9-19-2022

A Multi-Level Analysis of the Effects of Statistics Anxiety/ Attitudes on Trajectories of Exam Scores

Kelly Rhea MacArthur

Jonathan Santo

Journal of Statistics and Data Science Education

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ujse21

A Multi-Level Analysis of the Effects of Statistics Anxiety/Attitudes on Trajectories of Exam Scores

Kelly Rhea MacArthur & Jonathan B. Santo

To cite this article: Kelly Rhea MacArthur & Jonathan B. Santo (2023) A Multi-Level Analysis of the Effects of Statistics Anxiety/Attitudes on Trajectories of Exam Scores, Journal of Statistics and Data Science Education, 31:1, 102-112, DOI: 10.1080/26939169.2022.2093805

To link to this article: https://doi.org/10.1080/26939169.2022.2093805

JOURNAL OF STATISTICS AND DATA SCIENCE EDUCATION
2023, VOL. 31, NO. 1, 102–112
https://doi.org/10.1080/26939169.2022.2093805

A Multi-Level Analysis of the Effects of Statistics Anxiety/Attitudes on Trajectories of Exam Scores

Kelly Rhea MacArthur[a] and Jonathan B. Santo[b]

[a]Department of Sociology & Anthropology, University of Nebraska Omaha, Omaha, NE; [b]Department of Psychology, University of Nebraska Omaha, Omaha, NE

ABSTRACT
This study explores three understudied facets—quadratic effects, change over time, and gender as a moderator—of the otherwise well-documented relationships between statistics anxiety and academic performance. Using pre- and post- course survey data among a sample of 111 undergraduate students in Social Statistics courses at a U.S. Midwestern university, we employ hierarchical linear modeling (HLM) to test for relationships between change in the six dimensions of the Statistics Anxiety Rating Scale (STARS) and exam grades over the course of the semester. We find that exam grades decreased over time, but at different rates across gender and the six STARS dimensions. We also identify a quadratic relationship between self-concept and final exam grades, as well as several gender interactions. This study highlights the multifaceted and dynamic nature of statistics anxiety/attitudes, as its relationship with academic performance is not always negative, linear, stable over time, or uniform across gender. Supplementary materials for this article are available online.

ARTICLE HISTORY
Received May 2020
Accepted April 2022

KEYWORDS
Gender; STARS; Statistics anxiety; Statistics education research

1. Introduction

A considerable degree of research has established that anxiety is negatively associated with learning (Rosenfeld 1978). This may be particularly evident in statistics courses that seem to engender extreme levels of fear surrounding having to perform mathematical calculations. The literature supports this point that anxiety has a strong inverse relationship with various measures of academic performance in statistics classes, as two reviews of the causes and effects of statistics anxiety conclude that there is a robust negative relationship between statistics anxiety and academic outcomes (Chew and Dillion 2014; Onwuegbuzie and Wilson 2003). In fact, anxiety may be the strongest predictor of grades in statistics classes (Fitzgerald, Jurs, and Hudson 1996). Given that most universities have quantitative literacy requirements, statistics anxiety may present a barrier to graduation, at least inasmuch that students avoid or fail statistics classes as a result of anxiety. The critical thinking skills, moreover, which quantitative literacy facilitates, are necessary for the maintenance of a just, informed, and democratic society.

In a review of the causes and effects of statistics anxiety, Onwuegbuzie and Wilson (2003) note that the Statistics Anxiety Rating Scale (STARS) is the most widely used survey to measure statistics anxiety and attitudes. The STARS distinguishes between three different sources of statistics anxiety, that from exams, interpreting mathematical operations, and asking others for help (Cruise, Cash, and Bolton 1985). Although the STARS 51-item scale was originally employed as a single statistics anxiety scale, more recently researchers agree that the three types should each be measured separately (Chew and Dillion 2014). The STARS also includes an additional three subscales that measure students' attitudes toward statistics teachers, one's own perceived abilities to do mathematical computations, and the perceived worth of statistics in their occupational and everyday lives (Cruise, Cash, and Bolton 1985). The STARS instrument and its subscales have been extensively validated (Cruise, Cash, and Bolton 1985; Baloğlu 2002; Liu, Onwuegbuzie, and Meng 2012; Papousek et al. 2012). Research that uses the STARS suggest that the six dimensions of statistics anxiety/attitudes are differentially related to indicators of academic performance, such as exam scores and final course grade (Onwuegbuzie 1995; Onwuegbuzie and Seaman 1995; Hanna, Shevlin, and Dempster 2008; Keeley, Zayac, and Correia 2008; Macher et al. 2013; MacArthur 2020). For example, Keeley, Zayac, and Correia (2008) found that which subscale was related to test scores varied across each of the six tests. Hanna, Shevlin, and Dempster (2008) similarly found that while fears of asking for help and computational self-concept predict students' estimated exam scores, interpretation anxiety and perceived worth of statistics predict actual exam scores. It is well-established, therefore, that anxiety and academic performance are negatively associated, but it is less clear which of the STARS six dimensions are related to which academic outcomes.

Despite the vast literature on statistics anxiety and, in particular, studies that employ the STARS instrument, there is

CONTACT Kelly Rhea MacArthur kmacarthur@unomaha.edu Department of Sociology & Anthropology, University of Nebraska Omaha, Omaha, NE.
Supplementary materials for this article are available online. Please go to www.tandfonline.com/ujse.

limited research that examines how change over time in statistics anxiety is tied to academic performance. Thus, this study explores three understudied facets regarding the otherwise well-documented relationships between statistics anxiety/attitudes and academic performance among undergraduate Social Statistics students: (a) changes in the six STARS dimensions; (b) changes in exam grades; and (c) quadratic effects of statistics anxiety/attitudes on exam scores. Furthermore, we examine how the important structural variable of gender moderates all of these relationships.

1.1. Changes over Time and Quadratic Effects

The first neglected area of research we address herein is that most studies examine how a given measure of statistics anxiety at one time point affects academic performance. However, there is evidence that neither anxiety levels in general, nor statistics anxiety/attitudes using the STARS instrument in particular, are static. Studies that do evaluate change suggest that anxiety decreases and attitudes improve over time (Keeley, Zayac, and Correia 2008; Williams 2013; McGrath et al. 2015; MacArthur 2020), usually due to a specific intervention designed to do so. Results are somewhat mixed, however, and some types of statistics anxiety and attitudes may remain stable over the course of the semester (Townsend et al. 1998; Zieffler et al. 2008; Lesser, Pearl, and Weber 2016) or even worsen (Zanakis and Valenzi 1997). Additionally, to our knowledge, there is only one study that examines change in anxiety/attitudes as a variable hypothesized to be associated with academic performance, but it does not find any statistically discernible effects (MacArthur 2020). Although, one study shows that an improvement in students' attitudes toward the value of statistics is associated with better grades (Zanakis and Valenzi 1997). Given these inconsistent findings and the unknown effects of change in the absence of a particular intervention, this study fills these gaps in the literature that neglect how changes in anxiety/attitudes are related to academic performance.

The second gap in the literature that this study addresses is that most studies employ an outcome measure of academic performance that is assessed at one time point, such as final course grade. For various reasons, performance in a class, such as exam grades, may improve or worsen over the course of a semester. Both baseline levels of anxiety/attitudes, as well as change in anxiety/attitudes, are arguably two factors that affect these trajectories. Wisenbaker, Scott, and Nasser (2000) found that statistics attitudes assessed at the end of the course had an effect on course performance, but baseline attitudes did not. Zanakis and Valenzi (1997) found the opposite pattern in that math anxiety at the beginning of the semester, but not at the end of the semester, was associated with course grade. And, to provide another example, MacArthur (2020) found that baseline levels of fear of teachers negatively affected exam grades, but self-concept at the end of the semester was positively associated with exam grades. In addition to these studies suggesting that baseline and ending levels of statistics anxiety may differentially matter for academic performance, there is also evidence that statistics anxiety may not uniformly affect academic performance at different time points throughout the course. Keeley, Zayac, and Correia (2008) study supports this point in that they found that at the first test out of six, none of the STARS dimensions were related to test scores. Some of the STARS subscales, however, were related to Tests 2 through 5 and all six scales were correlated with Test 6. Despite that the studies discussed above emphasize that the timing of the assessment matters—of statistics anxiety, as well as of academic performance—there are no studies that simultaneously examine the role of changes in statistics anxiety/attitudes on changes in exam scores.

