ in M2; 0.429, $p < 0.05$ in M4; 0.454, $p < 0.05$ in M5; 0.387, $p < 0.05$ in M6). For example, the standardized average score in confidence in planning courses was 0.674 *SD* higher for women participants in the high-engagement group than women non-participants.

Table 7
OLS Regression of Self-Efficacy on the Interaction of Female with TD Program Engagement

	Planning Courses (b/se)	Using Teaching Methods (b/se)	Creating Learning Environ-ments (b/se)	Assessing Student Learning (b/se)	Interacting with Students (b/se)	Mastering Subject Knowledge (b/se)
	M1	M2	M3	M4	M5	M6
Female	-0.560***	-0.365**	-0.099	-0.332**	-0.514***	-0.363**
	(0.118)	(0.121)	(0.127)	(0.123)	(0.128)	(0.129)
TD Engagement (ref: non-participants)						
Low (<10 hours)	-0.278	-0.329	-0.060	-0.076	-0.287	-0.018
	(0.294)	(0.304)	(0.318)	(0.311)	(0.324)	(0.324)
Low Moderate (>10, <=25 hours)	-0.099	0.077	0.134	-0.231	-0.330	-0.062
	(0.305)	(0.315)	(0.330)	(0.323)	(0.336)	(0.336)
High Moderate (>25, <=55 hours)	-0.291	-0.063	-0.050	-0.529	-0.425	-0.200
	(0.307)	(0.318)	(0.332)	(0.322)	(0.335)	(0.335)
High (>55 hours)	-0.237	-0.074	0.009	-0.097	-0.493	-0.442
	(0.320)	(0.331)	(0.346)	(0.335)	(0.349)	(0.349)
Interaction						
Female X Low (<10 hours)	0.297	0.234	0.109	0.252	0.484**	0.158
	(0.158)	(0.163)	(0.171)	(0.166)	(0.173)	(0.174)
Female X Low Moderate (>10, <=25 hours)	0.220	0.205	0.082	0.139	0.231	-0.009
	(0.156)	(0.161)	(0.169)	(0.164)	(0.170)	(0.171)
Female X High Moderate (>25, <=55 hours)	0.437**	0.317	0.141	0.368*	0.342	0.213
	(0.162)	(0.167)	(0.175)	(0.170)	(0.176)	(0.178)
Female X High (>55 hours)	0.674***	0.595***	0.327	0.429*	0.454*	0.387*
	(0.172)	(0.178)	(0.186)	(0.181)	(0.188)	(0.189)
N	1265	1266	1264	1257	1257	1247
Adjusted R^2	0.233	0.187	0.094	0.145	0.106	0.090

* p < 0.05; ** p < 0.01; *** p < 0.001
Note. Only main and interaction effects are presented. Interaction effects of being female with race are not presented. All covariates are included in each model. Covariates included in the analysis but not reported in this table include citizenship, year doctoral studies began, primary career goal at start of doctoral studies, interest in teaching at start of doctoral studies, institution, principal field of study, interest in becoming a faculty member, required teaching development, and completed doctorate.

Although Table 4 shows that participation in non-intensive TD activities, on average, contributed little to improving STEM early career academics' college teaching self-efficacy, Table 8 indicates this is not the case for women. Women participants in non-intensive TD activities benefited more than women non-participants in confidence in planning courses (0.401, $p < 0.01$ in M1), using teaching methods (0.324, $p < 0.05$ in M2), and interacting with students (0.519, $p < 0.01$ in M5).

Table 8
OLS Regression of Self-Efficacy on the Interaction of Female with TD Program Type

	Planning Courses (b/se)	Using Teaching Methods (b/se)	Creating Learning Environ-ments (b/se)	Assessing Student Learning (b/se)	Interacting with Students (b/se)	Mastering Subject Knowledge (b/se)
	M1	M2	M3	M4	M5	M6
Female	-0.557***	-0.366**	-0.102	-0.330**	-0.506***	-0.362**
	(0.117)	(0.120)	(0.126)	(0.123)	(0.128)	(0.129)
TD Type (ref: non-participants)						
Non-intensive	-0.441	-0.436	-0.102	-0.334	-0.434	-0.152
	(0.274)	(0.282)	(0.296)	(0.288)	(0.300)	(0.301)
Intensive and Others	0.162	0.375	0.344	-0.002	-0.209	0.091
	(0.313)	(0.323)	(0.338)	(0.332)	(0.345)	(0.346)
Formal Course and Others	-0.158	0.105	-0.020	-0.179	-0.340	-0.255
	(0.282)	(0.291)	(0.305)	(0.296)	(0.307)	(0.308)
Interaction						
Female X Non-intensive	0.401**	0.324*	0.155	0.286	0.519**	0.238
	(0.149)	(0.153)	(0.161)	(0.156)	(0.162)	(0.164)
Female X Intensive and Others	0.180	0.177	0.013	0.262	0.288	-0.082
	(0.165)	(0.170)	(0.178)	(0.173)	(0.180)	(0.181)
Female X Formal Course and Others	0.446**	0.331*	0.199	0.258	0.266	0.227
	(0.145)	(0.149)	(0.157)	(0.152)	(0.158)	(0.159)
N	1273	1274	1272	1265	1265	1256
Adjusted R^2	0.236	0.194	0.095	0.144	0.109	0.091

* $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$
Note. Only main and interaction effects are presented. Interaction effects of being female with race are not presented. All covariates are included in each model. Covariates included in the analysis but not reported in this table include citizenship, year doctoral studies began, primary career goal at start of doctoral studies, interest in teaching at start of doctoral studies, institution, principal field of study, interest in becoming a faculty member, required teaching development, and completed doctorate.

