ASSESSING STATISTICS ANXIETY IN AN ONLINE OR HYBRID SETTING: THE ADAPTION AND DEVELOPMENT OF A NEW INSTRUMENT SASOH

LU LIU
University of La Verne
lliu2@laverne.edu

ABSTRACT

With the purpose of developing an instrument for measuring statistics anxiety in the online or hybrid setting, this study tested the newly developed instrument in two stages. Results on item selection and exploratory factor analysis based on pilot testing (n = 115) are presented. Results on classical item analysis, the confirmatory factor analysis, the measurement invariance test results, and the predictive and discriminant validity of the final model based on formal testing (n = 709) are presented. The resulting Statistics Anxiety Scale in the Online or Hybrid setting instrument (SASOH) has 27 items and four dimensions. The four dimensions are Class and Interpretation Anxiety (CI), Fear of Asking for Help Anxiety (FA), Online System Anxiety (OS), and Pre-Conception Anxiety (PC). The results of the confirmatory factor analysis revealed that the four-factor SASOH model represents an adequate description of statistics anxiety in an online or hybrid setting. Moreover, multiple-groups confirmatory factor analysis affirmed that the resulting model achieved at least partial measurement and structural invariance across gender and program. In addition, attitudes toward statistics significantly predicts the four factors of statistics anxiety, and the discriminant validity from mathematics anxiety was confirmed. Recommendations for future studies are also provided.

Keywords: *Statistics education research; Statistics anxiety; Online or hybrid education; Instrument development; Factor analysis; Multiple-groups CFA*

1. INTRODUCTION

Statistics anxiety is broadly defined as “the feelings of anxiety encountered when taking a statistics course or doing statistical analyses” (Cruise et al., 1985, p. 92). Zeidner (1991) defined it as a type of performance anxiety characterized by “extensive worry, intrusive thoughts, mental disorganization, tension, and physiological arousal” (p. 319). To further clarify statistics anxiety as a construct distinct from mathematics anxiety and attitudes toward statistics, Chew and Dillon (2014) suggested redefining statistics anxiety as:

> a negative state of emotional arousal experienced by individuals as a result of encountering statistics in any form and at any level; this emotional state is preceded by negative attitudes toward statistics and is related to but distinct from mathematics anxiety. (p. 199)

Past studies have examined statistics anxiety through four general approaches: (a) treating anxiety as a covariate, (b) treating anxiety as a predictor variable, (c) treating anxiety as an outcome variable, and (d) designing intervention strategies to deal with anxiety. As a covariate, statistics anxiety has been found to be associated with many motivational beliefs such as self-efficacy and intrinsic goal orientation (Baloğlu et al., 2017; Razavi et al., 2017) or intolerance of uncertainty and worry (Williams, 2013). As a predictor, statistics anxiety is often used to predict statistics achievement or learning difficulties (Hanna & Dempster, 2009; Lin et al., 2016; Onwuegbuzie & Seaman, 1995). Treating statistics anxiety as an outcome variable, researchers have found many antecedents of statistics anxiety such as the Big Five personality factors (Chew & Dillon, 2014), attentional bias (Chew et al., 2017), attitude toward statistics (Kinkead et al., 2016), and mathematics anxiety (Mji, 2009). With the goals of reducing statistics anxiety and increasing student performance, many interventions have been created, such as

Statistics Education Research Journal, 20(1), Article 2. https://doi.org/10.52041/serj.v20i1.96

the one-minute strategy (Chiou et al., 2014), active learning strategies (Rapp-McCall & Anyikwa, 2016), and innovative teaching methods (Einbinder, 2014).

Past research has asserted that nearly three quarters of the college students have some degree of statistics anxiety when taking statistics courses or training and students deem statistics courses as a major obstacle to degree attainment (Kinkead et al., 2016; Onwuegbuzie, 2004). The conclusions, however, of the above studies are based on the validity of existing instruments for measuring statistics anxiety. Although many studies have demonstrated validity and reliability evidence for existing instruments in measuring statistics anxiety in a traditional face-to-face setting, the application of such instruments in an online or hybrid setting has been largely ignored (Liu & Haque, 2017).

1.1. REVIEW OF EXISTING INSTRUMENTS

The statistical anxiety rating scale (STARS) developed by Cruise et al. (1985) is the most widely used measure for assessing statistics anxiety. The authors utilized principal components analysis and identified six components of statistics anxiety: (a) test and class anxiety, (b) interpretation anxiety, (c) fear of asking for help, (d) worth of statistics, (e) computational self-concept, and (f) fear of statistics teachers. Test and class anxiety concerns the anxiety involved in attending a statistics class or taking a test. Interpretation anxiety refers to the feelings of anxiety encountered when interpreting statistical data. Asking for help anxiety relates to the anxiety experienced when seeking help. Worth of statistics refers to a student’s attitude to the relevance of statistics to their study and future career. Computational self-concept relates to students’ self-evaluation of their ability to do mathematics. Fear of statistics teachers refers to students’ perception of their statistics instructors. The first three components measure statistics anxiety, and the next three components measure attitudes toward statistics. Together, the instrument has 51 items, and each item is on a 5-point Likert scale. Several studies have demonstrated evidence of validity and reliability of the instrument. For example, Baloğlu (2002) tested the concurrent validity and internal validity of the scale with 221 college students, and Hsiao (2010) supported the multidimensional construct validity of the scale.

The other widely used instrument for assessing statistics anxiety is the Statistics Anxiety Scale (SAS), which has three subscales: (a) examination anxiety, (b) asking for help anxiety, and (c) interpretation anxiety. The SAS contains several items adapted from the STARS scale with the goals of assessing only the three sub-scales for statistics anxiety and shortening the scales (Vigil-Colet et al., 2008). The definitions of the three dimensions are very similar to the three dimensions of the STARS measuring statistics anxiety. The SAS scale has a total of 24 items, with eight in each of the three sub-scales.