Thirdly, this study tests for nonlinear effects and, specifically, quadratic effects, between STARS constructs and exam grades, which only one other study has done. Arguably, the most prominent scholar of statistic anxiety, Onwuegbuzie advocated decades ago for testing whether the relationship between anxiety and academic performance is actually linear (Onwuegbuzie and Wilson 2003). Nonetheless, studies' analytic approaches continue to assume that the relationship is linear and negative. In the only exception, following the logic of the Yerkes-Dodson law that claims that there are floor and ceiling effects to which stress effects learning, Keeley, Zayac, and Correia (2008) tested whether this principle applies to statistics anxiety and course grades. They found that, indeed there is an optimal level of anxiety in that while those low and high in test anxiety perform poorly on exams, those who report moderate levels of test anxiety perform the best. A moderate degree of anxiety may inspire students to prepare more fastidiously compared to those who are very low in anxiety (Onwuegbuzie and Wilson 2003). Conversely, an extremely high degree of anxiety may interfere with students' cognitive abilities during an exam (Onwuegbuzie 2003). Even though test scores went down on average with each of the six tests as the material became more difficult, Keeley, Zayac, and Correia (2008) found that the relationship between test anxiety and scores became stronger and more curvilinear with each test in that a quadratic equation, as opposed to a linear relationship, was a better fit to the data with each subsequent test. Additionally, instead of a linear relationship, a quadratic equation was the best fit for perceived worth of statistics and interpretation anxiety for final exam scores. A few other studies confirm that at least some degree of anxiety may facilitate academic performance (Onwuegbuzie 1995; Onwuegbuzie and Seaman 1995; Haynes, Mullins, and Stein 2004; Gyuris and Everingham 2011; Mellanby and Zimdars 2011; Macher et al. 2013), although none of these studies explicitly test for quadratic effects between STARS dimensions and academic performance.

Given the theoretical justifications for testing for curvilinear effects, in conjunction with preliminary empirical evidence provided by Keeley, Zayac, and Correia (2008), we similarly hypothesize that there may be a mid-level degree of anxiety that maximizes students' exam scores over the course of the semester. Whereas Keeley, Zayac, and Correia (2008) focused on students' need for achievement as a moderator, here we elaborate on Keeley's analysis of curvilinear effects by exploring whether and how gender shapes the relationships—and changes in them—between statistics anxiety and academic performance.

1.2. The Role of Gender

There is a large body of research that attempts to address the causes and effects of the well-documented and persistent pattern of women's underrepresentation in Science, Technology, Engineering, and Math (STEM) fields. One explanation that is often cited in the literature is that women have more anxiety regarding math and statistics that impedes their confidence in their abilities to develop the quantitative literacy skills that a career in a STEM field necessitates (Zeldin and Pajares 2000; Stout et al. 2011). Most studies indicate that female college students do in fact report higher levels of statistics and math anxiety compared to male students (Benson 1989; Bradley and Wygant 1998; Onwuegbuzie 1998; Zettle and Raines 2000; Mills 2004; DeCesare 2007; Condron, Becker, and Bzhetaj 2018; MacArthur 2020), although some studies find mixed results or no gender differences (Zeidner 1990; Benson, Bandalos, and Hutchinson 1994; Townsend et al. 1998; Baloğlu 2003; Haynes, Mullins, and Stein 2004; Onwuegbuzie 2004; Van Gundy et al. 2006; Bui and Alfaro 2011; Wismath and Worrall 2015). Studies using the STARS constructs show differential effects across the six dimensions. Hedges (2017), for example, finds that women report higher test anxiety and worse self-concept than men, while Macher et al. (2013) find that women have elevated rates of anxiety about exams, interpretation, and self-concept. Whereas MacArthur (2020) finds that women report higher levels of exam and interpretation anxiety and worse attitudes toward self-concept, and statistics worth than men, but no gender differences for fear of asking for help or attitudes toward teachers. Thus, although the research is somewhat mixed, we expect that female students will report more statistics anxiety and worse attitudes than will men.

In addition to the documented gender differences in baseline mean levels of anxiety, there is evidence that the qualitative experience of statistics anxiety may vary by gender as well. For example, it is unclear whether women's and men's anxiety changes at different rates throughout the semester. Since female students tend to have lower self-confidence in their math abilities than men (Sax 1994; Haynes, Mullins, and Stein 2004; Condron, Becker, and Bzhetaj 2018), they may quickly gain confidence with the reduction of uncertainty that typically comes with exposure to the course material and perhaps a supportive instructor. In turn, women's anxiety may decrease at a faster rate than men's, as one study shows that women's test anxiety decreases across the course of a semester at a faster rate than men's (MacArthur 2020). Anxiety may also function differently for women and men in terms of academic performance as, for example, Mellanby and Zimdars (2011) show that anxiety operates in a facilitative manner in terms of exam grades among women, but not among men, meaning that anxiety and grades may have a quadratic relationship for women, but not for men. There is theoretical reason, therefore, to suspect that gender will shape all of the patterns that will be tested in the current study.

1.3. Study Aims

The purpose of this study is to explore several aspects of the relationship between statistics anxiety and academic performance that tend to be neglected in prior research, including: (a) baseline and ending levels of statistics anxiety/attitudes; (b) changes in anxiety/attitudes over the course of the semester; (c) changes in exam grades; and (d) quadratic relationships between changes in anxiety/attitudes and exam grades. Finally, we explore gender as a moderator for all of these relationships. In summary, the main goal of this study is to refine understandings of the nature of statistics anxiety by exploring several understudied aspects of the relationships between anxiety and academic performance among undergraduate students in statistics courses. Ultimately, the findings of this study can be used to help statistics instructors who are already relying on the guiding principles set forth in the *American Statistical Association's (ASA) Guidelines for Assessment and Instruction in Statistics Education* (GAISE) college report to improve their teaching by: (a) emphasizing statistical literacy; (b) using real data; (c) stressing conceptual understanding; (d) incorporating active learning; (e) analyzing data with technology; and (f) supporting student learning with assessments (American Statistical Association 2005; Carver et al. 2016).

2. Materials and Methods

2.1. Sample/Procedure

Data for this study were collected from a metropolitan public university in eight sections of the 200-level class, Social Statistics, all taught by the first author, who, at the time of data collection began, was an assistant professor of sociology in their second year of teaching statistics (of approximately eight years of collegiate teaching overall). Majors from all colleges can enroll in order to fulfill a quantitative literacy requirement, but only Sociology majors are required to take the 1-credit accompanying lab course that uses statistical software. The prerequisites for the course are sophomore-standing and a basic math class and class enrollment is capped at 25 students per section. Like most introductory level statistics courses, it begins by covering descriptive statistics (e.g., measures of central tendency and dispersion) and then progresses to inferential statistics (e.g., confidence intervals) and culminates in multiple regression techniques. Pre- (T1) and post- (T2) survey data were collected across four consecutive semesters from Fall 2016 through Spring 2018. Class sessions were held for 1 hr and 15 min twice a week. The T1 data were collected during the class after the first week of class when the enrollment period had closed. The T2 data were collected on the last day of class before final exam week. This study complied with all IRB regulations and the surveys were anonymous, voluntary, and students were given informed consent information (IRB#: 863-16-EX).

There is a considerable amount of missing data from T1 to T2 (or vice versa) due to a variety of reasons, including students who (a) dropped the course; (b) were absent from class on the day the T1 and/or T2 surveys were administered; (c) chose not to complete either/both the T1 or T2 survey; or (d) did not correctly write the last 4 digits of their student I.D. number on one of the surveys. Although it is not possible to calculate the exact response rate, there is complete data (matching T1 and T2 surveys and not missing on any of the analytic variables) for 112 students out of the 199 who were enrolled at T2, which constitutes an approximate 57% response rate. One additional student was dropped from the sample due to lacking any grade scores.