Discussion

At a time of increasing concern over the preparation of future STEM faculty for their role as college teachers, ours is the first study to use SCCT to examine the effects of doctoral TD on early career academics' teaching self-efficacy, which research shows is a strong predictor of successful teaching performance (Pajares, 1996; Ross, 2013; Sitzmann & Ely, 2011). The study examined whether early career academics' college teaching self-efficacy is affected by (1) any sort of participation in TD, (2) the degree of engagement in TD, and (3) the type of TD. Because self-efficacy sometimes varies by gender (Huang, 2013) and race/ethnicity (Byars-Winston, Estrada, Howard, Davis, & Zalapa, 2010; Lent et al., 2005), we also examined whether those characteristics interact with the effects of TD on college teaching self-efficacy.

Using regression analyses that accounted for key covariates, we found a significant and consistent connection between early career academics' participation in TD during their doctoral training—what SCCT calls learning experiences—and their college teaching self-efficacy. Moreover, the study revealed that participation in TD especially benefits women.

Effects of TD Participation on College Teaching Self-efficacy

The key driver of human agency, Bandura (1997) asserted, is one's beliefs about what one can accomplish. More so than knowledge, skill, and prior accomplishments, self-efficacy is a strong predictor of the degree of accomplishment that individuals eventually attain (Pajares, 1996; Sitzmann & Ely, 2011). Because teaching efficacy beliefs start to form during one's earliest teaching experiences and become more set over time (Woolfolk Hoy, 2004), it is important to help aspiring postsecondary faculty develop a strong sense of teaching self-efficacy during their doctoral training.

Our finding is significant, then, that participation in doctoral TD programs is closely linked with stronger college teaching self-efficacy across multiple dimensions of college teaching. TD participants were more confident than non-participants in their ability to handle various teaching activities, even after controlling for appropriate covariates such as gender, amount of teaching experience, and interest in becoming a faculty member. These findings are consistent with those of Prieto and Meyers (1999), who found that formal training in teaching for 176 psychology graduate teaching assistants increased their sense of self-efficacy for college teaching.

As for *why* TD is associated with a greater sense of college teaching self-efficacy, our survey data did not directly address this question. From the perspective of SCCT, we speculate that TD activities provide doctoral students with access to the four key sources of self-efficacy information (Bandura, 1997). For example, a course such as "Teaching in the STEM College Classroom" might afford doctoral students the opportunity to present a "teachable unit" (mastery experience), receive positive feedback (verbal persuasion), observe how classmates carried out their own teachable units (modeling or vicarious experience), and experience certain levels of anxiety as they present (emotional or physiological arousal). Determining whether and how TD programs do this, however, warrants further study.

Effects of the Degree of Engagement in TD on College Teaching Self-efficacy

Simply participating in TD is valuable in its own right, yet the amount of TD participation also matters. In her review of impact studies of K–12 teacher professional development, Desimone (2009) argued that "there is a research consensus on the main features of professional development that have been associated with changes in knowledge, practice, and, to a lesser extent, student achievement" (p. 183). Of the five critical features Desimone identified, one was duration, which refers not only to the number of hours spent in professional development but also the span of time over which the activity is spread. For our study, we focused on the number of hours that participants reported to have spent in TD—what we call degree of engagement.

In examining the degree of engagement in TD, which ranged from low (1–10 TD contact hours) to high (> 55 hours), we found that, in general, greater engagement led to greater gains in teaching self-efficacy for early career academics. As TD engagement increased (after accounting for covariates), participants were more confident about using teaching methods, planning courses, and at the highest level of engagement, assessing student learning. That TD programs focus on these three areas is not a surprise. Extensive study has found that three evidence-based approaches have the greatest impact on undergraduate student learning and achievement: engaging, student-centered instructional strategies; assessments requiring higher-order cognitive skills (as opposed to lower-level skills such as recollection or categorization); and course design that aligns learning objectives, assessment activities, and in-class instruction, sometimes referred to as backward design (Freeman et al., 2014; Holt, Young, Keetch, Larsen, & Mollner, 2015; Momsen, Long, Wyse, & Ebert-May, 2010; Wieman, 2014; Wood, 2009). Given the robust evidence supporting these three approaches, it is encouraging to find that TD programs have an appreciable effect on early career instructors' confidence in these areas. Whether TD programs intentionally focus on these approaches, however, requires further study.

Because confidence in assessing student learning is influenced at only a high level of TD engagement, we might also conclude that assessment is seen as an advanced dimension of teaching that is addressed after teaching methods and course design. If so, TD programs should consider addressing assessment sooner, for two reasons. First, recent research suggests that far too few introductory STEM courses use assessments that require higher-order cognitive skills (Momsen et al., 2010; Momsen et al., 2013). Second, sequencing assessment earlier among TD topics would better support efforts to teach doctoral students about backward design, of which assessment is an essential component.

Three of the six teaching components in our study—confidence in mastering subject knowledge, in interacting with students, and in creating learning environments—were not associated with TD at any level of engagement. The first, confidence in mastering subject knowledge, is not surprising because disciplinary subject knowledge is the teaching component that conventional doctoral training develops best (Walker et al., 2008). The lack of any relationship with the two other components suggests that TD activities may not provide doctoral students with authentic mastery experiences of working directly with undergraduate students. With fewer mastery experiences to build their self-efficacy beliefs, future faculty may struggle with dimensions that actively engage undergraduate students inside and outside the classroom.

Together, these findings about how TD engagement affects self-efficacy beliefs suggest that, in the push to improve undergraduate STEM education, TD programs may contribute to instructional and curricular reform by promoting evidence-based approaches. When it comes to broadening the participation of women and underrepresented minorities in undergraduate STEM, however, TD may not yet be playing the role it could. This possibility means TD providers might better explain to participants that effective undergraduate instruction includes not only knowing what to teach, how to teach, and how to assess whether students are learning, but also engaging students in inclusive learning environments.