There are several other instruments measuring statistics anxiety. Some examples are the Statistics Anxiety Inventory (SAI), the Student Anxiety and Attitudes in Business Statistics Scale (SAAIBS), and the Statistics Anxiety Measure (SAM). The SAI was modified from a mathematics scale and has 40 Likert-scale items (Zeidner, 1991). These items were separated into two dimensions: anxiety about statistics content and anxiety about statistics performance and problem-solving. The items in the anxiety about statistics content dimension cover the components of class anxiety and interpretation anxiety. The anxiety about statistics performance and problem-solving is similar to the test anxiety in STARS, which refers to the anxiety related to course performance such as worrying about taking the exams and quizzes. The SAAIBS has 36 items, divided into six dimensions: (a) student interest in and perceived worth of statistics, (b) anxiety when seeking help for interpretation, (c) computer usefulness and experience, (d) understanding, (e) test anxiety, and (f) mathematics anxiety (Zanakis & Valenzi, 1997). The computer usefulness and experience anxiety refers to the anxiety the student encountered when using a computer or statistical software to solve statistics problems. The SAM is a 44-item instrument developed with the goal of unifying the common dimensions noted in previous research. It has four sub-scales: (a) anxiety, (b) class, (c) mathematics, and (d) performance (Earp, 2007). The above is not an exhaustive list of existing instruments measuring statistics anxiety, rather they are the common ones. Other examples include the Statistics Anxiety Scale (SAS), which is another example of a modified mathematics scale, (Betz,1978; Pretorius & Norman, 1992) and the statistics anxiety scales at the input, processing, and output stages of learning statistics (Whitcome, 2004)

In sum, statistics anxiety is commonly conceptualized as a multidimensional construct. Based on the content of the items in the existing instruments, four common dimensions are presented:

Class/content Anxiety, Interpretation Anxiety, Asking for Help Anxiety, and Performance Anxiety (Table 1). Because researchers have suggested that it is important to distinguish statistics anxiety from attitudes toward statistics (Chew & Dillon, 2014; Chiesi & Primi, 2010), only the dimensions measuring statistics anxiety are considered in this study. Please note that not every instrument has all of the four dimensions and not all dimensions across the instruments are included in Table 1.

Table 1. Common Dimensions in the Existing Instruments

Dimensions	Instrument	Example Items
Class/Content Anxiety	STARS SAI SAAIBS SAM	Doing the coursework for a statistics course. I am worried about taking statistics.
Interpretation Anxiety	STARS SAI SAAIBS SAM SAS	Interpreting the meaning of a table in a journal article. Seeing a classmate carefully studying the results table of a problem he has solved.
Asking for Help Anxiety	STARS SAAIBS SAM SAS	Asking a private teacher to explain a topic that I have not understood at all. Going to the teacher’s office to ask questions.
Performance/Test/Examination Anxiety	STARS SAI SAAIBS SAM SAS	Walking into the room to take a statistics test. Realizing the day before an exam that I cannot do some problems that I thought were going to be easy.
Others/Miscellaneous		
Computer Usefulness and Experience	SAAIBS	Bad experience using a computer.
Attitudes Towards Math/Math Anxiety	SAM SAAIBS	Math is my least favorite subject.

1.2. THE NEED FOR A NEW INSTRUMENT

In Fall 2016, approximately 6 million students enrolled in at least one online course, and nearly 31.7% of students enrolled in degree-granting postsecondary institutions took at least one distance education course (U.S. Department of Education, 2018). Moreover, many universities offer free or low-cost online statistics courses for the public. For example, the public can access over 2,500 online courses from 140 top universities in edX, an online learning platform founded by Harvard University and Massachusetts Institute of Technology (https://www.edx.org/about-us). Currently, the platform offers hundreds of online statistics courses. Studies have documented that students in an online or hybrid setting have an even less favorable attitude toward statistics than students in traditional face-to-face programs and they procrastinate in the enrollment of statistics courses due to anxiety (DeVaney, 2010; Xu & Jaggars, 2014). Therefore, the diagnosis of statistics anxiety of students in an online or hybrid setting becomes even more critical.

Several items in the current instruments are not suitable for measuring statistics anxiety in an online or hybrid setting. For example, statistics anxiety items such as “Walking into the room to take a statistics test” and “Going to the teacher’s office to ask questions” are suitable for face-to-face programs, but not for online or hybrid programs that have few or no face-to-face interactions because physical classrooms or teachers’ offices are not available in such settings. Similarly, items such as “Asking someone in the computer lab for help in understanding a printout” or “Watching a student search through a load of computer printout from his/her research” may not be feasible either as computer lab and face-to-face interaction with other students are not available in an online or hybrid setting. Such contextual inappropriateness of these items in an online or hybrid environment has been pointed out in the few studies that attempted to use the existing instruments such as STARS among online graduate students (e.g., DeVaney, 2016; Hsu et al., 2009). Without an instrument measuring statistics anxiety in such a setting, however, past studies could only either avoid using these items or take out the wording related to “classroom,” “office,” “computer lab,” or “exam.” Simply removing those words or the whole items would probably dimininsh the evidence for overall validity and reliability of the instruments.

Moreover, the assessment in professional graduate degree programs is often not the same as the assessment in an undergraduate program. The current statistics anxiety instruments have been designed and validated using factor analysis, using undergraduate students, mainly in psychology programs. For example, 159 undergraduate students enrolled in a statistics course in a Psychology Department in Spain were the sample for the SAS development (Vigil-Colet et al., 2008), and 650 undergraduate psychology students in the UK were the sample to validate the STARS instrument (Hanna et al., 2008). The applicability of the current instruments in a broad range of programs, especially for the students in graduate or professional degree programs, has largely not been assessed. In the professional degree programs, the course emphasis may be more focused on the practical application of statistical knowledge including article critique, class discussions, class presentation, data analysis, and research projects instead of examinations. The same may be true of a variety of online statistics training programs offered by university extensions or organizations that may not even have quizzes or exams. Therefore, the current statistics anxiety subscale of test or examination anxiety with items such as “waking up in the morning on the day of a statistics test” and “going over a final examination in statistics after it has been marked” may not suitable for this emphasis either.