A missing data analysis using independent samples t-tests on all analytic variables revealed that those with lower exam scores were more likely to have missing data, as well as those who had more anxiety about interpretation and worse attitudes toward teachers and self-concept at T1. To deal with missing data, we deleted any observation from the dataset that was missing on any of the analytic variables (i.e., listwise deletion), which is the most robust method for dealing with data on the independent variables that are not missing at random (Allison 2001). The final analytic sample with no missing data consists of 111 students, 79 women and 32 men.

2.2. Measures

2.2.1. Dependent Variable: Academic Performance

Academic performance is measured by three exam scores. The first two exams were worth 100 points and the final exam, while not cumulative, was worth 150 points, but they are standardized for these analyses by using the percentage of points earned on each exam. All of the exams are mixed format (e.g., multiple choice, true/false, calculations, open-ended) and the second two exams are a mixture of in-class and take home in which they receive the take-home portion a week ahead of time. Taken together, the exams constitute a little under half of their total grade in the course.

2.2.2. Change over Time in Academic Performance

Change over time in the three exam grades was quantified by creating a variable coding the first exam as "−2", the second as "−1" and the final exam as "0". This allowed us to model differences between the students in how exam grades changed over the course of the semester in addition to delineating differences between the students' final grades (when time = 0).

2.2.3. The Statistics Anxiety Rating Scale (STARS)

The Statistics Anxiety Rating Scale (STARS; Cruise, Cash, and Bolton 1985) is a 51-item scale that contains six subscales that measure three types of statistics anxiety and three types of attitudes toward statistics. The first three subscales are sources of anxiety that are measured on a 3-point Likert scale with responses 1= "no anxiety," 2= "some anxiety," and 3= "strong anxiety." The other three scales are attitudes measured on a 5-point Likert scale with responses ranging from 1= "strongly agree" to 5= "strongly disagree." Some of the items were reverse coded so that higher scores on each of the scales indicate higher levels of anxiety and more negative attitudes and then all scales were computed by summing the average of each item and dividing by the number of items.

The three anxiety subscales, measured on a three-point Likert scale, are exam, interpretation, and asking for help anxiety. The Exam/Class Anxiety subscale contains 8 items that measure students' anxiety when taking a test or attending a class. Anxiety about interpreting statistical results, called Interpretation Anxiety, contains 11 items and Fear of Asking for Help is a 4-item scale that assesses anxiety around asking others for help with certain tasks.

The three attitudes sub-scales are perceptions about Fear of Statistics Teachers (5 items), Statistics Worth (16 items), and Computational Self-Concept (7 items), which are measured on a five-point Likert scale and reflect perceptions about statistics teachers, the usefulness of statistics, and their personal ability to do mathematical calculations. All Cronbach's alphas, for both T1 and T2 indicate good internal consistency, ranging from 0.82 to 0.94 for T1 and 0.80 to 0.94 for T2.

2.2.4. Change in Time in STARS Scales

For the portion of the analyses that examine change in these six constructs, we created change scores by subtracting the mean value of each T1 scale from the T2 value.

2.2.5. Gender

Gender is coded 1= Female and 0= Male. Students' gender was determined by their name and picture. It should be noted that this measure of gender is not ideal in that it only represents stereotypical presentations of gender and captures neither how students self-identify nor nonbinary gender identities.

2.2.6. Race/Ethnicity

Given that non-whites are shown to have higher rates of statistics anxiety (Onwuegbuzie 1998), we control for race. Race/Ethnicity is coded 1 = White and 0 = Nonwhite, as determined by their picture. As with gender, this measure of race/ethnicity does not reflect how students self-identify.

2.2.7. Semester

Semester is coded as an ordinal variable from 1 = Fall 2016 to 4 = Spring 2018. This variable controls for differences in the course and external factors in each semester, accounting for variation in anxiety/attitudes and grades for both factors we could not control, such as the weather and the structure of the classroom, as well as improvements the instructor made to the course as they gained experience with each semester.

2.3. Plan of Analysis

To test for mean differences, we used independent samples t-tests to compare STARS anxiety scores as a function of gender and dependent samples t-tests to assess change over time in STARS anxiety scores. For both sets of t tests, to account for differences in sample sizes between women and men, we do not assume equal variances.

All multi-level analyses were conducted using Hierarchical Linear Modeling (HLM), which is an appropriate approach given the nested nature of the data in that individuals' grades and anxiety at the end of the semester are not independent from their grades and anxiety earlier in the semester. In other words, individuals' later scores are nested within their earlier scores. HLM is a multi-level statistical technique that accounts for this nested data and allows us to examine, not only whether there are differences among individuals in how they change (Level 1 within-individual variation), but also if there are differences in change over time across individuals (Level 2 between-individual variation) (Scott, Simonoff, and Marx 2013).

In addition to change in exam grades, we also examine final exam grade as a dependent variable, as it may be of interest to statistics instructors to know where students end up in terms of their academic performance in order to inform pedagogy. In order to justify an HLM approach, we first analyze the unconditional model that includes only the dependent variable and we provide the intraclass correlation (ICC), which is a ratio of the proportion of within-individual variance (Level 1) to the between-individuals variance (Level 2); in a two-level model such as this one, the ICC is simply the percentage of the variance at Level 2, with values larger than 0.25 indicating the need for a multi-level approach (Guo 2005; Yuen and Santo 2018).

After we justify a nested, HLM approach by reporting the ICC, the within-individual predictor of time was included as a random effect. At the between-individual level we estimate, in a sequential fashion, six HLMs for each of the STARS constructs and for both of the dependent variables. Model 1 is the unconditional model that analyzes Level 1 within-individual variation in the dependent variable, either final exam grade or change in exam grades. Model 2 adds the STARS quadratic effect, which we calculated by squaring the STARS dimension at hand. Model 3 adds change in time in STARS as a predictor of grades and, finally, Model 4 adds the gender interactions.

All models control for gender, race/ethnicity, and semester. Only main effects that were statistically discernible (p <0.05) when originally entered into the model, having also reduced prediction error in the models by 1% or more, *and* improved the models (based on a Chi-square difference test) are interpreted and discussed below, but all coefficients are reported in the tables in Appendices A–F of the supplementary materials. Gender interactions are included herein if their addition reduced prediction error in the models by 1% or more and improved the models (again, based on a Chi-square difference test). Descriptive statistics were performed in SPSS version 24 and HLM analyses were conducted using the statistical modeling program HLM 8.

3. Results

3.1. Descriptive Statistics

The analytic sample of 111 students is primarily female (71%) and white (84%). Data came from across the four semesters relatively equally (ranging from 22% to 29%). As shown in Table 1, the averages for each of the three exams are 88.8 (SD = 9.5), 82.6 (SD = 11.1) and 72.8 (SD = 18.7).

In comparing the mean levels of each of the six STARS constructs as shown in Table 1, both women and men experience the greatest degree of exam anxiety and the worse attitudes toward the worth of statistics compared to the other two types of anxiety and the other two types of attitudes, respectively, at both T1 and T2. Also, for both T1 and T2, mean levels of the STARS constructs show that women report the lowest levels of anxiety regarding asking for help and the most positive attitudes toward statistics teachers. Men, however, report the least amount of anxiety concerning interpreting mathematical operations.

The results of the independent samples t-tests indicate several statistically discernible gender differences. At T1, women report greater exam (p <0.01) and interpretation anxiety (p <0.01), as well as more negative attitudes toward their self-concept (p <0.01) and the worth of statistics (p <0.01) than do men. At T2, there are no longer gender differences in exam anxiety and self-concept, but greater levels of interpretation anxiety ($p = 0.05$) and worth of statistics (p <0.01) among women remain.