Finally, we note that neither low nor low-moderate engagement had any effect on confidence in all six teaching dimensions. While low and low-moderate TD engagement may have positive effects on other outcomes not measured in this study, such as awareness of the complexity of teaching undergraduates and awareness of academic positions focused on undergraduate teaching, these levels of engagement add little or nothing to increase college teaching self-efficacy. Accounting for covariates, there was very little difference in college teaching self-efficacy measures of non-participants and those of participants with low or low-moderate engagement. This finding suggests at least 26 hours of engagement in TD are required to attain substantive gains in college teaching self-efficacy. Whether there is a threshold for TD effects on self-efficacy beliefs and other key outcomes must be studied more closely.

Effects of TD Type on College Teaching Self-efficacy

Drawing on previous research categorizing TD programs (Barger, Connolly, & Savoy, 2010), we studied the outcomes of three types of TD activities: non-intensive activities, intensive activities, and formal courses. The contribution that these three types of TD make to early career academics' college teaching self-efficacy resembles that of TD engagement—that is, the less-intensive activities have little to no effect on teaching self-efficacy, the middle-range intensive offerings contribute to self-efficacy about using specific teaching methods, and formal courses have a significant impact on early career academics' confidence in their ability to plan courses, assess student learning, and refine teaching methods. Our findings suggest participation in intensive TD activities or formal courses significantly contributes to improving STEM early career academics' college teaching self-efficacy in a way that non-intensive TD courses do not. Moreover, we report that TD engagement and TD type each has its own contribution to gains in college teaching self-efficacy. To explain why these three types of programs have differential effects on participants' college teaching self-efficacy, we believe formal courses may provide more access to the four sources of self-efficacy information (Bandura, 1997; Usher & Pajares, 2008) than non-intensive courses.

Interaction with Gender and Race/ethnicity

As noted, self-efficacy beliefs related to teaching and learning can vary by gender (Huang, 2013) and race/ethnicity (Byars-Winston et al., 2010; Lent et al., 2005). Although we found no interaction effects with race/ethnicity, the interaction effects with gender were generally significant for TD participation (four of six self-efficacy domains), high TD engagement (five of six self-efficacy domains), and TD type. These positive interaction effects can be interpreted in two ways: Women who participate in TD programs are more confident in college teaching than either (1) women who do not participate, or (2) men who participate in TD programs. The second interpretation, however, must be made with care given that women are already significantly less confident in college teaching than men (see Table 2) and that the main effects of being a woman are significantly negative and slightly bigger than interaction effects (see Table 6). Thus, this positive interaction effect does not necessarily indicate that women participants are more confident in college teaching than men participants. Rather, it may suggest that TD participation nearly cancels out women participants' initial, comparatively lower college teaching self-efficacy beliefs. Moreover, although our findings show that TD participation with more than 25 hours or in intensive/formal courses is generally associated with gains in college teaching self-efficacy, even fewer than 10 hours of participation in non-intensive TD activities are beneficial for women's college teaching self-efficacy.

The Contribution of Teaching Experience to College Teaching Self-efficacy

Although we intended to focus on the effects of TD on college teaching self-efficacy, our findings about the effects of actual teaching experience are worth mentioning. Our analyses show that teaching experience has a significant and unique effect on college teaching self-efficacy. Thus, combining TD with actual teaching experience has the greatest effect on college teaching self-efficacy.

Implications for Practice and Policy

Our study findings have important implications for the design and delivery of TD programs for doctoral students in STEM fields. One take-away from this study concerns the amount of time that doctoral students should spend in TD activities. Although it is useful to know that even a little bit of TD is associated with a greater sense of college teaching self-efficacy (compared with non-participants), our study shows that college teaching self-efficacy increases with higher levels of TD engagement. Because prior research shows a strong relationship between teachers' sense of self-efficacy, actual teaching performance, and desired student outcomes (Pajares, 1996; Ross, 1998, 2013), heightening college teaching self-efficacy among future faculty may be a critical pathway to improving STEM undergraduate teaching and learning. Thus, the findings from this study provide evidence that, for those doctoral students who hope to become faculty members and instructors, time in TD is generally well spent.

However, factors complicate early career academics allocating time to participate in TD. Skeptical advisors and peers may view doctoral students' time spent on TD as a waste of time or a way of avoiding one's research. In fact, faculty advisors in STEM fields sometimes stigmatize participation in TD by warning advisees away from such activity (Benbow, Byrd, & Connolly, 2011). Another factor that may thwart doctoral student involvement in TD is that research-intensive universities seldom offer such activities to doctoral students in a coordinated fashion. TD opportunities may be available within specific departments or colleges, or individual advisors may discuss teaching with their doctoral advisees on an ad hoc basis, but students often cannot count on a set of professional development opportunities being coordinated and offered on a systematic basis. As a result, organizing one's TD experiences is typically a "do-it-yourself" experience. The time it takes to find these opportunities and assess the potential return on an investment in them is a real cost to doctoral students. Because their time is valuable and guarded by their advisors, their families, and themselves, helping doctoral students to find TD programs and assess their potential value would make it much easier for them to participate in TD. Some graduate schools already organize TD opportunities into coherent frameworks, offering professional development at times that fit as easily as possible with doctoral students' schedules (e.g., evenings, weekends, or intensive weeks between semesters), and effectively market these opportunities, including what doctoral students—even those without faculty aspirations—may gain from participating. Michigan State University's PREP program is one example of such a systematic and developmentally focused professional development program (Vergara et al., 2013). In addition, several nationally organized projects offer comprehensive programs to help doctoral students prepare for their teaching responsibilities. Two examples are the Center for the Integration of Research, Teaching, and Learning (Austin et al., 2008; Pfund et al., 2012) and the Council of Graduate Schools' (2011) program on Preparing Future Faculty to Assess Student Learning. Programs that draw clear connections between the time spent in participation and the return on that investment are likely to gain greater buy-in from doctoral students and advisors alike.

