2016 NCTM RESEARCH CONFERENCE
April 11–13 • San Francisco

Paper Title: Gender and Doctoral Mathematics: Impactful Factors for Female Students' Success
Author(s): Emily Miller
Session Title: Gender and Doctoral Mathematics: Impactful Factors for Female Students' Success
Session Type: Interactive Paper Session
Presentation Date: Wednesday, April 13, 2016: 1:15 PM - 2:30 PM
Presentation Location: San Francisco, California

Women have long been underrepresented in mathematics. While progress has been made at the undergraduate level, with 44% of bachelor's degrees in mathematics being awarded to women in 2007 (Hill, Corbett, & St. Rose, 2010), similar improvement at the doctoral level has not occurred. For example, in 2013, only 27% of doctoral recipients were female (Vélez, Maxwell, & Rose, 2014). This imbalance is even further evident when one looks at gender ratios for mathematics faculty members at doctorate-granting institutions. For example, in 2006, only 12% of mathematics faculty at doctorate-granting institutions were female (Phipps, Maxwell, & Rose, 2007).

The underrepresentation of female mathematics doctorate students is well recognized, but the solution is not apparent. According to McAlpine and Norton (2006), retention and attrition are influenced by the "interaction of a constellation of dynamic factors" (p. 5). A multitude of factors, emerging from quantitative and qualitative research studies and theoretical arguments, have been hypothesized as contributing to the retention or attrition of female mathematics doctoral students. Proposed factors include those pertaining to students' background characteristics and external commitments (e.g., Becker, 1984; Herzig, 2010; Preckel, Goetz, Pekrun, & Kleine, 2008); relationships with their advisor, other professors, and students in their department (e.g., Bair & Haworth, 2004; Baird, 1993; Tinto, 1993); the quality and culture of the courses they take in the doctoral program (e.g., Hall & Sandler, 1982; Herzig, 2004b); the support they receive through assistantships or fellowships (e.g., Ehrenberg & Mavros, 1992; Ethington & Pisani, 1993); the presence or absence of other female students or role models (e.g., Blickenstaff, 2005; Robst, Keil, & Russo, 1998; Schroeder & Mynatt, 1993); and perceptions of biases or discrimination against female students (e.g., Berg & Ferber, 1983; Spencer, Steele, &

Quinn, 1999; Sue, 2010). For a more detailed description of the literature supporting the constructs included in this study, see Miller (2015b).

Most prior research concerned with the underrepresentation of women in advanced mathematics has focused on identifying factors that impact the *attrition* of women from mathematics doctoral programs (e.g., Herzig, 2002; Herzig, 2004a; Herzig, 2004b). Because the sample sizes in these studies have typically been small, it is unclear how generalizable these factors are in accounting for women's attrition. Moreover, additional factors may be at play in influencing retention, beyond those that contribute to attrition. Thus, it is important to identify factors associated with the *success* of women in mathematics doctoral programs. However, very few studies have examined the problem from this perspective.

Many constructs have been used to characterize the outcome of a student's experience in a doctoral program. Constructs such as *retention, completion*, and *persistence* have been used to represent doctoral degree attainment (e.g., Ampaw & Jaeger, 2011; Nerad & Miller, 1996; Tinto, 1993); *attrition* has been used to describe the trajectories of students who discontinued their studies (e.g., Bair & Haworth, 2004; Golde, 1998). These constructs are essentially binary in nature: either a doctoral student completes her program, or she does not. Although binary constructs are more easily defined and measured, the focus is then on the end result, and not on the confluence of decisions and experiences that contribute to that end result. In this way, "attrition has been conceptualized as a solitary event, rather than as the consequence of a dynamic process" (Kerlin, 1997, p. 21). There is more to be learned about a student's experience than simply their receipt of a diploma. Even for those who do obtain doctorates, did they *thrive* or did they merely *survive* until graduation?

Research Questions

Identifying those factors that play a significant role in doctoral program success is a necessary first step in addressing the gender gap in mathematics doctoral study. This study aims to identify factors that have the strongest association with student success, as reported by female and male graduates of mathematics doctoral programs currently employed at post-secondary institutions, using a large-sample quantitative survey methodology. Male participants are used as a comparison group for the responses from female participants. In particular, this study was designed to investigate the following research questions:

1. What factors have the strongest influence on mathematics doctoral program success for women who have earned their doctorate?
2. How do the factors influencing mathematics doctoral program success compare for men and women who have earned their doctorate?

A primary goal of this research is to identify those factors that are critical to women's success in obtaining a Ph.D. in mathematics so that future research can investigate how these factors interact to influence doctoral program success. Moreover, identification of critical factors can inform the redesign of doctoral programs to better facilitate women's success. Data collected about these factors will describe the experiences of successful female doctoral students, contrast these experiences with the experiences of successful male doctoral students, and evaluate the importance of each factor in participants' success in obtaining a doctorate in mathematics. In contrast to previous studies using a binary construct, this study draws on a more descriptive outcome measure, doctoral program *success*, to attempt to capture the complex nature of this dynamic process. In addition, data collected will be used to conduct group comparisons of critical factors for women who did and did not complete their doctoral programs.

Methods

This section begins with a description of the survey instrument used in the study. Then, the sample selection and resulting sample demographics are discussed, followed by a description of the steps taken to prepare the data for analysis using partial least squares structural equation modeling (PLS-SEM). This preparation included exploratory factor analysis to determine underlying latent constructs and multiple imputation for handling missing data. The section concludes by describing the analyses conducted to investigate each research question.