The results of the paired samples t-tests shown in Table 1 show that, among the full sample, the decrease of exam, interpretation, and asking for help anxiety from T1 to T2 are all statistically discernible (Mean change= −0.2, −0.1, and −0.2, respectively), while attitudes toward self-concept, teachers, and statistics worth do not change. Among women, all three types of anxiety (Mean change = −0.2, −0.2, and −0.2, respectively) and attitudes toward self-concept (Mean change = 0.2) improve, but, among men, neither anxiety nor attitudes change from the beginning to the end of the semester.

3.2. HLM of Academic Performance

After examining the zero order correlations (see Table 2), we ran the unconditional model in which only the dependent variable, exam grades, is included. As shown in Table 3, a majority, 64%, of the variability was within-individuals (Level 1), whereas 36% of the variance (ICC = 0.36) was between-individuals (Level 2), constituting a significant portion of the total variability ($\chi^2_{(110)} = 294.1$, $p < 0.01$) and justifying the multi-level

Table 1. Descriptive statistics: *T*-tests by change in academic performance and statistics anxiety/attitudes and gender.

	Full Sample ($N = 111$)			Women ($n = 79$)			Men ($n = 32$)				
Academic Performance	Mean (SD)			Mean (SD)			Mean (SD)				
Exam I	88.84 (9.45)			88.73 (9.37)			89.06 (9.65)				
Exam II	82.57 (11.06)			82.78 (11.21)			81.88 (11.33)				
Exam III	72.76 (18.65)			75.75 (17.12)			68.05 (18.12)				
Anxiety	T1 Mean (SD)	T2 Mean (SD)	Mean Change[b]	T1 Mean (SD)	T2 Mean (SD)	Mean Change	T1 (SD)	T2 (SD)	Mean Change	T1 Gender p value[c]	T2 Gender p value
Exam Anxiety	2.03 (0.48)	1.85 (0.48)	0.17***	2.14 (0.45)	1.90 (0.42)	0.24***	1.75 (0.45)	1.74 (0.58)	0.01	0.00	0.17
Interpretation Anxiety	1.63 (0.35)	1.49 (0.32)	0.14***	1.71 (0.35)	1.53 (0.31)	0.18***	1.44 (0.28)	1.40 (0.31)	0.05	0.00	0.05
Asking for Help Anxiety	1.55 (0.55)	1.39 (0.47)	0.16***	1.58 (0.56)	1.37 (0.46)	0.21***	1.49 (0.53)	1.44 (0.49)	0.05	0.46	0.48
Attitudes[a]											
Statistics Teachers	1.84 (0.68)	1.89 (0.82)	−0.06	1.83 (0.78)	1.88 (0.75)	−0.05	1.84 (0.78)	1.92 (0.97)	−0.08	0.96	0.84
Statistics Worth	2.20 (0.73)	2.27 (0.78)	−0.07	2.33 (0.79)	2.36 (0.83)	−0.03	1.87 (0.62)	2.05 (0.63)	−0.18	0.00	0.04
Self-Concept	2.17 (0.80)	2.07 (0.76)	0.09	2.29 (0.81)	2.13 (0.77)	0.16*	1.86 (0.69)	1.94 (0.73)	−0.08	0.01	0.22

[a]Higher values indicate more negative attitudes. [b]Asterisks indicate statistically discernible mean change from Time 1 to Time 2, paired samples t-tests, two–tailed, $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. [c]p values are for gender differences, independent samples t-tests, two-tailed.

Table 2. Zero-order correlation in the between-individual variables.

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.
1. Gender	1.0														
2. Ethnicity	0.1	1.0													
3. Semester	−0.2	−0.1	1.0												
4. T1 Exam Anxiety	0.4**	0.1	−0.0	1.0											
5. T2 Exam Anxiety	0.2	0.1	0.0	0.5**	1.0										
6. T1 Interpretation Anxiety	0.3**	0.2	−0.1	0.6**	0.3**	1.0									
7. T2 Interpretation Anxiety	0.2*	−0.1	−0.3**	0.3**	0.5**	0.4**	1.0								
8. T1 Asking for Help Anxiety	0.1	0.0	−0.1	0.3**	0.1	0.2*	0.2	1.0							
9. T2 Asking for Help Anxiety	−0.1	0.0	0.1	0.1	0.4**	0.0	0.4**	0.6**	1.0						
10. T1 Statistics Teachers	0.0	0.1	0.1	0.2*	0.2*	0.2*	0.2	0.1	0.2*	1.0					
11. T2 Statistics Teachers	−0.0	0.1	−0.0	0.2	0.4**	0.0	0.3**	−0.0	0.4**	0.5**	1.0				
12. T1 Statistics Worth	0.3**	0.2*	0.1	0.2*	0.2	0.2	0.1	0.1	−0.0	0.5**	0.2	1.0			
13. T2 Statistics Worth	0.2*	0.2*	0.1	0.1	0.3**	0.0	0.2*	−0.0	0.1	0.4**	0.5**	0.7**	1.0		
14. T1 Self-Concept	0.3**	0.2	−0.0	0.4**	0.3**	0.2*	0.1	0.1	0.1	0.5**	0.4**	0.7**	0.5**	1.0	
15. T2 Self-Concept	0.1	0.2*	−0.1	0.2*	0.5**	0.1	0.3**	0.1	0.3**	0.4**	0.6**	0.5**	0.7**	0.7**	1.0

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

Table 3. Initial modeling of academic performance.

	Unconditional Model	Change over time	Covariates
	b (t)	b (t)	b (t)
Constant	81.94 (74.75)*	74.14 (45.50)*	69.80 (11.44)*
Gender			3.70 (2.23)*
Race (1 = white)			3.15 (.58)
Semester			.76 (.52)
PRPE (χ 2)		54.98% (168.74)*	2.60% (13.73)*
Grade Change		−7.60 (−10.19)*	−8.17 (−2.89)*
Gender			2.22 (2.86)*
Race (1 = white)			.50 (.19)
Semester			.09 (.13)
PRPE (χ 2)			9.73% (25.89)*

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

modeling approach adopted here (Guo 2005; Yuen and Santo 2018).

3.3. Linear & Quadratic Effects of STARS

Six sets of HLMs were run to test for the effects of each STARS anxiety/attitudes factor separately. All coefficients for the main effects, quadratic effects, change over time, and gender interactions for each STARS dimension separately are shown in the tables in Appendices A–F of the supplementary materials. Below we discuss only the effects that the HLM analyses showed were: (a) statistically discernible when originally entered into the model (p <0.05); (b) improved model fit as indicated by a statistically discernible chi square statistics at p <0.05; (c) and reduced the proportion of error (PRPE) by at least 1%.

We found different effects across each of the three STARS anxiety dimensions, in which the coefficients in the tables in Appendices A–C of the supplementary materials are shown in bold. For exam anxiety, among those whose exam anxiety increased, they scored lower in final exam grades ($b = -12.9, t = -2.4$). Interpretation anxiety was associated with lower final exam grades ($b = -3.7, t = -2.1$) and a sharper decrease in exam grades over time ($b = -1.53, t = -2.2$). Moreover, among those whose interpretation anxiety increased, they scored lower in final exam grades ($b = 12.5, t = -2.3$). Regarding fear of asking for help anxiety, those higher in fear showed a stronger decrease in exam grades over time ($b = -1.7, t = -2.4$). In addition, those whose fear of asking for help increased over the course of the semester had a sharper decrease of exam scores over time ($b = -8.9, t = -1.8$).

We also identified different effects across the three types of STARS attitudes, as shown in the tables in Appendices D–F of the supplementary materials. First, regarding fear of statistics teachers, greater fear was associated with lower final exam grades ($b = -4.8, t = -2.6$) and a sharper decrease in exam grades over time ($b = -1.5, t = -1.7$). Second, among those whose negative attitudes toward the worth of statistics increased over the course of the semester, they had lower final exam scores ($-5.9, t = -2.3$). Third, we found that those with negative attitudes toward their computational self-concept had lower final exam grades ($b = -2.9, t = -2.1$) and, among those whose self-concept worsened over time, they scored lower in final exam grades ($b = -7.6, t = -3.1$).