At the same time, students who enroll in online courses also have a unique set of concerns different from those of students in a traditional setting. Many students deem the online courses as a computer automated system lacking personal interactions with instructors and peers (Boettcher & Conrad, 2010). As a result, some students come to the online courses with the pre-conceived assumptions of online learning as self-taught or self-regulated learning in which the students are obliged to take great responsibilities for their own learning (Schulze, 2009; Tichavsky et al., 2015). For the presumptuously difficult subjects such as statistics, such assumptions almost certainly increase student anxiety (Xu & Jaggars, 2014). Therefore, the new instrument must consider the unique characteristics of online learning.

Learning complex concepts such as statistics can be challenging. When more and more students enroll in hybrid or online education, the lack of instruments measuring statistics anxiety with evidence of validity and reliability in an online and hybrid setting cannot continue being ignored. Before interventions can be designed to help these students, a tool to measure the attributes of statistics anxiety among students in hybrid or online education is needed, especially in the graduate or professional degree programs where the statistical applications are emphasized.

2. PURPOSE

The purpose of this study was to develop an instrument for measuring statistics anxiety in an online or hybrid setting, SASOH, based on the *Standards for Educational and Psychological Testing* (2014) published by the joint committee of the American Educational Research Association, the American Psychological Association, and the National Council on Measurement in Education. The new instrument was designed to be used for statistics courses offered in an online or hybrid setting with no or few face-to-face interaction opportunities among the instructors and students and that emphasize statistical applications with few or no examination components. The predictive relationship from attitudes toward statistics and the discriminant validity from mathematics anxiety were assessed as well. The proposed instrument can be used for diagnostic, classification, progress, and modification-of-instruction purposes. Targeted treatment can be implemented based on the diagnosis, classification, and monitoring of students (DeVaney, 2010).

3. METHOD

3.1. PROCEDURE

The Standards for Educational and Psychological Testing published by the American Educational Research Association (AERA), the American Psychological Association (APA), and the National

Council on Measurement in Education (NCME) (2014) were used as the foundation for developing the statistics anxiety scale in the online or hybrid setting instrument (SASOH). Two stages were used in this work: The pilot testing stage included analysis of the participant characteristics, the item creation process, the item selection and exploratory factor analysis results, and the conceptual factor model. The formal testing stage included analysis of the participant and item characteristics, the classical item analysis, the confirmatory factor analysis, the measurement invariance test results, and the predictive and discriminant validity of the SASOH model.

In both stages, the same criteria were used to select the participants: (a) the participants were 18 years old or older, (b) the participants were currently enrolled in a higher education institution as a part-time or full-time undergraduate or graduate student, and (c) the participants were taking or had taken a statistics course in an online or hybrid setting. In comparison, the participants for the pilot testing stage were more homogeneous and taken from two universities, and the participants for the formal testing stage were more heterogenous as the sample was recruited from the whole U.S. using the Qualtrics data collection service.

At the pilot testing stage, 115 participants in hybrid graduate programs from two universities were included (n = 115). Sixty-three students were enrolled in a hybrid quantitative research method course and fifty-two students were enrolled in an online statistics course. Out of the 115 students in the pilot sample, 113 completed all items of the survey including the demographics information. The sample was composed of 28 males (24.3%) and 86 females (74.8%). The ethnicity distribution was 37 Whites (32.2%), 35 Hispanics (30.4%), 12 Asians (10.4%), 12 students in the "Two or more races" group (10.4%), and 19 students in the "Black or African-American" or "Other" category. In terms of age, the largest percentage of the sample was in the "30-49 years old" group (52.2%), about a third of the sample was in the "25-29 years old" group (33.9%), and 12.2 percent of the sample was in the "50-64 years old" group. Although the number of participants satisfies the sample size recommendation of at least 100 subjects needed for the factor analysis (Gorsuch, 1983; Kline, 1994), the author acknowledges the relatively small subject to item ratio at this stage of the study.

At Stage II using a ratio of number of participants to number of items of 20 to 1 as the guideline (Hair et al., 1995), a sample size of more than 540 was needed for the CFA. Therefore, a sample size of 540 or more was aimed for in the second instrument testing stage. At this stage, a total of 633 students were recruited from the U.S. using the data collection service provided by Qualtrics. Qualtrics is an enterprise survey technology company, and it has been providing online samples for over five years (https://www.qualtrics.com/research-services/). In the comparison studies between the traditional data collection methods and the ones using commercial platforms such as Qualtrics, Survey Monkey, and Mechanical Turk, researchers have shown that using these platforms is an efficient and affordable method to collect national representative samples (Boas et al., 2020; Heen et al., 2014). The Qualtrics research service was selected because of its large network of data collection agents and its quality validation process. For example, the Qualtrics research team checked the time participants used to fill in the survey to eliminate the speeders (e.g., respondents who took far less time in completing the survey relative to other respondents) and created certain demographic quotas to ensure a balanced sample based on the study needs. To reach the target sample of this study, pre-qualification questions were selected carefully to ensure that only participants who satisfied the selection criteria were allowed to complete the survey. Then the survey was soft launched with only 10% of the target sample size first for an opportunity to catch any issues with the survey design. During this step, the Qualtrics research team sent the initial results to review the participants' characteristics and whether they satisfied the sample selection criteria. After confirmation, the survey was fully launched in the following three weeks. Each week, an update of the data collection progress report was provided, and the quality of the data was reviewed. At the beginning of the last week of the data collection, a gender quota was used to collect more data from male respondents to ensure a more balanced sample because 75% of the respondents were females at that point of the data collection. Additionally, in addition to the 633 students, another group of 76 graduate students in a hybrid statistics course were recruited from a local university. Together, the sample size was 709 (n = 709) for this study at Stage II. See Table 2 for the demographic profile of the participants at both stages.