Survey Instrument

The instrument used to collect data was an electronic survey consisting of three sections. The first section contained three items designed to ensure that participants satisfied the selection criteria. First, participants were asked if they had obtained a doctorate in mathematics or applied mathematics. If the response was "Yes," the participant continued with the survey. If the response was "No," then participants were asked if they had ever enrolled in a doctoral program with the intent to earn such a degree. If the response was again "No," the survey ended. Finally, participants were asked if they were currently enrolled in a mathematics doctoral program. If the response was "Yes," the survey ended. Participants who had enrolled in a doctoral program in pure or applied mathematics, but did not complete their degree, were directed to a slightly different version of the survey to collect information about their experiences with attrition from their doctoral program. This version of the survey was conceptually identical, but included a "Not Applicable" option for some items. This allowed participants to distinguish factors they did not experience because of lack of opportunity in the program from factors they did not experience because of their departure from the program.

The second section of the survey asked participants to indicate their level of agreement with 62 statements using a five-point Likert scale, with *1 = Strongly Disagree*, *2 = Disagree*, *3 = Neither Agree, Nor Disagree*, *4 = Agree* and *5 = Strongly Agree*. The statements were designed to represent the key factors identified from a systematic review of the literature and were organized into ten blocks based on hypothesized themes (Miller, 2015b). The ten blocks evaluated student attributes, prior educational experiences, external (non-academic) commitments, institutional support experiences, interactions with professors, interactions with peers, academic relationship with advisor, programmatic structure, quality of coursework, and gender ratios within the program. An additional block evaluated the outcome construct, doctoral program success, and contained five items. Approximately half of the 62 items in this section were worded in the negative; data for these questions were reverse coded to improve the reliability of the items. Items worded in the positive were assigned Likert values such that the higher end of the scale aligned with greater success. For items worded in the negative, the lower end of the scale aligned with greater success. Since the participants were asked to recall experiences that may have occurred decades prior, the items were ordered chronologically to aid in memory recall. The questions within each block were randomized so that participants' responses were less susceptible to order effects.

In the third and final section of the survey, participants were asked demographic questions, such as their gender, their current job title, and the highest mathematics degree their employing institution grants. This section concluded with items about aspects of the participants' training in mathematics. For example, participants were asked to identify the length of time spent in their doctoral program, the gender of their doctoral advisor, and when they earned their doctorate. For the full set of demographic questions, see the survey in Appendix A.

Ten mathematics education doctoral students and faculty members piloted the survey. These pilot participants did not have doctorates in pure or applied mathematics and so did not detract from the desired sample. However, most held advanced degrees in mathematics and were, therefore, able to provide knowledgeable feedback on the survey. The pilot process had three aims: (1) to garner feedback on the content validity and clarity of the survey items; (2) to estimate the time required for participants to complete the survey; and (3) to ensure that the online data collection proceeds as planned. Pilot participants reported taking approximately 15 minutes to complete the survey, as expected, and no issues with the online data collection process were discovered. Minor feedback on the wording of survey items was received and incorporated to improve the survey. The survey instrument can be found in Appendix A.

Sample

The target population for this study was mathematics faculty members employed at tertiary institutions in the United States, who either hold a doctorate in mathematics (pure or applied) or who had at one time enrolled in a doctoral program with the intent to earn such a degree. Those currently enrolled in mathematics doctoral programs were excluded from completing the survey. The experiences of these participants are incomplete, and therefore, are not comparable to the target population.

Although graduates of mathematics doctoral programs have several employment options available to them, most assume positions in academia. For instance, in the 2012–2013 academic year, 65.8% of doctoral recipients in the mathematical sciences accepted academic positions (Vélez, Maxwell, & Rose, 2014). Since these data include the career paths of statistics and biostatistics graduates, fields in which over 70% of graduates accept positions in industry or government, this percentage is likely higher when limited to those earning mathematics doctorates. Moreover, less than six percent of all new doctorate recipients in mathematics reported being unemployed, and only five percent of recent female graduates reported unemployment. Therefore, although the sample selected for this study may not be generalizable to all graduates of mathematics doctoral programs, specifically those unemployed or employed in industry or governmental positions, it is generalizable to a large majority of doctorate recipients in mathematics. Since unemployed graduates or graduates employed in non-academic positions would be nearly impossible to recruit in a systematic manner, participation in this study was limited to participants employed in academia.

To obtain a sample from this population, a sampling frame was used from the report, "Statistical Abstract of Undergraduate Programs in the Mathematical Sciences in the United States" (Blair, Kirkman, & Maxwell, 2013). The sampling frame contains an exhaustive list of two- and four-year colleges and universities granting degrees in mathematics in the United States, separated into four strata by institution type: associate's colleges, baccalaureate colleges, master's colleges/universities, and doctoral/research universities. Any institutions not listed in one of the preceding four categories (e.g., tribal colleges) were not included in the institution sample. Table 1 presents the distribution of institutions of the four types.

Table 1
Number (Percent) of Institutions by Type in the Sampling Frame

	Number (percent) of institutions
Associate's colleges	1031 (42.78%)
Baccalaureate colleges	553 (22.95%)
Master's colleges/universities	565 (23.44%)
Doctoral/research universities	261 (10.83%)
Total	2410 (100.00%)

Based upon the distribution of institutions of each of the four types, a corresponding proportion of faculty members within each stratum were sampled. For instance, approximately

11% of the total number of institutions in the sample fall into the doctoral/research university category. Therefore, a corresponding percentage of invitations for the survey were sent to faculty at that type of institution. To achieve this, institutions in each stratum were randomly ordered. According to the random ordering, all available participants at the highest listed institutions were selected for the sample, until the required sample size was obtained. Contact information for mathematics faculty employed at the selected institutions was collected through an Internet search.