To aid in the interpretation of the only quadratic effect we found, that between computational self-concept and final exam grades ($b = 2.9, t = 2.3$), we constructed a figure to show exam scores of those with “very negative” (2 standard deviations above the mean), “negative” (1 standard deviation above the mean), “neutral” (at the mean), and “positive” (1 standard deviation below the mean) attitudes. As Figure 1 shows, both students with “positive” and “very negative” attitudes toward one’s computational self-concept perform better on final exams. That is, there is an inverse relationship between computational self-concept and final exam grades except among students with

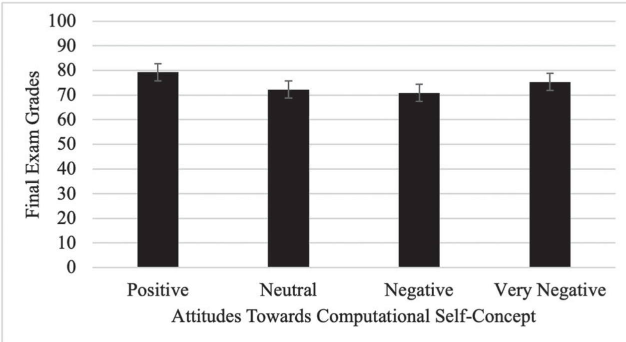

Figure 1. Quadratic Effect of Computational Self-Concept on Final Exam Grades (with 95% confidence intervals).

"very negative" attitudes, at which point the association turns positive.

3.4. Gender Interactions

Also shown in Appendices A-F are the results once gender is added to the model. Across the six STARS dimensions, we identified two statistically discernible interaction effects with gender that also improve model fit and the reduction in error, that of fear of asking for help and fear of statistics teachers (See bolded coefficients in Appendices C and D of the supplementary materials). First, men who score higher in asking for help anxiety scored lower on exams than women with similar levels of asking for help anxiety ($b = 2.5, t = 1.4$). Men whose fear of asking for help increased scored lower on the final exam than women whose fear increased at comparable rates ($b = 11.4, t = 3.0$). Second, men who scored higher in fear of statistics teachers scored lower on exams than women with similar levels of fear ($b = 2.3, t = 1.5$) and, among men whose fear of teachers increased, they scored lower on the final exam compared to women with similar levels of teacher fear ($b = 2.9, t = 1.4$).

To further illustrate how gender moderates the relationship between change in asking for help/fear of statistics teachers and final exam grades, we calculated "low," "average," "high," and "very high" levels of asking for help anxiety with one standard deviation below the mean, the mean, one standard deviation above the mean, and two standard deviations above the mean, respectively. As Figure 2 illustrates, two processes are at work regarding gender, fear of asking for help, and academic performance. Among men, increases in fear of asking for help are associated with lower final exam grades. Moreover, higher values of fear of asking for help is more detrimental on final exam grades among men compared to women.

Figure 3 shows final exam scores as a function of change in fear of statistics teachers by gender. Men who are higher in fear of statistics teachers scored lower than women with similar levels of teacher fear. Figure 3 also shows how men whose fear of statistics teachers increased over the course of the semester scored lower on the final exam than women whose fear similarly increased.

4. Discussion

4.1. Summary of Findings

The main objective of this study is to refine understandings of the relationship between statistics anxiety/attitudes and academic performance. Multi-level modeling allowed for us to test the main effect of each STARS scale score in addition to the quadratic effect. Above and beyond those effects, we tested for change over time and then whether there were gender differences in these effects. We found that, on average, exam grades for the full sample decreased over the course of the semester. Across the six STARS dimensions, however, and across gender, students' exams scores decreased at different rates and were not always linear. Taken together, results indicate that the relationship between statistics anxiety/attitudes and academic performance is not always negative, linear, stable over the course of the semester, or uniform across gender. The results of this study, therefore, challenge several assumptions that the majority of past studies' analytic approaches neglect or imply.

4.2. Effect of STARS on Academic Performance

We found that several of the STARS scales were associated with both exam grades and final exam grade in particular. Results for the main effects of the STARS dimensions on final exam grades indicate that those with high interpretation anxiety and negative attitudes toward statistics teachers and their computational self-concept perform less well on the final exam. In regards to the main effects of the STARS constructs on the degree of decrease of exam grades, we find that those who reported higher levels of interpretation and asking for help anxiety, in addition to worse attitudes toward statistics teachers, experience more of a drop in their exam grades over the course of a semester.

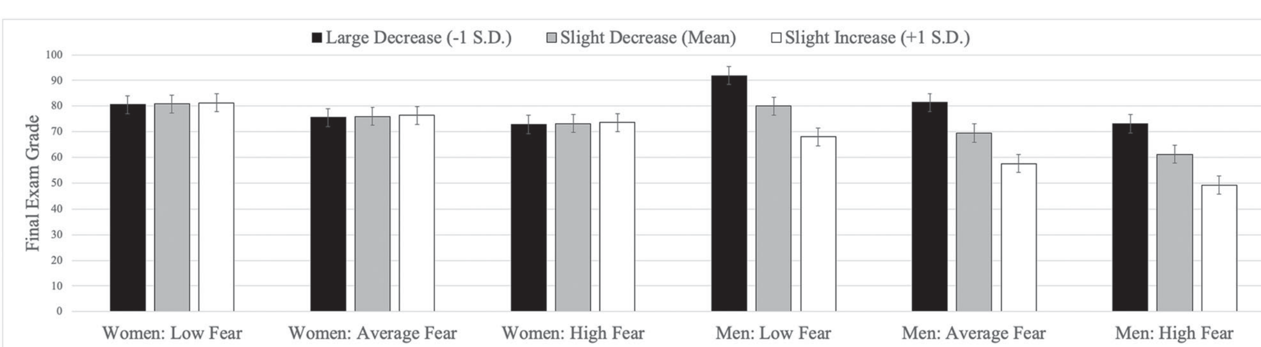

Figure 2. Change in Fear of Asking for Help and Final Exam Grades by Gender (with 95% confidence intervals).

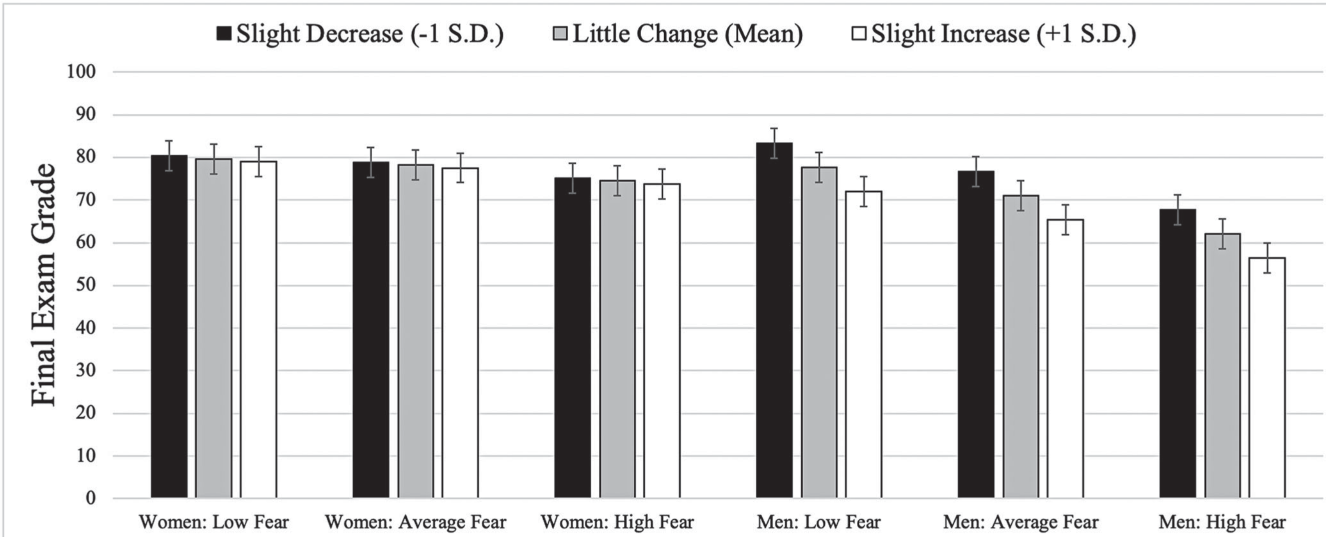

Figure 3. Change in Fear of Statistics Teachers and Final Exam Grades by Gender (with 95% confidence intervals).