Table 2. Demographic profiles for the two stages of study

	Stage I Pilot Testing (n = 115)		Stage II Formal Testing (n = 709)	
	n	%	n	%
Gender				
Male	28	24.3%	288	40.6%
Female	86	74.8%	408	57.5%
Other			5	0.7%
Missing	1	0.9%	8	1.1%
Ethnicity				
Asian	12	10.4%	57	8.0%
Black or African American	8	7.0%	119	16.8%
Hispanic of any race	35	30.4%	90	12.7%
White	37	32.2%	330	46.5%
Two or more races	12	10.4%	29	4.1%
Other	11	9.6%	36	5.1%
Missing			48	6.8%
Age				
18-24 years old			180	25.4%
25-29 years old	39	33.9%	161	22.7%
30-49 years old	60	52.2%	275	38.8%
50-64 years old	14	12.2%	43	6.1%
65 years and over			2	0.3%
Missing	2	1.7%	48	6.8%
Program				
Traditional Undergraduate			66	9.3%
Traditional Graduate			53	7.5%
Hybrid Undergraduate			192	27.1%
Hybrid Graduate	115	100%	256	36.1%
Online Undergraduate			73	10.3%
Online Graduate			64	9.0%
Missing			5	0.7%

Out of the 709 students in the final sample (n = 709), 91.8% completed all items of the survey and the rest completed at least 70% of all items. The sample was composed of 288 males (40.6%), 408 females (57.5%), and 5 in the "other" category (0.7%). Also, it was composed of 57 Asians (8.0%), 119 African-Americans (16.8%), 90 Hispanics (12.7%), 330 Whites (46.5%), and 65 in the "Two or more races" or "Other" category (9.1%). In terms of age, the largest group in the sample was the "30-49 years old" group (38.8%), about a quarter of the sample was in the "18-24 years old" (25.4%), a little less than a quarter was in the "25-29 years old" group (22.7%), and less than 7 percent of the sample was in the combination of "50-64 years old" and "65 years and over" groups (6.4%).

Additionally, the information about the subjects' program and their experiences with statistics courses and the online format were collected. A total of 119 students were currently enrolled in a traditional face-to-face program (16.8%), 448 students were enrolled in a hybrid program (63.2%), and 137 students were enrolled in an online program (19.3%), but all of them were taking or had taken a statistics course in a hybrid or online setting based on the selection criteria. A little less than 50% of the subjects were enrolled in a part-time or full-time undergraduate program (n = 331) and a little more than 50% were enrolled in a part-time or full-time graduate program (n = 373). In terms of the number of statistics courses they had taken, about half had only taken 1 or 2 courses (n = 401), a little over a quarter had taken 3 to 5 courses (n = 195), 11% had taken 6 to 10 courses (n = 76), and 5% had taken 11 or more courses (n = 33). In terms of the number of online or hybrid courses that they had taken, 39.0% had only taken 1 or 2 courses (n = 277), 31.6% had taken 3 to 5 courses, (n = 224), 18.3% had taken 6 to 10 courses (n = 130), and 10.3% had taken 11 or more courses (n = 73). The average level of experience in online learning and computer technology is 3.84 (SD = 0.87), which is in between "neutral" and "experienced" in a 5-category Likert-scale from "extremely inexperienced" to "extremely experienced."

3.2. ANALYTIC APPROACH

The extent to which the proposed items measure statistics anxiety in an online or hybrid setting was examined using Mplus v. 8.2. (Muthén & Muthén, 1998/2017). In the pilot testing stage, descriptive statistics and exploratory factor analysis (EFA) were used to select the items for the Stage II testing. In the formal testing stage, classical item analysis, confirmative factor analysis (CFA), and multiple-groups CFA were conducted to check if the resulting model represents an adequate description of statistics anxiety in an online or hybrid setting and to evaluate if the model is measurement invariant between males and females and between undergraduates and graduates. The model fit indices including chi-square, comparative fit index (CFI), Tucker-Lewis index (TLI), root mean square error of approximation (RMSEA), and standardized root mean squared residual (SRMR) were used for the criteria to decide on the fit of the models. The cutoff scores for both CFI and TLI are 0.95 or higher, for RMSEA is 0.06 or lower, for SRMR is 0.08 or lower (Hu & Bentler, 1999).

After the final four-factor SASOH model was accepted, the across-group equivalence of the models in program groups (undergraduates vs. graduates) were evaluated. As one of the two methods used to evaluate the measurement invariance and population heterogeneity, multiple-groups CFA has several advantages over the Multiple Indicators Multiple Causes (MIMIC) modeling because it can examine all aspects of invariance and heterogeneity (Brown, 2015). In testing the multiple-groups CFA invariance, the steps laid out in Bowen and Masa (2015) and Brown (2015) were followed. In detail,

Step 1. Adequate fit of the CFA model in each group, tests the CFA model for each group separately and ensures that the four-factor SASOH model is acceptable for both males and females;

Step 2. Equal form or Configural invariance model, tests equal factor structures, in which the number of factors and pattern of indicator-factor loadings are equivalent across groups;

Step 3. Equal factor loadings or Metric invariance model, tests the equivalent factor loadings between the two groups;

Step 4. Equal item intercepts or Scalar invariance model, tests the equivalence of item intercepts between groups;

Step 5. Equal residual variance invariance or Strict factorial invariance, tests whether the two groups have the same residual variances;

Step 6. Equal factor variance, tests whether the factor variance is the same between two groups;

Step 7. Equal factor means, tests whether the factor means are the same across the groups. Steps 2-5 are the tests for measurement invariance and Steps 6-7 are the tests for structural invariance.