Power analysis. There are different recommendations regarding the necessary sample size for PLS-SEM analyses to be sufficiently powered. According to Cohen (1992, as cited in Hair, Jr., Hult, Ringle, & Sarstedt, 2014), the sample size is dependent upon the statistical power, the significance level, the minimum value of R^2 desired, and the maximum number of indicators pointing to a single construct in the path diagram of the structural equation model. Based on these considerations, and with the hypothesized latent construct structure created by identifying themes from the literature, each analysis would require at least 166 participants, or 332 participants overall.[1] Others recommend using the results of *a priori* power analyses for multiple regression, which would recommend a sample size of at least 64 per analysis, or 128 overall, as computed with G*Power power analysis software[2] (Faul, Erdfelder, Lang, & Buchner, 2007; Hair, Sarstedt, Pieper, & Ringle, 2012). Still others advocate using the "ten times rule" and obtaining 10 times the maximum number of indicators leading to any one

[1] Assuming a power of 80 percent, with a significance level of .05, a minimum R^2 of .10, and a maximum of 7 items defining a single construct.

[2] Assuming a power of 80 percent, with a significance level of .05, an effect size f^2 of .10, and 7 predictors.

construct (Barclay, Higgins, & Thompson, 1995; Nunnally, 1967). For the hypothesized latent construct structure, which had a maximum of seven indicators per construct, this would result in a required sample size of at least 70 for each model (or 140 overall). Using the most conservative of these guidelines, the minimum sample size recruited needed to be at least 332, with at least 166 men and at least 166 women. In order to obtain a sample of this size, a total of 6887 invitations were sent to solicit responses to the survey.

Survey responses received. Of the 6887 invitations, 1084 responses were received. Of these, 988 were complete. After discarding the responses of those who did not fit the criteria and those with more than 15% missing data, the analytic sample size consisted of 662 responses. Of these 662 responses, 163 were from female doctorate recipients and 417 were from male doctorate recipients. The remaining responses were from 40 men and 42 women who had enrolled in, but did not complete, a mathematics doctoral program. As will be discussed later, an exploratory factor analysis revealed a different latent construct structure than was hypothesized, in which there was a maximum of six indicators per latent construct. Therefore, revising the *a priori* estimates, and using the most conservative estimate for the required sample size, 157 participants were required for each analysis, or 314 overall. Therefore, the obtained sample size is sufficient to detect significant differences within the data. Table 2 presents the number and percent of survey invitations, responses received, and the response rate for each type of institution. Tables 3 and 4 present the number (and percent) of participants of each gender by job title and institution type, respectively. Table 5 presents participants' time since degree obtainment (or time since leaving) by degree completion status.

Table 2

Number (Percent) of Survey Invitations and Responses Received by Institution Type

	Number (percent) of invitations	Number (percent) of responses	Response rate
Associate's colleges	2867 (41.63%)	63 (9.52%)	2.20%
Baccalaureate colleges	1537 (22.32%)	262 (39.56%)	17.05%
Master's colleges/universities	1660 (24.10%)	125 (18.88%)	7.53%
Doctoral/research universities	823 (11.95%)	212 (32.02%)	25.76%
Total	6887	662	---

Note. Overall response rate = 9.61%.

Table 3

Number (Percent) of Survey Responses Received by Gender and Job Title

	Number of male participants	Number of female participants	Total
Full professor	162 (73.30%)	59 (26.70%)	221
Associate professor	122 (72.62%)	46 (27.38%)	168
Assistant professor	97 (61.39%)	61 (38.61%)	158
Post doctorate	10 (83.33%)	2 (16.67%)	12
Adjunct professor	13 (76.47%)	4 (23.53%)	17
Lecturer or instructor	35 (56.45%)	27 (43.55%)	62
Other	17 (73.91%)	6 (26.09%)	23
Total	456 (68.99%)	205 (31.01%)	661

Note. One participant left this item blank.

Table 4

Number (Percent) of Survey Responses by Gender and Institution Type

	Number of male participants	Number of female participants	Total
Associate's colleges	36 (57.14%)	27 (42.86%)	63
Baccalaureate colleges	173 (66.03%)	89 (33.97%)	262
Master's colleges/universities	99 (79.20%)	26 (20.80%)	125
Doctoral/research universities	149 (70.28%)	63 (29.72%)	212
Total	457 (69.03%)	205 (30.97%)	662

Table 5

Number of Survey Responses by Time Since Degree (Leaving) and Completion Status

Time since degree (leaving)	Obtained doctorate	Enrolled, but did not complete doctorate	Total
0 to 9 years	216 (88.52%)	28 (11.48%)	244
10 to 19 years	157 (84.86%)	28 (15.14%)	185
20 to 29 years	89 (88.12%)	12 (11.88%)	101
30 to 39 years	65 (87.84%)	9 (12.16%)	74
40 to 49 years	44 (89.80%)	5 (10.20%)	49
50 or more years	9 (100.00%)	---	9

Total	580 (87.61%)	82 (12.39%)	662

Previous research reports that male students complete their doctorates in less time than female students (Herzig, 2004a). In contrast, participants in this study did not reflect this trend. While male participants (n_M = 409, M_M = 5.61 years, SD_M = 1.59) completed their doctorates in slightly less time than female participants (n_F = 161, M_F = 5.65 years, SD_F = 1.35) in this sample, the difference was not significant ($t(568) = -0.253$, $p = .800$). Also of interest, over 92% of the sample reported having a male advisor, which aligns with previously reported research (Miller, 2015a).