The findings we report control for how overall the students' scores dropped over the semester and thus we can detail what predicts scores overall and whether or not scores were likely to decrease more or less. We therefore, deem the STARS aspects that are associated with both overall final exam scores and rate of decrease in exam scores as especially salient for academic performance. By this measure, of the three types of anxiety and three types of attitudes, interpretation anxiety and fear of statistics teachers are particularly important. These findings contradict most studies that employ the STARS scales that find that exam anxiety is the most influential for academic outcomes (Onwuegbuzie 1998; Onwuegbuzie and Wilson 2003; Haynes, Mullins, and Stein 2004). However, results here are consistent with studies that similarly show that anxiety around interpreting mathematical calculations is also an especially relevant type of statistics anxiety/attitude (Walsh and Ugumba-Agwunobi 2002; Onwuegbuzie and Wilson 2003; Williams 2010).

On the other hand, the finding that negative attitudes toward statistics teachers is particularly salient for academic performance is not in alignment with previous research using the STARS. In fact, for one of the exams, Keeley, Zayac, and Correia (2008) found that of the six STARS scales, only fear of teachers was not associated with exam grades. In one of the few studies, if not the only, to employ the STARS and explicitly focus on the important role of attitudes toward teachers in reducing statistics anxiety, Williams (2010) found that all six of the STARS anxiety/attitudes scales are related to various aspects of teacher immediacy (verbal and nonverbal behaviors that make people feel close to one another). This study and Williams (2010) suggest that the teacher-student relationship may be an effective focal point for anxiety-reducing interventions, such as increasing immediacy behaviors like smiling and addressing students by name (Williams 2010).

When change over time in the STARS scales were examined for their associations with change in exam grades throughout the semester, we found that as asking for help anxiety increased, there was a sharper decrease in exam grades. For final exam scores, those with more anxiety surrounding exams and asking for help, as well as negative attitudes toward the worth of statistics and computational self-concept, performed worse on the final exam. Following the same logic as above regarding assessing "importance" by their association with multiple outcomes, we argue that change in fear of asking for help is a particularly important aspect of change in statistics anxiety/attitudes to consider for academic performance. While prior research has established the importance of exam anxiety (Onwuegbuzie 1998; Onwuegbuzie and Wilson 2003; Haynes, Mullins, and Stein 2004), perceived worth of statistics (Zanakis and Valenzi 1997; Onwuegbuzie 2004; Liu, Onwuegbuzie, and Meng 2012), and self-concept (Townsend et al. 1998; Marsh and Martin 2011), the results of this analysis provide a new area of emphasis, that of fear of asking for help. Especially given that we find that both fear of teachers and fear of asking for help are associated with multiple outcomes, statistics instructors might want to consider ways to reduce those fears, such as finding a way for students to ask questions anonymously (Chew and Dillon 2014).

In the only study other than Keeley, Zayac, and Correia (2008) to directly test for quadratic effects, we identified one notable quadratic effect in that poor computational self-concept is associated with lower final exam grades, but at very low and very high levels of poor self-concept, this relationship reverses and has a positive effect on final exam grade (see Figure 2). Thus, while Keeley, Zayac, and Correia (2008) found that there was an optimal mid-level amount of test anxiety for exam grades, we found that the extreme levels of very positive and very negative attitudes toward self-concept are associated with better final exam grades. It is possible that those low in self-concept did better on the final exam because of their low confidence, rather than in spite of it. We also found that those whose negative attitudes toward their computational self-concept increased scored lower in final exam grades compared to those whose attitudes did not increase as much. Perhaps students' self-concept had been becoming increasingly worse because they had been doing increasingly poorly in the class thus far. When the stakes are the highest in terms of passing the class with the final exam, students who had the worst self-concept probably crammed and studied more and consequently did better on the final exam than those who did not study as much because they were more confident in their abilities. Those who were extremely high in their self-confidence, of course, were probably so because they indeed held the knowledge and skills necessary to do well on the final exam. Rather than trying to completely eliminate statistics anxiety, these findings suggest that statistics instructors may want to seek to identify what levels of anxiety are most detrimental for performance.

4.3. The Role of Gender

All students' exam scores, on average, went down with each exam. Women's and men's exam grades, however, decreased at different rates as a function of the six STARS dimensions. Women's exam scores did not decrease as much as men's did and they had higher final exam scores. It is notable that while women's confidence in their own abilities to do well in math and statistics classes is lower than men's in this study, as well as in the literature more broadly (Sax 1994; Campbell and Beaudry 1998; Correll 2001; Mills 2004; DeCesare 2007; Condron, Becker, and Bzhetaj 2018; MacArthur 2020), they actually do better in terms of grades. This is in alignment with other studies that show that the gap in confidence between women and men is not due to actual academic performance (Bradley and Wygant 1998; Correll 2001). Since it is not an objective ability, more research is needed to determine what accounts for women's lower self-concept as it pertains to math and statistics. On the one hand, this study shows that women's computational self-concept improves by the end of the semester. On the other hand, combatting women's internalization of the inaccurate stereotypes that characterize girls as less equipped for math and science fields than boys, will inevitably prove to be difficult, as these stereotypes are created and reinforced throughout the socialization process, education system, and beyond. Nonetheless, this study shows that women's self-concept is amenable to change and, regardless, they perform well academically compared to their male counterparts.

We find a few other notable gender interactions. First, men who have negative attitudes toward statistics teachers did substantially worse on exams than women with similar attitudes toward teachers. Second, men whose attitudes toward teachers worsened scored lower than women with comparable worsening of attitudes toward teachers (see Figure 3). Third, men whose fear of asking for help increased showed a sharper decline in exam grades over the course of the semester and, as a result, scored lower on the final exam (see Figure 1). Thus, having poor attitudes toward statistics teachers and more anxiety about asking for help is more salient for men than for women. Although it is speculation, perhaps ideals of masculinity that encourage restrictive societal norms around men's expression do not allow for men to ask for help and thus their relationships with teachers (and classmates) are hindered. As a result, men's grades suffer. In contrast, perhaps when women fear teachers or asking for help, there is more societal room to express and cope with those feelings and thus they do not let it affect their academic performance as much. These findings suggest that statistics instructors may want to pay particular attention to, and develop strategies to help overcome, male students' fears of them and asking for help.

4.4. Practical Implications

All together, the results of this multi-level analyses indicate that each of the six STARS dimensions are important for academic performance, although in different ways. First, what is important for exam grades overall may not be for improvement in exam grades. Second, change in each of the STARS constructs may be differentially related to exam scores overall and improvement in exam scores. Third, some degree of anxiety may positively affect academic performance. Given that the results vary across the six STARS scales and across main effects, change over the course of the semester, and linearity, this study calls attention to some of the STARS dimensions that typically do not receive as much scholarly focus, namely that of interpretation anxiety, fear of asking for help, and attitudes toward statistics teachers. Furthermore, this study reinforces the important role of gender in shaping the relationships between statistics anxiety/attitudes.