A second implication of this study stems from the finding that TD is especially helpful to women doctoral students. This result is important because, in STEM fields, more women than men leave doctoral programs without degrees (Council of Graduate Schools, 2008), which has serious consequences for "sustaining our nation's scientific and technical prowess, building our domestic talent pool, and diversifying our professional, academic, and policy workforce" (Bernstein, 2011, p. 31). TD programs, as small-scale, "proximal" interventions (Bekki, Smith, Bernstein, & Harrison, 2013) may provide women doctoral students with not only greater pedagogical skill and understanding (Connolly & Lee, 2014) but also with opportunities for the kinds of community that are key to their doctoral persistence. Moreover, women as early career faculty typically face greater challenges than their male counterparts (Allan, 2011; Mason, Wolfinger, & Goulden, 2013; Trautvetter, 1999; Trower & Bleak, 2004). Individuals who use doctoral TD to prepare for teaching roles may experience less stress and more balance among their academic responsibilities (Benbow et al., 2011).

Third, this study shows that different types of TD lead to different outcomes. Providing an institutional map of TD programs could not only help doctoral students find programs faster, but also help them find programs with TD outcomes of interest, such as course design or assessing student learning. Such a map of opportunities could be organized by key features of TD programs such as duration, content, format, and selectivity (Barger et al., 2010; Connolly et al., 2010). In service to examining how different types of TD serve different outcomes, future research might explore the outcomes that early career academics do obtain from single or short-term TD opportunities (Zakrajsek, 2010).

Fourth, because combining TD with teaching experience has the greatest effect on college teaching self-efficacy, STEM doctoral students should, time permitting, be encouraged to participate in both types of activities during their doctoral training and especially in those that purposefully integrate theory and practice. Exploring how to combine teaching experience and TD for optimal learning and efficiency would be a useful direction for future research.

Fifth, our findings suggest that SCCT could prove to be a useful framework for designing, studying, and evaluating TD programs. Drawing upon SCCT research, Hackett (2013) has asserted that educational and training interventions ought to purposefully cultivate self-efficacy, which in turn promotes academic- and career-related interests, choices, performance, and satisfaction. Hackett identified five characteristics of programs that intentionally foster self-efficacy: such programs (1) provide opportunities for participants to build competencies; (2) provide access to the four sources of self-efficacy information; (3) address participants' unrealistic expectations about the outcomes of learning experiences; (4) minimize contextual factors that inhibit self-efficacy and enhance those that facilitate it; and (5) help participants clarify academic and career goals. These characteristics could easily be applied to TD programs for STEM doctoral students by, for example, helping doctoral students clarify their career goals. Given the relationship between self-efficacy and eventual performance, we suggest that TD administrators consider using these characteristics to design, deliver, and evaluate their programs.

Conclusion

Of the various ways to prepare doctoral students as future STEM academics, TD programs have become more popular at U.S. research universities. More data about their impact on participants and on undergraduate teaching may help these programs become commonplace. Understanding the effects of TD programs may help them gain wider acceptance. Key stakeholders should know these three findings about TD: a strong predictor of successful teaching performance is teaching self-efficacy; doctoral training, being a particularly influential stage in one's academic formation, is a crucial time to develop confidence in one's teaching abilities; and, as we found in this study, participation in doctoral TD contributes significantly to gains in college teaching self-efficacy.

Appendix A
Variable Descriptions

Variables	Description
Self-efficacy	Perceived college teaching self-efficacy, which measures respondents' self-reported confidence with college teaching. Each of six measured factors included five items for a total of 30 items. A five-point Likert scale was employed for each item: 1. Not at all confident, 2. Slightly confident, 3. Somewhat confident, 4. Very confident, 5. Extremely confident. These measures are from the Year 3 (2011) survey. *Italic items* below were dropped through exploratory factor analysis.
1. Planning courses	Five items include confidence in q1) setting learning goals, q2) selecting reading materials, q3) designing assignments, q4) planning class activities, and *q5) determining grading criteria*; Cronbach α = 0.906.
2. Using teaching methods	Five items include confidence in q6) using various teaching strategies, *q7) clearly communicating expectations to students*, q8) engaging students in learning, q9) providing students opportunities to practice skills, and q10) promoting student collaboration; Cronbach α = 0.904.
3. Creating learning environments	Five items include confidence in q11) encouraging students to ask questions, q12) encouraging students to express ideas, q13) encouraging participation from women and minorities, *q14) encouraging students to respect one another,* and *q15) managing student-instructor disagreements;* Cronbach α = 0.910.
4. Assessing student learning	Five items include confidence in *q16) developing assessments consistent with learning goals,* q17) accurately assessing students' knowledge, q18) grading assignments using criteria, q19) providing students constructive suggestions, and q20) providing students prompt feedback; Cronbach α = 0.878.
5. Interacting with students	Five items include confidence in q21) fostering students' independent thinking, q22) addressing sensitive issues in ways that help students to deal with them maturely, q23) fostering students' confidence in their ability to learn, *q24) working with problem students outside classroom,* and *q25) recognizing students who are not achieving to their fullest potential*; Cronbach α = 0.876.
6. Mastering subject knowledge	Five items include confidence in q26) providing students an overview of discipline, q27) demonstrating passion for the material being taught, q28) staying current in subject knowledge, q29) helping students understand the relevance of learning, and q30) enriching teaching with research; Cronbach α = 0.875.
TD Participation	The teaching development (TD) program refers to activities to enhance pedagogical knowledge for doctoral students through seminars, courses, workshops, symposiums, and discussion groups. TD participation was measured as "Yes" or "No." Although each respondent was asked about TD program participation for each specific program in Year 1 survey, the Year 3 survey was asked whether the respondent participated in any TD program.
TD Engagement	Based on total TD hours, four levels of TD engagement were determined: 1. Non-participant (no TD or 0 TD contact hours), 2. Low engagement (1 ~ 10 TD contact hours), 3. Low-Moderate engagement (11 ~ 25 TD contact hours), 4. High-Moderate engagement (26 ~ 55 TD contact hours), 5. High engagement (more than 55 TD contact hours).
Amount of Teaching Experience	The amount of college teaching experience was determined as four levels based on total semesters or quarters of diverse teaching related activities (i.e., teaching assistant, lab assistant, guest lecturer, instructor, research mentor) that respondents experienced during graduate education and/or postdoctoral training: 1. Low (<= 1 semester), 2. Low-Moderate (> 1, <= 3 semesters), 3. High-Moderate (> 3, <= 6 semesters), and 4. High (> 6 semesters). The median was around 4.5 quarters or 3 semesters (mean: 4.5 semesters or 6.7 quarters). Quarter units were converted to semester units based on 1.5-to-1 quarter-semester ratio.