Data Collection

Participants were invited by e-mail to participate in the study. The e-mail contained a link to the electronic survey on Qualtrics. Data was collected through the Qualtrics website, a service for building surveys and collecting data electronically (Qualtrics Labs Inc., Provo, UT). After one week, an e-mail reminder was sent to encourage those who had not yet completed the survey to do so. Data collection was conducted in three waves, with additional invitations to participate in the survey being sent in each wave until the necessary minimum number of participants of each gender was met.

Data Analysis

Creating an Overall Scale for Institutional Support Experiences

One of the 10 blocks on the survey assessed participants' experiences with institutional supports (e.g., teaching assistantships, fellowships), in terms of both the academic benefits and the time demands. Since doctoral students may receive different forms and durations of institutional support, participants were asked to evaluate only those sources of support they had received. Consequently, a consistent measure for the evaluation of the particular institutional support each participant experienced was needed. In order to summarize each participant's various sources of support and their evaluation of each, two new items were calculated (one for academic benefits and one for time demands). Each new item represents the weighted average of the participant's evaluations of each source from which they received funding, weighted by the duration of the funding:

$$InstSup1 = \frac{1}{T}\sum_{i=1}^{n} L_i t_i$$

where n is the number of sources from which the participant received funding, T is the total duration of the funding received while in a doctoral program, L_i is the participant's Likert evaluation of the academic benefits (or time demands) of i^{th} funding source, and t_i is the duration for the i^{th} funding source in years.

Formulating Latent Constructs with Exploratory Factor Analysis

After data collection, relationships between the 62 Likert items on the survey were tested with exploratory factor analysis to formulate latent constructs for the PLS-SEM analyses. Exploratory factor analysis was conducted in SPSS (IBM Corp.) through a principal components extraction with a varimax rotation. The Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was .863, above the recommended minimum value of .6, indicating that the sample size of 553 participants with no missing data was adequate for factor analysis (Hutcheson & Sofroniou, 1999). Additionally, Bartlett's test of sphericity was significant (χ^2 (1891) = 13988.41, $p < .01$), indicating that correlations exist within the data, making it suitable for factor analysis (Dziuban & Shirkey, 1974).

According to Kaiser's criterion, 17 factors exist with eigenvalues greater than 1. Collectively, these 17 factors explain 63.91 percent of the variance in the data. Considering loadings above .4 as significant (Stevens, 2002), the following six survey items did not load above .4 on any of the 17 factors: *PersChar4*, *PersCons4*, *ContPrep3*, *InstSup4*, *Fairness4*, and the weighted average of the participant's evaluations of the time demands of each funding source, weighted by the duration of the funding (similar to *InstSup1*).

After removing these six items, the data were reanalyzed for the remaining 56 items. Again, the KMO measure of sampling adequacy (.850) and Bartlett's test of sphericity (χ^2 (1540) = 12495.14, $p < .01$) indicated the data were suitable for factor analysis. After removing the items with low loading values from the analysis, 16 factors (15 predictor constructs and one outcome construct) were detected with eigenvalues greater than 1. These factors explained 64.89 percent of the variance in the data. It is these 16 factors that constituted the latent constructs for the PLS-SEM analyses.

Multiple Imputation for Missing Data

Before any additional analyses were conducted, any participants with greater than 15 percent missing data (i.e., participants who did not respond to at least 10 items on the survey) were discarded from the sample. For the remaining missing data, multiple imputation with five iterations was employed to maximize the useable sample size. Multiple imputation is the preferred method for dealing with missing data, since it does not bias the resulting data set as severely as other methods, such as mean imputation (Sinharay, Stern, & Russell, 2001) and does not drastically decrease the useable sample size for analyses, as with casewise deletion (Hair et al., 2014). The imputation was conducted using the "Fully conditional specification" option in SPSS (IBM Corp.), also known as Markov Chain Monte Carlo (MCMC) imputation, meaning that five separate data sets were created, each with different imputed missing values based on predictions from the observed data. Then, each PLS-SEM analysis was conducted five times, with the final results being pooled from the five sets of results according to Rubin's (1987) rules.

Partial Least Squares Structural Equation Modeling

Data analyses were conducted using partial least squares structural equation modeling (PLS-SEM). For the purposes of this study, only data from participants with doctorates was analyzed. PLS-SEM, which is conceptually similar to multiple regression, allows for evaluation of causal relationships between latent constructs, instead of only observable variables (Hair, Ringle, & Sarstedt, 2011). Although the analysis technique is most widely used for business applications, its use in social science research has become more commonplace in recent years (e.g., Monteiro, Wilson, & Beyer, 2013; Velayutham, Aldridge, & Fraser, 2012).

The use of PLS-SEM allowed for the investigation of the comparative effects on success of various factors associated with doctoral study in mathematics. This analysis attempts to "maximize explained variance in the dependent constructs [while evaluating] the data quality on the basis of measurement model characteristics" (Hair et al., 2011). As opposed to covariance-based structural equation modeling (CB-SEM), PLS-SEM is more appropriate for this study for several reasons. First, it is a more suitable choice for exploratory analyses and reduces some of the biases inherent in CB-SEM (Hair et al., 2014). Second, PLS-SEM has been shown to have greater statistical power than CB-SEM and thus, has the ability to detect significant differences when utilizing a smaller sample than with CB-SEM (Hair et al., 2014). Third, PLS-SEM is based on less restrictive assumptions for the distribution of the data. For instance, normality is not assumed; PLS-SEM analyses have been shown to be robust to skewed or kurtotic data with sufficiently large sample sizes (Hair et al., 2012; Hair et al., 2014). In order to conduct the PLS-SEM analyses, the software program SmartPLS was used (Ringle, Wende, & Becker, 2014).