Because the weighted least square mean and variance (WLSMV) was used as the estimator, rescaled likelihood ratio tests were conducted to compare the nested models in the above steps by using the DIFFTEST function in Mplus. Moreover, Modification Indices were checked for any modifications needed.

4. RESULTS

4.1. STAGE I: PILOT TESTING OF THE ITEMS

Expert Feedback and Item Development Procedure The initial items for the pilot study were created from four sources: (a) literature review, (b) items from existing instruments, (c) syllabi of online or hybrid statistics course or training, and (d) expert feedback. The literature review was not only from the quantitative studies, but also qualitative studies in which the researchers interviewed the students in the online or hybrid statistics courses about their anxiety of these courses (e.g., DeVaney, 2010; Kinkead et al., 2016; Malik, 2015). Based on the past literature, it is clear that statistics anxiety is a multi-dimensional construct. Therefore, the items were created based on the related dimensions. Several items from existing instruments, especially the STARS, were adopted or modified in the pilot testing. The author also searched the online or hybrid statistics course syllabi available online and checked the course content and assessment methods for some commonalities.

Five experts who are professors in the statistics and quantitative research methods field were consulted in the pilot testing stage of constructing the scale. The experts have extensive experience in teaching statistics in both traditional and online settings. They were asked to provide feedback about the attributes reflective of the construct. They offered recommendations on both the content areas and the clarity of the language in the wording of the items. In terms of the content areas, the experts suggested adding some items in two content areas: (a) the negative mindset or pre-perceptions when taking statistics training such as “much of the content is self-taught” and “quantitative-minded people hang out together” and (b) technology or online-learning-related areas such as “the course will primarily deal with computers” and “I have statistics questions that I need answered right away.” These areas match with the study results in qualitative research such as Kinkead et al. (2016) and Malik (2015) as well. Also, because online or hybrid statistics courses tend to emphasize the application or interpretation aspect of the statistics instruction, class/content anxiety was merged with interpretation anxiety as few items were created for the former anxiety.

Additionally, the experts suggested improving the clarity of the wording of the items. For example, “Asking a private teacher to explain a topic that I have not understood at all” was changed to “Asking instructors to explain a topic that I do not understand at all” to reflect the culture of college programs. The experts suggested eliminating some items such as “studying statistics” and “enrolling in a statistics course” because they were too general or abstract.

In total, 45 items across the four dimensions of Class and Interpretation anxiety (CI), Fear of Asking for Help Anxiety (FA), Online System anxiety (OS), and Pre-Conceptions Anxiety (PC) tested at this step. The CI was defined as the anxiety the learners encountered when taking a statistics class or interpreting statistical data; the FA was defined as the anxiety the learners experienced when asking an instructor for help understanding the statistical concepts or course requirements such as assignments. The OS was defined as the anxiety the learners faced when using the online computer system, statistical software, and the lack of personal contact during the statistics learning process; the PC was defined as the negative mindset or preconception that the learners associated with taking statistics courses. Some items were adapted from the existing instruments where appropriate, and some new items were written based on the literature, course syllabi, and expert feedback.

Item Selection and Exploratory Factor Analysis Procedure First, the response distributions of the individual items were examined. The highly skewed and unbalanced items were eliminated or modified based on the descriptive frequency test (Clark & Watson, 1995). For example, the item “It is much easier if I can ask a statistician to run the analysis for me” had low distinction because almost every participant chose “agree” or “strongly agree” with this item. In contrast, in response to the item “Reading an advertisement for a car which includes figures on miles per gallon, depreciation, etc.” most participants answered with “No Anxiety” or “Little Anxiety.” In a follow-up consultation, experts agreed that these items should be eliminated as the students with low level of statistics anxiety may still agree with the first item and the second item was not as closely related to the four dimensions as other items and it was too easy.

By default, Mplus uses the oblique rotation of GEOMIN (Muthén & Muthén, 1998/2017) in the EFA. For the analysis, the statement “type = efa 1 5” was specified in Mplus for the study. It asked the program to produce output for a 1, 2, 3, 4, and 5 factor solution to check the number of factors that fit the data the best. From the scree plot, four eigenvalues were larger than 1. Their values were 13.19, 2.31, 1.63, and 1.49, and they explained 39.96%, 7.01%, 4.94%, and 4.52% of the total variance, respectively. Although the first eigenvalue was much larger than the second one, the four-factor solution was accepted in the EFA test based on the significant chi-square different test results among the one-factor, two-factor, three-factor, and four-factor models.

Once the four-factor model was accepted, the next step was to check the fit of the items under each factor. The items with weak loadings (i.e., less than 0.3) were eliminated (Hair et al., 1995). Additionally, some items loaded on multiple factors with moderate-sized loadings were eliminated. For example, item “Getting a low grade despite putting in all the time and effort” loaded on three factors with a loading of 0.474, 0.343, and 0.345 respectively. Because such items do not discriminate well among the factors and made it difficult to interpret the meaning of the final scale, they were eliminated from the final scale (Kline, 2000). It is important to note that the substantial meaning of the items was checked first to make sure that they were not the unique contributors of a factor before any eliminations during the process (Pett et al., 2003). The finalized conceptual model of the four-factor SASOH model has four dimensions: (a) Class and Interpretation Anxiety (CI), (b) Fear of Asking for Help Anxiety (FA), (c) Online System Anxiety (OS), and (d) Pre-Conception Anxiety (PC).