Appendix A, cont.

Gender	Gender was measured as male and female.
Race/ethnicity	Originally, race/ethnicity was measured as American Indian or Alaska native, Asian, Black or African American, Hispanic or Latino, more than one race or ethnicity, native Hawaiian or other Pacific Islander, and White. However, because of the low proportion of some minority groups, these seven categories were merged into three categories: White, Asian, and other minority.
Citizenship	Citizenship status at the beginning of doctoral studies was measured as U.S. citizen, U.S. permanent resident, and other, and finally merged into U.S. citizen and permanent resident, and other.
Year Doctoral Studies Began	This variable measures a year that respondent began their doctoral studies.
Primary Career Goal at Start of Doctoral Studies	Primary career goal at the beginning of doctoral program was measured as faculty career at a college or university; research career in government, industry or business; start your own business; undecided; and other goal. These categories merged into faculty career, research career, and other.
Interest in Teaching at Start of Doctoral Studies	Interest in teaching undergraduate students at start of doctoral studies was measured using a five-point Likert scale: 1. Not at all interested, 2. Slightly interested, 3. Somewhat interested, 4. Very interested, and 5. Extremely interested.
Principal Field of Study	Principal fields of study were measured as 1. Engineering; 2. Physical Sciences; 3. Earth, Atmospheric, and Ocean Sciences; 4. Mathematical Sciences; 5. Computer Sciences; 6. Agricultural Sciences; 7. Biological Sciences; 8. Psychology; 9. Social Sciences; 10. Health Fields; and 11. Other.
Institutions	Participants' doctorate-granting institutions include Arizona State University, the University of Washington–Seattle, and the University of Wisconsin–Madison.
Interest in Becoming a Faculty Member	Whether respondents considered applying for a faculty job in the future when they were late-stage doctoral students. Measured at Year 1 (2009) as Yes, No, and Not Sure.
TD Participation is Required	Whether advisor, department, graduate school, etc., required respondent to participate in TD. Measured at Year 1 survey as Yes and No.
Completed Doctorate	Respondents' academic status at Year3 (2011) survey, measured as 1. Currently enrolled in a Ph.D. program, 2. Graduated from a Ph.D. program with a Ph.D., and 3. Previously enrolled in a Ph.D. program but no longer pursuing a Ph.D.

References

Allan, E. J. (2011). Women's status in higher education: Equity matters. *ASHE Higher Education Report, 37*(1).

Anderson, W. A., Banerjee, U., Drennan, C. L., Elgin, S. C. R., Epstein, I. R., Handelsman, J., … Warner, I. M. (2011). Changing the culture of science education at research universities. *Science, 331,* 152–153.

Association of American Universities. (2014). *Framework for systemic change in undergraduate STEM teaching and learning.* Retrieved from https://stemedhub.org/groups/aau/File:AAU_Framework_030114.pdf

Austin, A. E. (2010). Reform efforts in STEM doctoral education: Strengthening preparation for scholarly careers. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. 25, pp. 91–128). Dordrecht, Netherlands: Springer.

Austin, A. E., Connolly, M. R., & Colbeck, C. L. (2008). Strategies for preparing integrated faculty: The Center for the Integration of Research, Teaching, and Learning. In C. L. Colbeck, K. O'Meara, & A. E. Austin (Eds.), *Educating integrated professionals: Theory and practice on preparation for the professoriate* (New Directions for Teaching and Learning No. 113, pp. 69–81). San Francisco, CA: Wiley.

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: Handbook of theory and research* (Vol. 21, pp. 397–456). New York, NY: Springer.

Austin, A. E., & Wulff, D. H. (2004). The challenge to prepare the next generation of faculty. In D. H. Wulff & A. E. Austin (Eds.), *Paths to the professoriate: Strategies for enriching the preparation of future faculty* (pp. 3–16). San Francisco, CA: Jossey-Bass.

Bandura, A. (1986). *Social foundations of thought & action: A social cognitive theory*. Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W. H. Freeman.

Bandura, A. (2005). The evolution of social cognitive theory. In K. G. Smith & M. A. Hitt (Eds.), *Great minds in management* (pp. 9–35). Oxford, England: Oxford University Press.

Bandura, A. (2006). Guide for constructing self-efficacy scales. In F. Pajares & T. Urdan (Eds.), *Self-efficacy beliefs of adolescents* (Vol. 5, pp. 307–337). Greenwich, CT: Information Age.