Although other programs, such as LISREL (Jöreskog & Sörbom, 2006) are more commonly used to conduct CB-SEM analyses, SmartPLS is uniquely suited for conducting PLS-SEM analyses.

Methods Used to Investigate Research Question 1

In order to identify the factors with the strongest association with female participants' doctoral program success, a PLS-SEM analysis was conducted using the data from only female participants who had obtained doctorates. Analysis of these data allowed for a determination of the strength of the associations between the latent constructs and the outcome, doctoral program success. These associations, reported as pooled path coefficients, were then compared to determine which latent constructs had the strongest impact on doctoral program success. This analysis was conducted using the following conventional specifications: a path weighting scheme, which maximizes the value of R^2 for the latent variables; a raw data transformation to standardize the input data; an initial value of +1 to initialize the analysis; a threshold stopping criterion of 0.00001 to ensure stabilization of the results; and a maximum of 300 iterations for convergence (Hair et al., 2014).

Methods Used to Investigate Research Question 2

In order to compare factors associated with participants' doctoral program success for male and female participants with doctorates, multi-group analysis was conducted (Hair et al., 2014). Multi-group analysis compares pairs of path coefficients for latent variables for different samples: in this case, female participants and male participants (Kock, 2014). Path coefficients were compared for significance using two methods: the pooled standard error method, which assumes the standard errors of the two samples are not significantly different; and the Satterthwaite method, which does not make assumptions about the standard errors of the data (Kock, 2014).

Results

RQ 1: Impactful Factors in the Success of Female Participants

Research Question 1 aimed to identify the relative impact of the 15 latent constructs on mathematics doctoral program success of female students. Five constructs were significantly predictive of the outcome construct (i.e., the pooled path coefficient was statistically significant). These constructs were Personal Characteristics, Personal Considerations, Academic Support from Advisor, Academic Benefits of Institutional Support, and Obstacles Faced. Interestingly, these factors include a mix of personal factors and institutional or program-level factors.

Of the five significant constructs, Obstacles Faced was most impactful in the successful of female participants. Its pooled path coefficient of 0.257 implies that a one standard deviation increase in a participant's evaluation of the obstacles faced while enrolled in their doctoral program would result in over a quarter of a standard deviation increase in their evaluation of their doctoral program success. In comparing the path coefficients, the obstacles a participant faces in her doctoral program are nearly twice as impactful as a student's personal considerations, such as familial and financial responsibilities. The remaining 10 constructs did not reach statistical significance. The pooled path coefficients can be found in Table 7.

Table 7
Pooled Path Coefficients for Female and Male Participants

	Pooled path coefficients	
Construct	Female participants	Male participants
Personal Characteristics	0.196**	0.138***
Personal Considerations	0.144*	0.061
Content Preparation	-0.002	0.090*
Sense of Belonging	-0.059	0.052
Academic Support from Advisor	0.205**	0.283***
Interactions with Others in the Department	0.141	0.066
Quality and Availability of Courses	0.141	0.169**
Academic Benefits of Institutional Support	0.176**	0.061
Professor Gender Ratios	0.013	-0.002

Student Gender Ratios	0.026	0.040
Ratios for Student Success	-0.077	0.059
Fairness of Policies	0.014	0.082*
Obstacles Faced	0.257***	0.050
Unwanted Attention Due to Gender	0.091	0.052
Opinions About Success Due to Gender	0.056	0.006

Note. *$p < .05$, **$p < .01$, ***$p < .001$

RQ 2: Comparison of Significant Factors for Female and Male Participants

Research Question 2 sought to investigate differences in the importance of the identified factors for female and male participants with doctorates. One way to determine how factors associated with doctoral program success compare for female and male participants is to compare those constructs whose path coefficients reached statistical significance in the model for the female participants and the model for the male participants. Academic Support from Advisor and Personal Characteristics were significantly predictive of doctoral program success for both genders. Personal Considerations, Academic Benefits from Institutional Support, and Obstacles Faced were predictive only for female students, while Content Preparation, Quality and Availability of Courses, and Fairness of Policies were predictive only for male students.

Additionally, recall that two types of multi-group analyses were conducted to compare the path coefficients for female participants to those of male participants. Table 8 presents the results of the multi-group analysis for all 15 predictor constructs. Only one comparison reached statistical significance and that was for the construct Obstacles Faced. Thus, the construct Obstacles Faced had a significantly stronger relationship with doctoral program success for female participants than for male participants (p = .001).

Table 8
Multi-group Analysis for Female and Male Participants

Construct	*t*-value (Pooled standard error method)	*t*-value (Satterthwaite method)
Personal Characteristics	0.7338	0.6828

Personal Considerations	1.0109	1.0501
Content Preparation	-1.1633	-1.2018
Sense of Belonging	-1.2414	-1.2701
Academic Support from Advisor	-0.8984	-0.9150
Interactions with Others in the Department	0.7484	0.7397
Quality and Availability of Courses	-0.2689	-0.2820
Academic Benefits from Institutional Support	1.2934	1.3946
Professor Gender Ratios	0.1707	0.1831
Student Gender Ratios	-0.1745	-0.1815
Ratios for Student Success	-1.6302	-1.5218
Fairness of Policies	-0.8506	-0.8113
Obstacles Faced	3.2253**	3.2968**
Unwanted Attention Due to Gender	0.4482	0.3952
Opinions About Success Due to Gender	0.5746	0.5212