Findings here may aid statistics instructors who seek to improve their instruction by incorporating the recommendations in the GAISE College report (Carver et al. 2016). For example, both goals of easing exam anxiety, which was found to be particularly salient here, and of following the sixth GAISE recommendation to encourage student learning with assessments might be achieved with multiple low-stakes assessments rather than an emphasis on traditional-style exams. The second GAISE recommendation, to use "real data" to teach statistical concepts, might also help relieve interpretation anxiety or computational self-concept, both of which were found here to be associated with lower final exam scores (among other effects). Furthermore, reducing statistics anxiety might explain the efficacy of an implementation of a GAISE recommendation, but future research needs to test if there are mediational effects of statistics anxiety between pedagogical interventions to comply with GAISE recommendations and learning outcomes. In short, efforts to reduce statistics anxiety and improve attitudes may be complementary and mutually reinforcing with those suggested in the GAISE.

4.5. Limitations

Generalizability of these findings is compromised for several reasons, most notably because the data for this study were all collected in classes taught by the same professor of a single course within one discipline. Thus, it is impossible to disentangle the effects of one particular professor versus generalizable results. For example, it is possible that exam anxiety was not as prominent here as past studies have found (Onwuegbuzie 1998; Onwuegbuzie and Wilson 2003; Haynes, Mullins, and Stein 2004; Condron, Becker, and Bzhetaj 2018) because the instructor made several concerted efforts to reduce anxiety surrounding exams, including making a portion of them take-home; providing a formula/tables sheet; and allowing students to make a "cheat sheet" that they could use during the exam. Additionally, this was not a random sample and all data were collected at one urban Midwestern university in the United States that is not representative of all colleges/students. Socio-demographic variables that are known to be correlated with statistics anxiety, such as age and socio-economic status (Onwuegbuzie and Wilson 2003), were not collected. And, the operationalization of two of the socio-demographics, race/ethnicity and gender, do not reflect how students self-identify, which may be consequential for both statistics anxiety and academic performance.

In addition to socio-demographics, there are a few other potential correlates for which we were not able to control. Measures of prior academic achievement, such as reading ability or previous mathematics/statistics experience are associated with statistics anxiety and, therefore, could have altered results found here (Baloğlu 2003; Collins and Onwuegbuzie 2007, Zanakis and Valenzi 1997). Furthermore, even with the longitudinal data employed here, it is not possible to establish the causal order between academic performance and anxiety, which likely function in a perpetuating cycle in which poor grades exacerbate anxiety levels, which in turn negatively affect academic performance further. A final limitation is that there may be a selection effect in that those with the most severe anxiety may have dropped the course in week one before T1 data collection or before the last week of class at T2, which is supported by the missing data analysis that showed more anxious students and those with lower exam scores were more likely to be selected out of the sample. Conversely, levels of statistics anxiety may be *over*estimated compared to other studies because of the student population of this sample. Students at this metropolitan state college may be more likely to have anxiety around rigorous courses such as statistics compared to more socio-economically advantaged students (DeCesare 2007).

4.6. Future Research

Despite its limitations, this study fills several gaps in the literature on statistics anxiety and provides new suggested areas for which research on statistics anxiety to focus. First, this study challenges assumptions about the linearity of the relationships between different types of statistics anxiety/attitudes and academic performance. Despite a call made decades ago to examine whether there may be a moderate, optimal level of anxiety for academic performance (Onwuegbuzie and Wilson 2003), there remains a dearth of research that examines quadratic effects of statistics anxiety on grades. Future research should verify if, as was found here, high levels of anxiety indeed leads to improved grades. In either case, it appears that Keeley, Zayac, and Correia (2008) are correct in their assertion that “anxiety is not a fire that needs to be stamped out for students to be successful in a statistics class” (p. 13).

In addition to quadratic effects of statistics anxiety on academic performance, the second suggested new emphasis is that of change, as this study is, to our knowledge, the first study to explore how change in the six STARS scales over the course of the semester is associated with both academic performance overall and changes over time. In contrast to most previous studies, future research should neither assume that the effects of different types of statistics anxiety/attitudes are uniform nor that they are stable over the course of the semester. Furthermore, although gender is an important and appropriate focal point, if statistics instructors seek to develop pedagogical strategies to address statistics anxiety/attitudes, future research should work to identify other socio-demographic factors such as race/ethnicity, socio-economic status, and age that may account for changes over time in anxiety/attitudes and their effects on academic performance. This study also shows that there is variation in the effects of the STARS dimensions on different outcomes: final exam grades, exam grades overall, and change in exam grades over the course of the semester. Future research should, therefore, simultaneously look at multiple outcomes and, perhaps, effects of statistics anxiety beyond academic performance that may be of interest to instructors, such as student well-being. As was done with this study, and future research should confirm patterns identified here, the benefit of employing a longitudinal, multi-level analysis is that it highlights the multifaceted and dynamic nature of statistics anxiety/attitudes and academic performance.

Supplementary Materials

Appendices A–F provide the coefficients for the four multilevel models for each of the STARS constructs.

Disclosure Statement

No potential competing interest was reported by the authors.

References

Allison, P. D. (2001), *Missing Data* (Vol. 136). Thousand Oaks, CA: Sage Publications.

American Statistical Association. (2005), “Guidelines for Assessment and Instruction in Statistics Education.” Available at *http://www.amstat.org/education/gaise/index.cfm*.

Baloğlu, M. (2002), “Psychometric Properties of the Statistics Anxiety Rating Scale,” *Psychological Reports*, 90, 315–325.

Baloğlu, M. (2003), “Individual Differences in Statistics Anxiety Among College Students,” *Personality and Individual Differences*, 34, 855–865.

Benson, J. (1989), “Structural Components of Statistical Test Anxiety in Adults: An Exploratory Model,” *The Journal of Experimental Education*, 57, 247–261.

Benson, J., Bandalos, D., and Hutchinson, S. (1994), “Modeling Test Anxiety among Men and Women,” *Anxiety, Stress, and Coping*, 7, 131–148.

Bradley, D. R., and Wygant, C. R. (1998), “Male and Female Differences in Anxiety about Statistics are not Reflected in Performance,” *Psychological Reports*, 82, 245–246.

Bui, N. H., and Alfaro, M. A. (2011), “Statistics Anxiety and Science Attitudes: Age, Gender, and Ethnicity Factors,” *College Student Journal*, 45, 573–586.

Campbell, J. R., and Beaudry, J. S. (1998), “Gender Gap Linked to Differential Socialization for High-Achieving Senior Mathematics Students,” *The Journal of Educational Research*, 91, 140–147.

Carver, R., Everson, M., Gabrosek, J., Horton, N., Lock, R., Mocko, M., Rossman, A., Holmes Roswell, G., Velleman, P., Witmer, J., and Wood, B. (2016), Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016.

Chew, P. K., and Dillon, D. B. (2014), “Statistics Anxiety Update: Refining the Construct and Recommendations for a New Research Agenda,” *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 9, 196–208.

Collins, K. M., and Onwuegbuzie, A. J. (2007), “I Cannot Read My Statistics Textbook: The Relationship between Reading Ability and Statistics Anxiety,” *The Journal of Negro Education*, 76, 118–129.

Condron, D. J., Becker, J. H., and Bzhetaj, L. (2018), “Sources of Students’ Anxiety in a Multidisciplinary Social Statistics Course,” *Teaching Sociology*, 46, 346–355.

Correll, S. J. (2001), "Gender and the Career Choice Process: The Role of Biased Self-Assessments," *American Journal of Sociology*, 106, 1691–1730.

Cruise, R., Cash, R., and Bolton, D. (1985), "Development and Validation of an Instrument to Measure Statistical Anxiety," in ASA Proceedings of the Section on Statistical Education Section, pp. 92–97, Alexandria, VA: American Statistical Association.

DeCesare, M. (2007), "Statistics Anxiety" among Sociology Majors: A First Diagnosis and Some Treatment Options," *Teaching Sociology*, 35, 360–367.

Fitzgerald, S. M., Jurs, S. J., and Hudson, L. M. (1996), "A Model Predicting Statistics Achievement among Graduate Students," *College Student Journal*, 30, 361–366.

Guo, S. (2005), "Analyzing Grouped Data with Hierarchical Linear Modeling," *Children and Youth Services Review*, 27, 637–652.