Barger, S. S., Connolly, M. R., & Savoy, J. N. (2010, April). *A model of highly effective teaching-focused professional development programs*. Paper presented at the meeting of the American Educational Research Association, Denver, CO.

Baulch, B., & Quisumbing, A. (2011). Testing and adjusting for attrition in household panel data. *CPRC Toolkit Note*. Retrieved from Chronic Poverty Research Centre website: http://www.chronicpoverty.org/uploads/publication_files/CPRC TOOLKIT NOTE ON ATTRITION(revised).pdf

Becher, T., & Trowler, P. (2001). *Academic tribes and territories: Intellectual enquiry and the cultures of discipline*. Philadelphia, PA: Open University.

Bekki, J. M., Smith, M. L., Bernstein, B. L., & Harrison, C. (2013). Effects of an online personal resilience training program for women in STEM doctoral programs. *Journal of Women and Minorities in Science and Engineering, 19*, 17–35.

Benbow, R. J., Byrd, D., & Connolly, M. R. (2011). *The Wisconsin longitudinal study of doctoral and postdoctoral teaching development: Key findings.* Madison, WI: Wisconsin Center for Education Research, University of Wisconsin–Madison.

Bernstein, B. L. (2011). Managing barriers and building supports in science and engineering doctoral programs: Conceptual underpinnings for a new online training program for women. *Journal of Women and Minorities in Science and Engineering, 17*, 29–50.

Betz, N. E., & Hackett, G. (2006). Career self-efficacy theory: Back to the future. *Journal of Career Assessment, 14*(1), 3–11. doi: 10.1177/1069072705281347

Braxton, J. M., & Hargens, L. L. (1996). Variation among academic disciplines: Analytical frameworks and research. In J. C. Smart (Ed.), *Higher education: Handbook of theory and practice* (Vol. 11, pp. 1–46). New York, NY: Agathon.

Byars-Winston, A., Estrada, Y., Howard, C., Davis, D., & Zalapa, J. (2010). Influence of social cognitive and ethnic variables on academic goals of underrepresented students in science and engineering: A multiple-groups analysis. *Journal of Counseling Psychology, 57*(2), 205–218. doi: 10.1037/a0018608

Chism, N. V. N. (2007). *Peer review of teaching: A sourcebook* (2nd ed.). Bolton, MA: Anker.

Coalition for Reform of Undergraduate STEM Education. (2014). *Achieving systemic change: A sourcebook for advancing and funding undergraduate STEM education.* Retrieved from http://www.aacu.org/sites/default/files/files/publications/E-PKALSourcebook.pdf

Connolly, M. R. (2012). *Postsecondary employment patterns of STEM doctorate recipients.* [Unpublished analysis of data from public version of Survey of Doctorate Recipients.]

Connolly, M. R., & Lee, Y.-G. (2014, April). *The effects of doctoral teaching development programs on STEM doctoral students' college teaching competency.* Paper presented at the meeting of the American Educational Research Association, Philadelphia, PA.

Connolly, M. R., Savoy, J. N., & Barger, S. S. (2010, May). *Future-faculty professional development programs for doctoral students in science, technology, engineering, and mathematics: An exploratory classification scheme.* Paper presented at the meeting of the American Educational Research Association, Denver, CO.

Council of Graduate Schools. (2008). *Ph.D. completion and attrition: Analysis of baseline demographic data from the Ph.D. completion project*. Washington, D.C.: Author.

Council of Graduate Schools. (2011). *Preparing future faculty to assess student learning.* Washington, D.C.: Author.

Cuddeback, G., Wilson, E., Orme, J. G., & Combs-Orme, T. (2004). Detecting and statistically correcting sample selection bias. *Journal of Social Service Research, 30*(3), 19–33. doi: 10.1300/J079v30n03_02

DeNeef, A. L. (2002). *The Preparing Future Faculty program: What difference does it make?* Washington, D.C.: Association of American Colleges and Universities.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher, 38*(3), 181–199. doi: 10.3102/0013189x08331140

Feldon, D. F., Peugh, J., Timmerman, B. E., Maher, M. A., Hurst, M., Strickland, D., … Stiegelmeyer, C. (2011). Graduate students' teaching experiences improve their methodological research skills. *Science, 333*, 1037–1039. doi: 10.1126/science.1204109

Fitzgerald, J., Gottschalk, P., & Moffitt, R. (1998). The impact of attrition in the Panel Study of Income Dynamics on intergenerational analysis. *Journal of Human Resources, 33*, 300–344.

Foster, E. M., & Fang, G. Y. (2004). Alternative methods for handling attrition: An illustration using data from the fast track evaluation. *Evaluation Review, 28*(5), 434–464. doi: 10.1177/0193841X04264662

Freeman, S., Eddy, S. L., McDonough, M., Smith, M. K., Okoroafor, N., Jordt, H., & Wenderoth, M. P. (2014). Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences, 111*(23), 8410–8415.

Gaff, J. G., & Lambert, L. M. (1996, July/August). Socializing future faculty to the values of undergraduate education. *Change, 28*, 38–45.

Gardner, S. K., & Mendoza, P. (Eds.). (2010). *On becoming a scholar: Socialization and development in doctoral education*. Sterling, VA: Stylus.

Golde, C. M., & Dore, T. M. (2001). *At cross purposes: What the experiences of today's doctoral students reveal about doctoral education*. Philadelphia, PA: The Pew Charitable Trusts.

Goldsmith, S. S., Haviland, D., Dailey, K., & Wiley, A. (2004). *Preparing Future Faculty Initiative: Final evaluation report*. Washington, D.C.: WestEd Evaluation Research Program and Abt Associates.