Note. $^*p < .05$, $^{**}p < .01$, $^{***}p < .001$

Discussion

This study was conducted to investigate the experiences of successful female mathematics doctoral students and to compare these experiences to that of male doctoral students. Much of the previous research in this area has focused specifically on issues of *attrition* of female students, utilizing small samples and qualitative methodologies. In contrast, this study used a large, representative sample of mathematics faculty members and focuses on factors associated with doctoral program *success*. While previous studies provided detailed descriptions of individuals' experiences, the findings were not generalizable. Moreover, it was unclear which factors were most critical in explaining retention and attrition. For this study, the use of structural equation modeling, combined with the inclusion of male participants as a comparison group, allows for more nuanced claims to be made than in previous work. For instance, previous studies made claims about the importance of the relationship between a student and her advisor based on the participant's qualitative self-report (e.g., Herzig, 2004b; Herzig, 2010; Hollenshead et al., 1994). However, it was unknown how influential this factor is in comparison to other factors reported by the participant as important. In this study, factors influencing doctoral program success were compared quantitatively to confirm that the quality of the advisor-advisee relationship was, in fact, one of the most influential factors, regardless of gender. Moreover, because of the representativeness of the sample, these results are generalizable to the population of mathematics faculty members in the United States.

Analyses of the data collected reveal that female participants with doctorates found aspects of their personal lives, the academic support they received from their advisor, the academic benefits of the institutional support they received (in the form of assistantships and fellowships), and the obstacles they faced on their path to their doctorate to be most impactful on their doctoral program success. Obstacles included both personal or individual obstacles (struggling with confidence) and programmatic or institutional obstacles (passing benchmark exams).

Differences were also detected in the experiences of male and female doctoral graduates. Only one of the 15 factors – Obstacles Faced – reached significance in comparing the latent constructs in the multi-group analysis. This means that obstacles faced were a significantly stronger predictor of doctoral program success for women than for men. One hypothesis that arises from this finding is that women might be more inclined to interpret obstacles faced as a detriment to their doctoral program success because they tend to have lower mathematics self-efficacy than men. Another hypothesis deals with different tendencies for the attributions of success and failure by men and by women: a woman’s success is more often attributed to luck or effort, while a man’s success is more often attributable to innate ability (Lott, 1985). Conversely, women’s failures tend to be associated with personal shortcomings, such as ability, while men’s failures are usually attributed to external circumstances, such as bad luck (Lott, 1985). Furthermore, the constructs that reached significance in the separate male and female

PLS-SEM models differed. Personal considerations (such as family responsibilities), opportunities to learn from teaching or research assistantships, and overcoming obstacles were predictive of success for female participants only. The significance of assistantship assignments for female participants, but not for male participants, could be a problem of inequity or of the perception of inequity. Female doctoral students may, in fact, be assigned assistantships with inferior opportunities to learn due to biased or inequitable practices. Alternatively, female students may perceive that they are not able to gain the same level of academic benefit from their assistantships as male students because male students may be able to form stronger bonds with lead course instructors or principal investigators, who are likely also male. Female students should thus be prepared to advocate for their own learning in this area by requesting a range of funding opportunities, including both teaching and research assistantships, during their doctoral program. For male participants, content preparation, coursework, and the fairness of policies within the department had a stronger influence on doctoral program success than for female participants. However, it is noteworthy that both a student's relationship with his or her advisor and personal characteristics were predictive in both models, suggesting that improvements or additional supports in these areas would be beneficial for all students, regardless of gender.

The results of this study suggest five key recommendations for doctoral programs and for female students. Doctoral programs could use these findings to empower their female students to become better advocates for their own learning. First, this research provides additional support for the finding from previous research that the role of the advisor has a strong influence on doctoral student success (e.g., Bair & Haworth, 2004; Fagen & Wells, 2004; Miller, 2015a; Tinto, 1993). The importance of developing and maintaining a productive advisor-advisee relationship resonated across both of the two main analyses. A supportive advisor was a prominent factor in explaining the success of both female and male participants. Therefore, departments could emphasize the importance of advising as part of their tenure review process to reward faculty for devoting time to improving these relationships (Bair & Haworth, 2004) and increase their awareness of the importance of their role as an advisor through training or professional development. Female students should be aware of the importance of choosing a suitable advisor and make this decision with great care. This choice should likely be based on considerations including, but not limited to, alignment of areas of research interest, compatibility of personality types, and potentially even discussions with a professor's former doctoral students. Second, additional female faculty members could be hired in order to provide visible female mentors and role models to female students. Third, a culture could be created within the program where both doctoral advisors and faculty members sponsoring students as teaching or research assistants are encouraged to mentor students and focus on their students' development as future faculty members and scholars. Although personal factors are outside the scope of a program's control, doctoral programs could provide supports in order to give admitted students the greatest possible chance of success. The fourth recommendation, to equalize available time for schoolwork for students with and without families, is for support or provisions for childcare to be integrated into the institutional structure. Finally, for students with financial concerns, programs could refer students to free or low-cost financial advisement in their area to help with budgeting, student loans, and programs available to assist students and low-income individuals.

Many of the findings presented here provide additional support for claims made in other studies. However, the comparisons of the relative importance of each factor for female students and between female and male students provide a unique contribution to understanding the mechanisms underlying success in doctoral mathematics. In the future, institutions could administer this survey to their current students (as part of a yearly review) or recent graduates (as part of an exit interview), creating a feedback mechanism to guide changes within the mathematics doctoral program. Furthermore, the survey could be modified and used by other researchers to investigate similar issues of retention and attrition for doctoral students in other STEM fields. Revisions that may need to be made include the addition of items pertaining to availability of laboratory time, space, and resources, and the quality and productivity of interactions between laboratory group members.