4.2. STAGE II: FORMAL TESTING OF THE INSTRUMENT

Classical Item Analysis The descriptive item statistics including the percentage distribution among the five categories, the item means, standard deviations, and item-total correlations are provided in Appendix A. The items with the highest means were Item 23, "I heard so many people had problems with statistics which makes me nervous" (M = 3.35, SD = 1.19), and Item 26, "I worry about how much of the content I have to learn on my own" (M = 3.34, SD = 1.18). The item with the lowest mean is Item 18, "Having to deal primarily with computers during the statistics course" (M = 2.40, SD = 1.23). The item discrimination indices such as item-total correlations ranged from 0.44 to 0.74, which satisfied the commonly used criterion that the range of the correlations should be between 0.3 and 0.9 (Rodriguez & Albano, 2017).

The Cronbach alpha for the whole instrument with all 27 items was 0.95. For the four sub-scales, the Cronbach alpha values were 0.924 for CI, 0.921 for FA, 0.816 for OS, and 0.804 for PC. All of the Cronbach alpha coefficients were over 0.7, which satisfied the criterion usually considered to identify strict internal consistency (Hair et al., 1995).

CFA Results Based on the literature review and the previous scales measuring statistics anxiety, a four-factor model was specified in which 12 items loaded onto the first factor (CI), 5 items loaded onto the second factor (FA), 5 items loaded onto the third factor (OS), and 5 items loaded onto the fourth factor (PC). All 27 items utilized a 5-point response scale with higher scores reflecting higher levels of statistics anxiety. The first item in each factor was used as the marker indicator for the analysis. The four-factor model contained no double-loading items, and all measurement errors were presumed to be uncorrelated. The factors were permitted to be correlated based on the past literature.

Out of the 709 participants who completed the survey, 680 (95.9%) completed all 27 items. Therefore, the final sample size for the CFA test was 680. No univariate or multivariate outliers were detected based on the leverage indices for each participant. The skewness and kurtosis of the items approximated a normal distribution. For example, the skewness was close to zero, and the kurtosis was between -2 and +2 (Gravetter & Wallnau, 2014). Also, because they had four or more response categories, these items could be treated as continuous variables (Bentler & Chou, 1987). Therefore, the ML estimator could be used in the CFA testing. When calculating the z-value of skewness or kurtosis divided by the standard error, several items were greater than 1.96 ($p < 0.05$). Therefore, the WLMSV estimator (weighted least square parameter estimates using a diagonal weight matrix with standard errors and mean) that demonstrated the best performance in the CFA of the categorical variables such as the Likert-scale items (Brown, 2015) were used in the CFA in this study.

The four-factor model fit was acceptable by every criterion except the significant χ^2, likely due to the large sample. The fit statistics were as follows: $\chi^2(318) = 992.196$, $p < 0.001$; CFI = 0.966; TLI = 0.963; RMSEA = 0.055; SRMR = 0.034. Thus, the 27 items appeared to measure four separate but related constructs. Further examination of local fit via normalized residual covariances and modification indices yielded no interpretable remaining relationships. For example, the modification indices suggested loading item 24, "I do not know what to expect in a statistics course," and item 16, "Asking instructors for help when trying to interpret a results table," on the OS factor. Such modifications, however, were not substantively meaningful according to the expert feedback and literature review. Thus, this four-factor model was retained. Unstandardized and standardized parameter estimates are presented in Table 3 and 4. All factor loadings and the factor covariance were statistically significant. As shown in Table 3, standardized loadings for the CI items ranged from 0.683 to 0.799 (with R^2 values for the amount of item variance accounted for by the factor ranging from 0.466 to 0.638), standardized loadings for the FA factor ranged from 0.836 to 0.908 (with R^2 values of 0.699 to 0.824), standardized loadings for the OS factor ranged from 0.604 to 0.786 (with R^2 values of 0.365 to 0.618), and standardized loadings for the PC factor ranged from 0.642 to 0.772 (with R^2 values of 0.412 to 0.596). The factor loading estimates suggest that the items were strongly related to their purported factors. Moreover, the four factors were moderately to highly related as well (range of r_s = 0.507 to 0.856). The four-factor model represents an adequate description of statistics anxiety in an online or hybrid setting. The final four-factor SASOH model with standardized estimates with the standard errors are presented in Appendix B).

Table 3. Parameter Estimates from the Four-factor SASOH Model (*n* = 680)

CI	**Unstandardized**		**Standardized**		**FA**	**Unstandardized**		**Standardized**	
	Estimate	***SE***	***Estimate***	***SE***		***Estimate***	***SE***	***Estimate***	***SE***
Q1	1.000	0.000	0.683	0.020	Q4	1.000	0.000	0.836	0.014
Q2	1.082	0.031	0.739	0.018	Q8	1.036	0.020	0.866	0.012
Q3	1.097	0.033	0.749	0.017	Q16	1.069	0.020	0.893	0.011
Q5	1.071	0.036	0.731	0.018	Q19	1.085	0.020	0.908	0.010
Q6	1.089	0.032	0.743	0.018	Q22	1.033	0.021	0.864	0.012
Q7	1.148	0.033	0.784	0.015					
Q10	1.083	0.034	0.739	0.018					
Q11	1.165	0.034	0.796	0.014					
Q14	1.079	0.034	0.736	0.017					
Q15	1.075	0.035	0.734	0.018					
Q20	1.170	0.034	0.799	0.015					
Q21	1.085	0.034	0.741	0.018					
OS					**PC**				
Q9	1.000	0.000	0.604	0.027	Q23	1.000	0.000	0.772	0.020
Q12	1.273	0.061	0.769	0.019	Q24	0.958	0.039	0.739	0.024
Q13	1.190	0.055	0.719	0.021	Q25	0.915	0.037	0.707	0.023
Q17	1.301	0.063	0.786	0.017	Q26	0.898	0.036	0.693	0.025
Q18	1.246	0.057	0.753	0.019	Q27	0.832	0.036	0.642	0.026