Gonyea, R. M. (2005). *Self-reported data in institutional research.* In P. D. Umbach (Ed.), *Survey research: Emerging issues* (New Directions for Institutional Research No. 127, pp. 73–89). San Francisco, CA: Jossey-Bass.

Hackett, G. (2013). Social cognitive career theory of career choice [PowerPoint slides]. Downloaded from http://www.umkc.edu/provost/student-retention/documents/Social%20Cognitive%20Career%20Theory%20February%202013.ppt

Handelsman, J., Miller, S., & Pfund, C. (2006). *Scientific teaching*. New York, NY: W. H. Freeman.

Hativa, N. (2014). *Student ratings of instruction: A practical approach to designing, operating, and reporting* (2nd ed.). United States: Oron.

Hershock, C., Groscurth, C. R., & Milkova, S. (2011). Approaches to preparing future faculty for teaching. In C. E. Cook & M. Kaplan (Eds.), *Advancing the culture of teaching on campus* (pp. 97–117). Sterling, VA: Stylus.

Holt, E. A., Young, C., Keetch, J., Larsen, S., & Mollner, B. (2015). The greatest learning return on your pedagogical investment: Alignment, assessment or in-class instruction? *PLoS ONE, 10*(9): e0137446. doi: 10.1371/journal.pone.0137446

Hopwood, N., & Stocks, C. (2008). Teaching development for doctoral students: What can we learn from activity theory? *International Journal for Academic Development, 13*, 187–198. doi: 10.1080/13601440802242358

Huang, C. J. (2013). Gender differences in academic self-efficacy: A meta-analysis. *European Journal of Psychology of Education, 28*, 1–35. doi: 10.1007/s10212-011-0097-y

Kalish, A., Robinson, S. S., Border, L., Chandler, E., Connolly, M. R., Eaton, L. J., … von Hoene, L. (2011, Spring). Steps toward a framework for an intended curriculum for graduate and professional students: How we talk about what we do. *Studies in Graduate and Professional Student Development, 14,* 163–173.

Klassen, R. M., & Usher, E. L. (2010). Self-efficacy in educational settings: Recent research and emerging directions. In T. C. Urdan & S. A. Karabenick (Eds.), *The decade ahead: Theoretical perspectives on motivation and achievement* (Vol. 16, pp. 1–33). Bingley: Emerald.

Kober, N. (2015). *Reaching students: What research says about effective instruction in undergraduate science and engineering*. Washington, D.C.: National Academies Press.

Lent, R. W., & Brown, S. D. (2006). On conceptualizing and assessing social cognitive constructs in career research: A measurement guide. *Journal of Career Assessment, 14*, 12–35. doi: 10.1177/1069072705281364

Lent, R. W., Brown, S. D., & Hackett, G. (1994). Toward a unifying social cognitive career theory of career and academic interest, choice, and performance. *Journal of Vocational Behavior, 45*, 79–122. doi: 10.1006/Jvbe.1994.1027

Lent, R. W., Brown, S. D., & Hackett, G. (2000). Contextual supports and barriers to career choice: A social cognitive analysis. *Journal of Counseling Psychology, 47*, 36–49. doi: 10.1037//0022-0167.47.1.36

Lent, R. W., Brown, S. D., Sheu, H. B., Schmidt, J., Brenner, B. R., Gloster, C. S., … Treistman, D. (2005). Social cognitive predictors of academic interests and goals in engineering: Utility for women and students at historically black universities. *Journal of Counseling Psychology, 52*, 84–92. doi: 10.1037/0022-0167.52.1.84

Lindley, L. D. (2006). The paradox of self-efficacy: Research with diverse populations. *Journal of Career Assessment, 14*, 143–160. doi:10.1177/1069072705281371

Lowman, J. (1995). *Mastering the techniques of teaching* (2nd ed.). San Francisco, CA: Jossey-Bass.

Major, C. H., & Dolly, J. P. (2003). The importance of graduate program experiences to faculty self-efficacy for academic tasks. *The Journal of Faculty Development, 19*, 89–100.

Mason, M. A., Wolfinger, N. H., & Goulden, M. (2013). *Do babies matter? Gender and family in the ivory tower*. New Brunswick, NJ: Rutgers University Press.

Miller, R. B., & Hollist, C. S. (2007). Attrition bias. In N. Salkind (Ed.), *Encyclopedia of measurement and statistics* (Vol. 1, pp. 57–60). Thousand Oaks, CA: Sage.

Momsen, J. L., Long, T. M., Wyse, S. A., & Ebert-May, D. (2010). Just the facts? Introductory undergraduate biology courses focus on low-level cognitive skills. *CBE—Life Sciences Education, 9*(4), 435–440.

Momsen, J. L., Offerdahl, E., Kryjevskaia, M., Montplaisar, L., Anderson, E., & Grosz, N. (2013). Using assessments to investigate and compare the nature of learning in undergraduate science courses. *CBE—Life Sciences Education, 12*(2), 239–249.

Murray, H. G. (2007). Low-inference teaching behaviors and college teaching effectiveness: Recent developments and controversies. In R. P. Perry & J. C. Smart (Eds.), *The scholarship of teaching and learning in higher education: An evidence-based perspective* (pp. 145–200). Dordrecht, Netherlands: Springer.

National Research Council. (2003). *Evaluating and improving undergraduate teaching in science, technology, engineering, and mathematics*. Washington, D.C.: National Academies Press.

National Research Council. (2012). *Discipline-based education research: Understanding and improving learning in undergraduate science and engineering*. Washington, D.C.: National Academies Press.

Pajares, F. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research, 66*, 543–578. doi: 10.3102/00346543066004543

Palmer, M. S. (2011, Spring). Graduate student professional development: A decade after calls for national reform. *Studies in Graduate and Professional Student Development, 14,* 1–17.