Now that factors influencing the success of female mathematics doctoral students have been empirically investigated in a more generalizable manner than has previously been done, small-scale interventions can begin to be implemented at individual institutions to see if modifications to these factors result in greater success for female students. Because of the length of time required to obtain a doctorate, these studies would need to collect longitudinal data over for a minimum of five years before assessments of efficacy could be made. If substantial improvements occur, these interventions could then be scaled up to include more institutions over time.

Additionally, it is an open question as to how these results would differ if the sample were expanded beyond those doctoral recipients employed in academia. Different factors may be important for success for those whose career goals will lead to employment in government or in industry. However, obtaining a representative sample from that portion of the population of mathematics doctorate recipients would be challenging. A potential starting point would be to track the career trajectories of recent graduates in order to determine if any trends exist in where these doctoral recipients accept employment. Then, once this is known, it may be easier to recruit a somewhat representative sample of those employed outside academia. Results from this population could then be compared to the results presented here to make recommendations to promote the success of all women, regardless of their career aspirations.

Conclusion

The research described here is a key step in formulating a set of best practices for retaining female mathematics doctoral students. This has the potential to make a significant impact in narrowing the gender gap both in participation and in success for mathematics doctoral students. An increased number of female graduates from mathematics doctoral programs should eventually lead to a more balanced gender ratio for mathematics faculty members. This, in turn, could have the effect of encouraging more women to become interested in and study mathematics, diversifying the discipline to the benefit of all involved (Hill et al., 2010). Over time, with increased participation and a greater number of female mathematicians, mathematics educators, and mathematics teachers as role models, gendered stereotypes of mathematical competency may become a thing of the past.

References

Akter, S., D'Ambra, J., & Ray, P. (2011). An evaluation of PLS based complex models: The roles of power analysis, predictive relevance and GoF index. Proceedings of the Seventeenth Americas Conference on Information Systems, August 2011.

Allen, M.J., & Yen, W.M. (1979). *Introduction to measurement theory.* Long Grove, IL: Waveland Press.

Ampaw, F.D., & Jaeger, A.J. (2011). Understanding the factors affecting degree completion of doctoral women in the science and engineering fields. *New Directions for Institutional Research, 2011*(152), 59 – 73.

Bair, C. R., & Haworth, J. G. (2004). Doctoral student attrition and persistence: A meta-synthesis of research. In J. C. Smart (Ed.), *Higher Education: Handbook of Theory and Research* (Vol. XIX, pp. 481–534). Netherlands: Kluwer Academic Publishers.

Baird, L. L. (1993). Using research and theoretical models of graduate student progress. *New Directions for Institutional Research, 1993*(80), 3 – 12.

Barclay, D.W., Higgins, C.A., & Thompson, R. (1995). The partial least squares approach to causal modeling: Personal computer adoption and use as illustration. *Technology Studies, 2*(2), 285 – 309.

Becker, J. R. (1984). The pursuit of graduate education in mathematics: Factors that influence women and men. *Journal of Educational Equity and Leadership*, *4*(1), 39 – 53.

Becker, J. R. (1990). Graduate education in the mathematical sciences: Factors influencing women and men. In L. Burton (Ed.), *Gender and Mathematics: An International Perspective* (pp. 119–130). London, England: Cassell Educational Limited.

Berg, H. M., & Ferber, M. A. (1983). Men and women graduate students: Who succeeds and why? *The Journal of Higher Education*, *54*(6), 629 – 648.

Blair, R., Kirkman, E.E., & Maxwell, J.W. (2013). *Statistical abstract of undergraduate programs in the mathematical sciences in the United States*. Providence, RI: American Mathematical Society.

Blickenstaff, J.C. (2005). Women and science careers: Leaky pipeline or gender filter? *Gender and Education, 17*(4), 369 – 386.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Dziuban, C.D., & Shirkey, E.C. (1974). When is a correlation matrix appropriate for factor analysis? *Psychological Bulletin, 81*(6), 358 – 361.

Eccles, J.S. (2002). Achievement. In J. Worell (Ed.), *Encyclopedia of Women and Gender: Sex Similarities and Differences and the Impact of Society on Gender* (pp. 43 – 53). San Diego, CA: Academic Press.

Ehrenberg, R.G., & Mavros, P.G. (1992). Do doctoral students' financial support patterns affect their times-to-degree and completion probabilities (Working Paper No. 4070). Retrieved from National Bureau of Economic Research website: http://www.nber.org/papers/w4070

Ethington, C.A., & Pisani, A. (1993). The RA and TA experience: Impediments and benefits to graduate study. *Research in Higher Education, 34*(3), 343 – 354.

Fagen, A.P., & Wells, K.M.S. (2004). The 2000 national doctoral program survey: An on-line study of students' voices. In D.H. Wulff & A.E. Austin (Eds.), *Paths to the Professoriate: Strategies for Enriching the Preparation of Future Faculty* (pp. 77 – 91). San Francisco, CA: Jossey-Bass.

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175 – 191.

Golde, C.M. (1998). Beginning graduate school: Explaining first-year doctoral attrition. *New Directions for Institutional Research, 1998*(101), 55 – 64.

Hair Jr., J.F., Hult, G.T.M., Ringle, C.M., & Sarstedt, M. (2014). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Los Angeles: Sage.

Hair, J.F., Ringle, C.M., & Sarstedt, M. (2011). PLS-SEM: Indeed a silver bullet. *Journal of Marketing Theory and Practice, 19*(2), 139 – 151.