Gyuris, E., & Everingham, Y. (2011), "Maths Anxiety over Two Campuses in a First Year Introductory Quantitative Skills Subject at a Regional Australian University – Establishing a Baseline," in: *Proceedings of the Australian Conference on Science and Mathematics Education*, eds. M. Sharma, A. Yeung, T. Jenkins, E. Johnson, G. Rayner and J. West, pp. 73–80, Sydney: UniServe Science.

Hanna, D., Shevlin, M., and Dempster, M. (2008), "The Structure of the Statistics Anxiety Rating Scale: A Confirmatory Factor Analysis using UK Psychology Students," *Personality and Individual Differences*, 45, 68–74.

Haynes, A. F., Mullins, A. G., & Stein, B. S. (2004), "Differential Models for Math Anxiety in Male and Female College Students," *Sociological Spectrum*, 24, 295–318.

Hedges, S. (2017), "Statistics Student Performance and Anxiety: Comparisons in Course Delivery and Student Characteristics," *Statistics Education Research Journal*, 16, 320–336.

Keeley, J., Zayac, R., and Correia, C. (2008), "Curvilinear Relationships between Statistics Anxiety and Performance among Undergraduate Students: Evidence for Optimal Anxiety," *Statistics Education Research Journal*, 7, 4–15.

Lesser, L. M., Pearl, D. K., and Weber, J. J. III, (2016), "Assessing Fun Items' Effectiveness in Increasing Learning of College Introductory Statistics Students: Results of a Randomized Experiment," *Journal of Statistics Education*, 24, 54–62.

Liu, S., Onwuegbuzie, A. J., and Meng, L. (2012), "Examination of the Score Reliability and Validity of the Statistics Anxiety Rating Scale," *The Journal of Educational Enquiry*, 11, 29–42.

MacArthur, K. R. (2020), "Avoiding Over-Diagnosis: Exploring the Role of Gender in Changes over Time in Statistics Anxiety and Attitudes," *Numeracy*, 13, 1–26.

Macher, D., Paechter, M., Papousek, I., Ruggeri, K., Freudenthaler, H. H., and Arendasy, M. (2013), "Statistics Anxiety, State Anxiety during an Examination, and Academic Achievement," *The British Journal of Educational Psychology*, 83, 535–549.

Marsh, H. W., and Martin, A. J. (2011), "Academic Self-concept and Academic Achievement: Relations and Causal Ordering," *The British Journal of Educational Psychology*, 81, 59–77.

McGrath, A. L., Ferns, A., Greiner, L., Wanamaker, K., & Brown, S. (2015), "Reducing Anxiety and Increasing Self-efficacy within an Advanced Graduate Psychology Statistics Course," *Canadian Journal for the Scholarship of Teaching and Learning*, 6, 1–17.

Mellanby, J., and Zimdars, A. (2011), "Trait Anxiety and Final Degree Performance at the University of Oxford," *Higher Education*, 61, 357–370.

Mills, J. D. (2004), "Students' Attitudes toward Statistics: Implications for the Future," *College Student Journal*, 38, 349–362.

Onwuegbuzie, A. J. (1995), "Statistics Test Anxiety and Female Students," *Psychology of Women Quarterly*, 19, 413–418.

Onwuegbuzie, A. J. (1998), "The Dimensions of Statistics Anxiety: A Comparison of Prevalence Rates among Mid-Southern University Students," *Louisiana Educational Research Journal*, 23, 23–40.

Onwuegbuzie, A. J. (2003), "Modeling Statistics Achievement among Graduate Students," *Educational and Psychological Measurement*, 63, 1020–1038.

Onwuegbuzie, A. J. (2004), "Academic Procrastination and Statistics Anxiety," *Assessment & Evaluation in Higher Education*, 29, 3–19.

Onwuegbuzie, A. J., and Seaman, M. A. (1995), "The Effect of Time Constraints and Statistics Test Anxiety on Test Performance in a Statistics Course," *The Journal of Experimental Education*, 63, 115–124.

Onwuegbuzie, A. J., and Wilson, V. A. (2003), "Statistics Anxiety: Nature, Etiology Antecedents, Effects, and Treatments – A Comprehensive Review of the Literature," *Teaching in Higher Education*, 8, 195–209.

Papousek, I., Ruggeri, K., Macher, D., Paechter, M., Heene, M., Weiss, E. M., Schulter, G., and Freudenthaler, H. H. (2012), "Psychometric Evaluation and Experimental Validation of the Statistics Anxiety Rating Scale," *Journal of Personality Assessment*, 94, 82–91.

Rosenfeld, R. A. (1978), "Anxiety and Learning," *Teaching Sociology*, 5, 151–166.

Sax, L. J. (1994), "Mathematical Self-Concept: How College Reinforces the Gender Gap," *Research in Higher Education*, 35, 141–166.

Scott, M. A., Simonoff, J. S., and Marx, B. D., Eds. (2013), *The SAGE Handbook of Multilevel Modeling*, Thousand Oaks, CA: Sage.

Stout, J. A., Dasgupta, N., Hunsinger, M., and McManus, M. A. (2011), "STEMing the Tide: Using Ingroup Experts to Inoculate Women's Self-Concept in Science, Technology, Engineering, and Mathematics (STEM)," *Journal of Personality and Social Psychology*, 100, 255–270.

Townsend, M. A., Moore, D. W., Tuck, B. F., and Wilton, K. M. (1998), "Self-Concept and Anxiety in University Students Studying Social Science Statistics within a Co-Operative Learning Structure," *Educational Psychology*, 18, 41–54.

Van Gundy, K., Morton, B. A., Liu, H. Q., and Kline, J. (2006), "Effects of Web-Based Instruction on Math Anxiety, the Sense of Mastery, and Global Self-Esteem: A Quasi-Experimental Study of Undergraduate Statistics Students," *Teaching Sociology*, 34, 370–388.

Walsh, J. J., and Ugumba-Agwunobi, G. (2002), "Individual Differences in Statistics Anxiety: The Roles of Perfectionism, Procrastination and Trait Anxiety," *Personality and Individual Differences*, 33, 239–251.

Williams, A. S. (2010), "Statistics Anxiety and Instructor Immediacy," *Journal of Statistics Education*, 18, 1–18.

Williams, A. S. (2013), "Worry, Intolerance of Uncertainty, and Statistics Anxiety," *Statistics Education Research Journal*, 12, 48–59.

Wisenbaker, J. M., Scott, J. S., and Nasser, F. (2000), "Structural Equation Models Relating Attitudes about and Achievement in Introductory Statistics Courses: A Comparison of Results from the US and Israel," in *9th International Congress on Mathematics Education*, Tokyo, Japan.

Wismath, S. L., and Worrall, A. (2015), "Improving University Students' Perception of Mathematics and Mathematics Ability," *Numeracy*, 8, 1–17.

Yuen, N. C., and Santo, J. B. (2018), "Multilevel Modeling: Applications for Life- Span Research," in: *The SAGE Encyclopedia of Lifespan Human Development*, ed. M. Borstein, pp. 1456–1459, Thousand Oaks, CA: SAGE Publications.

Zanakis, S. H., and Valenzi, E. R. (1997), "Student Anxiety and Attitudes in Business Statistics," *Journal of Education for Business*, 73, 10–16.

Zeidner, M. (1990), "Does Test Anxiety Bias Scholastic Aptitude Test Performance by Gender and Sociocultural Group?" *Journal of Personality Assessment*, 55, 145–160.

Zeldin, A. L., & Pajares, F. (2000), "Against the Odds: Self-Efficacy Beliefs of Women in Mathematical, Scientific, and Technological Careers," *American Educational Research Journal*, 37, 215–246.

Zettle, R. D., and Raines, S. J. (2000), "The Relationship of Trait and Test Anxiety with Mathematics Anxiety," *College Student Journal*, 34, 246–259.

Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K., & Chang, B. (2008), "What Does Research Suggest about the Teaching and Learning of Introductory Statistics at the College Level? A Review of the Literature," *Journal of Statistics Education*, 16, 1–26.