R-Square for Item Variances

Items	R-Square	Items	R-Square
Q1	0.466	Q15	0.539
Q2	0.546	Q16	0.798
Q3	0.561	Q17	0.618
Q4	0.699	Q18	0.567
Q5	0.535	Q19	0.824
Q6	0.552	Q20	0.638
Q7	0.615	Q21	0.549
Q8	0.750	Q22	0.746
Q9	0.365	Q23	0.596
Q10	0.546	Q24	0.547
Q11	0.633	Q25	0.499
Q12	0.592	Q26	0.480
Q13	0.517	Q27	0.412
Q14	0.542		

Table 4. Factor Variances and Covariance (n = 680)

Factor Covariance									
CI with FA	0.367	0.022	0.643	0.025	FA with OS	0.315	0.023	0.624	0.027
CI with OS	0.353	0.023	0.856	0.014	FA with PC	0.328	0.025	0.507	0.032
CI with PC	0.415	0.023	0.786	0.019	OS with PC	0.341	0.024	0.731	0.023
Factor Variances									
CI	0.466	0.028	1.000	0.000	FA	0.699	0.024	1.000	0.000
OS	0.365	0.033	1.000	0.000	PC	0.596	0.032	1.000	0.000

4.3. MEASUREMENT AND STRUCTURAL INVARIANCE

At the first step, the four-factor SASOH model was tested for the two groups separately. For undergraduates, the fit statistics were as follows: $\chi^2(318) = 631.048$, $p < 0.001$; CFI = 0.963; TLI = 0.960; RMSEA = 0.056; SRMR = 0.041. For graduates, the fit statistics were as follows: $\chi^2(318) = 695.311$, $p < 0.001$; CFI = 0.966; TLI = 0.962; RMSEA = 0.058; SRMR = 0.042. Adequate fit of the model for both groups was met based on the previously stated criteria. Therefore, no modifications were made at this step.

For the configural invariance model in Step 2, the four-factor SASOH model was estimated simultaneously in each group with undergraduates as the reference group. Fit statistics were as follows: $\chi^2(636) = 1326.878$, $p < 0.001$; CFI = 0.965; TLI = 0.961; RMSEA = 0.057; SRMR = 0.042. Therefore, the configural invariance model had good fit, and it served as the baseline model for the subsequent tests.

In Step 3, the metric invariance model was tested by constraining all factor loadings to be equal across groups, but all item thresholds were freely estimated. Fit statistics were as follows: $\chi^2(659) = 1252.083$, $p < 0.001$; CFI = 0.970; TLI = 0.968; RMSEA = 0.052; SRMR = 0.044. The p-value for the rescaled likelihood ratio test with $\chi^2_{\text{diff}}(23) = 37.271$ was lower than 0.05. The modification indices suggested freeing the factor loading for Q9 between groups. Therefore, at Step 3_1, factor loadings were constrained equal between the two groups except item 9 “Not seeing instructors teach physically”. The p-value for the change in the chi-square value became above 0.05 [$\chi^2_{\text{diff}}(22) = 29.556$, $p = 0.130$]. Therefore, partial metric invariance held between undergraduates and graduates.

In Step 4, the item thresholds were constrained to be equal across groups on top of all equal factor loadings except item 9 in a scalar invariant model. In comparison to the partial metric invariance model above, the p-value for the change in the chi-square value was above 0.05 [$\chi^2_{\text{diff}}(104) = 106.838$, $p = 0.405$]. Based on the modification indices, no more modifications were needed. Therefore, the full scalar invariance model was maintained.

In Step 5, equal residual variance invariance or strict factorial invariance model was tested. The p-value for rescaled likelihood ratio test with $\chi^2_{\text{diff}}(26) = 36.766$ was above 0.05. The residual variance for Q7 was further set to be free in Step 5_1 based on the modification indices. The p-value for rescaled likelihood ratio test with $\chi^2_{\text{diff}}(25) = 29.454$ was still above 0.05, which indicated that residual variance for item 7 “Interpreting statistical concepts in the discussion forums” was different between undergraduates and graduates (e.g. the residual variance is smaller for graduates than undergraduates). Therefore, partial residual variance invariance or strict factorial invariance model held at this step.

In Step 6 and 7, the structural invariance was tested. To test for the factor variances invariance, the factor variances in graduates were constrained to be equal to the factor variances in undergraduates. That resulted in no difference in fit [$\chi^2_{\text{diff}}(4) = 3.552$, $p = 0.470$]. Therefore, the factor variances invariance was confirmed. To test for the factor means invariance, the factor means in graduates were constrained to be equal to the factor means in undergraduates. That also resulted in no difference in fit [$\chi^2_{\text{diff}}(4) = 5.556$, $p = 0.235$]. Based on the modification indices, however, the mean of factor FA was set free while others were constrained to be equal. The p-value for the change in the chi-square value was still above 0.05 [$\chi^2_{\text{diff}}(3) = 1.614$, $p = 0.656$]. Therefore, partial structural invariance was obtained across undergraduates and graduates with the latter having less fear of asking for help anxiety than the former.

In conclusion, the four-factor SASOH model can be considered at least partially invariant for program groups based on the overall good fit, and 20% or fewer noninvariant factor loadings and thresholds (Dimitrov, 2010). The partial measurement invariance and structural invariance were obtained across undergraduates and graduates. The results revealed some differences between the two groups in item 7, item 9, and the FA factor. For example, item 7 explained more variances in the Class and Interpretation Anxiety (CI) factor for graduates than undergraduates (e.g., the residual variance is smaller for graduates than undergraduates) and graduates had less Fear of Asking for Help Anxiety (FA) than undergraduates overall. See Table 5 for all the results of the program groups.