Hair, J.F., Sarstedt, M., Pieper, T.M., & Ringle, C.M. (2012). The use of partial least squares structural equation modeling in strategic management research: A review of past practices and recommendations for future applications. *Long Range Planning, 45*(5-6), 320 – 340.

Hall, R.M, & Sandler, B.R. (1982). *The classroom climate: A chilly one for women?* Washington, D.C.: Association of American Colleges.

Herzig, A. H. (2002). Talking the talk: Graduate women in the disciplinary culture of mathematics. *Annual Meeting of the American Education Research Association*. New Orleans, LA.

Herzig, A. H. (2004a). Becoming mathematicians: Women and students of color choosing and leaving doctoral mathematics. *Review of Educational Research*, *74*(2), 171 – 214.

Herzig, A. H. (2004b). "Slaughtering this beautiful math": Graduate women choosing and leaving mathematics. *Gender and Education*, *16*(3), 379 – 395.

Herzig, A. H. (2010). Women belonging in the social worlds of graduate mathematics. *The Montana Mathematics Enthusiast*, *7*(2&3), 177 – 208.

Hill, C., Corbett, C., & St. Rose, A. (2010). *Why so few? Women in science, technology, engineering, and mathematics.* American Association of University Women.

Hutcheson, G., & Sofroniou, N. (1999). *The multivariate social scientist: Introductory statistics using generalized linear models*. Thousand Oaks, CA: Sage Publications.

IBM Corp. Released 2013. IBM SPSS Statistics for Macintosh, Version 22.0. Armonk, NY: IBM Corp.

Jöreskog, K.G. & Sörbom, D. (2006). LISREL 8.8 for Windows [Computer software]. Skokie, IL: Scientific Software International, Inc.

Kerlin, R. (1997). *Breaking the silence: Toward a theory of women's doctoral persistence* (Unpublished doctoral dissertation). University of Victoria, British Columbia.

Kock, N. (2014). Advanced mediating effects tests, multi-group analyses, and measurement model assessments in PLS-based SEM. *International Journal of e-Collaboration, 10*(1), 1 – 13.

Lott, B. (1985). The devaluation of women's competence. *Journal of Social Issues, 41*(4), 43 – 60.

McAlpine, L., & Norton, J. (2006). Reframing our approach to doctoral programs: An integrative framework for action and research. *Higher Education Research & Development, 25*(1), 3 – 17.

Miller, E. (2015a). Examining factors associated with the success of women in mathematics doctoral programs. *Journal of Women and Minorities in Science and Engineering, 21*(1), 47 – 85.

Miller, E. (2015b). Factors contributing to the success of female mathematics doctoral students: A review of the literature. Unpublished manuscript. School of Education, University of Delaware, Newark, Delaware.

Monteiro, H., Wilson, J., & Beyer, D. (2013). Exploration of predictors of teaching effectiveness: The professor's perspective. *Current Psychology, 32*(4), 329 – 337.

Nerad, M., & Miller, D.S. (1996). Increasing student retention in graduate and professional programs. *New Directions for Institutional Research, 1996*(92), 61 – 76.

Nunnally, J.C. (1967). *Psychometric Theory*. New York: McGraw-Hill.

Phipps, P., Maxwell, J.W., & Rose, C.A. (2007). 2006 annual survey of the mathematical sciences in the United States (Third report). *Notices of the American Mathematical Society, 54*(10), 1333 – 1344.

Preckel, F., Goetz, T., Pekrun, R., & Kleine, M. (2008). Gender differences in gifted and average ability students: Comparing girls' and boys' achievement, self-concept, interest, and motivation in mathematics. *Gifted Child Quarterly, 52*(2), 146 – 159.

Qualtrics. (2015). Qualtrics: Online Survey Software & Insight Platform. Available from http://www.qualtrics.com/

Ringle, C.M., Wende, S., & Becker, J.-M. (2014). SmartPLS 3. Hamburg: SmartPLS. Retrieved from http://www.smartpls.com

Robst, J., Keil, J., & Russo, D. (1998). The effect of gender composition of faculty on student retention. *Economics of Education Review, 17*(4), 429 – 439.

Rubin, D.B. (1987). *Multiple imputation for nonresponse in surveys.* New York: Wiley.

Schroeder, D., & Mynatt, C. (1993). Female graduate students' perceptions of their interactions with male and female major professors. *Journal of Higher Education*, *64*(5), 555 – 573.

Sinharay, S., Stern, H.S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods, 6*(4), 317 – 329.

Spencer, S., Steele, C., & Quinn, D. M. (1999). Stereotype threat and women's math performance. *Journal of Experimental Social Psychology*, *35*(1), 4 – 28.

Stevens, J. (2002). *Applied Multivariate Statistics for the Social Sciences.* Mahwah, NJ: Lawrence Erlbaum Associates.

Sue, D.W. (2010). *Microaggressions in everyday life: Race, gender, and sexual orientation.* Hoboken, NJ: Wiley.

Tinto, V. (1993). Toward a theory of doctoral persistence. *Leaving college: Rethinking the causes and cures of student attrition* (pp. 230 – 243). Chicago: University of Chicago Press.

Velayutham, S., Aldridge, J.M., & Fraser, B. (2012). Gender differences in student motivation and self-regulation in science learning: A multi-group structural equation modeling analysis. *International Journal of Science and Mathematics Education, 10*(6), 1347 – 1368.

Vélez, W.Y., Maxwell, J.W., & Rose, C. (2014). Report on the 2012-2013 new doctoral recipients. *Notices of the AMS, 61*(8), 874 – 884.