Pfund, C., Mathieu, R., Austin, A., Connolly, M. R., Manske, B., & Moore, K. (2012, December 10). Advancing STEM undergraduate learning: Preparing the nation's future faculty. *Change: The Magazine of Higher Learning, 44,* 64–72.

President's Council of Advisors on Science and Technology. (2012). *Engage to excel: Producing one million additional college graduates with degrees in science, technology, engineering, and mathematics.* Retrieved from http://www.whitehouse.gov/sites/default/files/microsites/ostp/pcast-engage-to-excel-final_feb.pdf

Prieto, L. R., & Altmaier, E. M. (1994). The relationship of prior training and previous teaching experience to self-efficacy among graduate teaching assistants. *Research in Higher Education, 35*, 481–497. doi: 10.1007/Bf02496384

Prieto, L. R., & Meyers, S. A. (1999). Effects of training and supervision on the self-efficacy of psychology graduate teaching assistants. *Teaching of Psychology, 26*, 264–266. doi: 10.1207/S15328023top260404

Ross, J. A. (1998). The antecedents and consequences of teacher efficacy. In J. Brophy (Ed.), *Advances in research on teaching* (Vol. 7, pp. 49–74). Greenwich, CT.

Ross, J. A. (2013). Teacher efficacy. In J. Hattie & E. M. Anderman (Eds.), *International guide to student achievement* (pp. 266–267). New York, NY: Routledge.

Santiago, A. M., & Einarson, M. K. (1998). Background characteristics as predictors of academic self-confidence and academic self-efficacy among graduate science and engineering students. *Research in Higher Education, 39*, 163–198. doi: 10.1023/A:1018716731516

Schönwetter, D. J., & Ellis, D. (2010). Taking stock: Contemplating North American graduate student professional development programs and developers. In J. E. Miller & J. E. Groccia (Eds.), *To improve the academy* (Vol. 29, pp. 3–17). San Francisco, CA: Jossey-Bass.

Seymour, E., & Hewitt, N. M. (1997). *Talking about leaving: Why undergraduates leave the sciences*. Boulder, CO: Westview.

Sitzmann, T., & Ely, K. (2011). A meta-analysis of self-regulated learning in work-related training and educational attainment: What we know and where we need to go. *Psychological Bulletin, 137*(3), 421–442. doi: 10.1037/a0022777

Theall, M., Mullinix, B., & Arreola, R. A. (2009). Promoting dialogue and action on meta-professional skills, roles, and responsiblities. In L. B. Nilson & J. E. Miller (Eds.), *To improve the academy* (Vol. 28, pp. 115–138). San Francisco, CA: Jossey-Bass.

Trautvetter, L. C. (1999). Experiences of women, experiences of men. In R. J. Menges (Ed.), *Faculty in new jobs: A guide to settling in, becoming established, and building institutional support* (pp. 59–87). San Francisco, CA: Jossey-Bass.

Trower, C. A., & Bleak, J. L. (2004). Study of new scholars. *Gender: Statistical Report [Universities]*. Cambridge, MA: Harvard Graduate School of Education.

Tschannen-Moran, M., & Woolfolk Hoy, A. (2001). Teacher efficacy: Capturing an elusive construct. *Teaching and Teacher Education, 17*, 783–805. doi: 10.1016/S0742-051x(01)00036-1

Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research, 78*, 751–796. doi: 10.3102/0034654308321456

Vergara, C. E., Urban-Lurain, M., Campa, H., Cheruvelil, K. S., Ebert-May, D., Fata-Hartley, C., & Johnston, K. (2013). FAST-Future Academic Scholars in Teaching: A high-engagement development program for future STEM faculty. *Innovative Higher Education, 39*(2), 1–15. doi: 10.1007/s10755-013-9265-0

Walker, G. E., Golde, C. M., Jones, L., Bueschel, A. C., & Hutchings, P. (2008). *The formation of scholars: Rethinking doctoral education for the twenty-first century*. San Francisco, CA: Jossey-Bass.

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

Weisbuch, R. (2004). Toward a responsive Ph.D.: New partnerships, paradigms, practices, and people. In D. H. Wulff & A. E. Austin (Eds.), *Paths to the professoriate: Strategies for enriching the preparation of future faculty* (pp. 217–235). San Francisco, CA: Jossey-Bass.

Widener, A. (2011, August). Time to teach: Is there time for teaching and research in a postdoc experience? *HHMI Bulletin, 24*(3), 24–29, 48.

Wieman, C. E. (2014). Large-scale comparison of science teaching methods sends clear message. *Proceedings of the National Academy of Sciences, 111*(23), 8319–8320.

Wood, W. B. (2009). Innovations in teaching undergraduate biology and why we need them. *Annual Review of Cell and Developmental Biology, 25*, 93–112.

Woolfolk Hoy, A. (2004). Self-efficacy in college teaching. *Essays on teaching excellence: Toward the best in the academy* (Vol. 15, pp. 8–11). Fort Collins, CO: The POD Network.

Wulff, D. H., & Austin, A. E. (Eds.). (2004). *Paths to the professoriate: Strategies for enriching the preparation of future faculty*. San Francisco, CA: Jossey-Bass.

Wurgler, E., VanHeuvelen, J. S., Rohrman, S., Loehr, A., & Grace, M. K. (2013). The perceived benefits of a preparing future faculty program and its effect on job satisfaction, confidence, and competence. *Teaching Sociology, 20*, 1–11. doi: 10.1177/0092055x13507782

Zakrajsek, T. D. (2010). Important skills and knowledge. In K. H. Gillespie, D. R. Roberston, & Associates (Eds.), *A guide to faculty development* (2nd ed., pp. 83–98). San Francisco, CA: Jossey-Bass.