4.4. RELATIONSHIP WITH ATTITUDES TOWARD STATISTICS AND MATHEMATICS ANXIETY

Based on the definition of statistics anxiety (Chew & Dillon, 2014), the predictive relationship from the attitudes toward statistics and the discriminant validity from the mathematics anxiety to statistics anxiety were assessed as well. Attitudes toward statistics was measured using the items from the STARS instrument. In the original instrument, 28 items are used to measure three dimensions (a) worth of statistics, (b) computational self-concept, and (c) fear of statistics teachers. For this study, 9 items were randomly selected from the three dimensions, and the sum scores were used for the test. The Cronbach's alpha coefficient for these items is 0.879. A multivariate regression analysis was conducted to test the predictive relationship between attitudes toward statistics as the independent variable and the four factors of statistics anxiety as dependent variables. The results show that attitude toward statistics significantly predicts the four factors. For CI, $F(1, 678) = 258.164$, $p < 0.001$, adjusted $R^2 = 0.275$; for FA, $F(1, 678) = 135.168$, $p < 0.001$, adjusted $R^2 = 0.165$; for OS, $F(1, 678) = 234.494$, $p < 0.001$, adjusted $R^2 = 0.256$; and for PC, $F(1, 678) = 420.695$, $p < 0.001$, adjusted $R^2 = 0.382$. Therefore, attitudes toward statistics significantly predicted the four factors of statistics anxiety.

The Pearson correlation test was conducted among the four factor scores and the sum score of mathematics anxiety. Math Anxiety was measured using the Abbreviated Math Anxiety Scale (AMAS) created by Hopko, Mahadevan, Bare, and Hunt (2003). It has nine 5-point Likert-scale items with 1 = "very low anxiety" and 5 = "very high anxiety". The sum scores range from 9 to 45. The range of correlation coefficients between the four factors and math anxiety were within 0.308 and 0.396 while the range of coefficients among the four factors were within 0.591 and 0.926. Therefore, all the convergent coefficients were at least 0.195 greater than the discriminant coefficients. Based on the rule that convergent correlations should be much higher than the discriminant correlations (Campbell & Fiske, 1959), the results showed that the discriminant validity from mathematics anxiety was confirmed.

Table 5. Tests of Measurement Invariance of SASOH in Undergraduates and Graduates (n=668)

Model [a]	**χ^2/*df***	**χ^2difference/Δ*df*** [b]	**CFI/ TLI**	**RMSEA (90% CI)**	**SRMR**	**Invariance met**
Step 1: Single Group Solutions						
Undergraduate (n = 313)	631.048/318(.00)		0.963/0.960	0.056 (0.050, 0.062)	0.041	
Graduate (n = 355)	695.311/318(.00)		0.966/0.962	0.058 (0.052, 0.064)	0.042	
Step 2-6: Measurement Invariance						
Step 2: Equal form	1326.878/636(.00)		0.965/0.961	0.057 (0.053, 0.061)	0.042	
Step 3: Equal factor loadings	1252.083/659(.00)	37.271/23(.03)	0.970/0.968	0.052 (0.048/0.056)	0.044	
Step 3_1: Equal factor loadings Partial[c]	1243.462/658(.00)	29.556/22(.13)	0.970/0.968	0.052 (0.047/0.056)	0.044	Yes
Step 4: Equal indicator intercepts	1338.024/762(.00)	106.838/104(.40)	0.971/0.973	0.048 (0.043, 0.052)	0.044	Yes
Step 5: Equal item residual variance	1338.024/762(.00)	36.766/26 (.08)	0.971/0.973	0.048 (0.043, 0.052)	0.044	Yes
Step 5_1: Equal item residual variance[d]	1326.104/761(.00)	29.454/25(.25)	0.971/0.973	0.047 (0.043, 0.051)	0.044	Yes
Step 6: Equal Factor Variance Invariance	1144.655/765(.00)	3.552/4(.47)	0.981/0.982	0.039 (0.034, 0.043)	0.044	Yes
Step 7: Equal Factor Means	1107.162/769(.00)	5.556/4(.23)	0.983/0.984	0.036 (0.031, 0.041)	0.045	Yes
Step 7_1: Equal Factor Means Partial[e]	1099.713/768(.00)	1.614/3(.66)	0.983/0.985	0.036 (0.031, 0.041)	0.044	Yes

Note. *p*-value in parentheses. [a] Estimator = Weighted Least Square Parameter Estimates using a Diagonal Weight Matrix with Standard Errors and Mean-(WLSMV). [b]The chi-square value for MLM, MLMV, MLR, ULSMV, WLSM and WLSMV cannot be used for chi-square difference testing in the regular way. MLMV, WLSMV, and ULSMV difference testing is done using the DIFFTEST option. The DIFFTEST option assumes the models are nested. [c]Based on the modification indices, factor loadings were constrained equal between the two groups except Q9. [d]Based on the modification indices, residual variance for Q7 was set to be free. [e]Based on the modification indices, the means of FA were set to be different for the two groups.

5. DISCUSSION AND CONCLUSION

With the purpose of developing an instrument for measuring statistics anxiety in the online or hybrid setting where there are no or few face-to-face interaction opportunities among the instructors and students and where the statistical applications with few or no examination components were emphasized, this study filled the gap in the previous literature: the lack of an existing instrument measuring statistics anxiety in such a setting. With more and more universities offering online or hybrid statistics courses instead of traditional face-to-face ones, the need for such instrument in an online or hybrid setting cannot be further ignored. Considering the current COVID-19 pandemic, when most universities and schools have been pushed to offer online instructions, the need for such instrument has never been more urgent.

The resulting SASOH model has four dimensions with 27 items. The four dimensions are Class and Interpretation anxiety (CI), Fear of Asking for Help Anxiety (FA), Online System Anxiety (OS), and Pre-conception Anxiety (PC). Compared to the common dimensions in the existing instruments (Table 1), the CI is a combination of Class/Content and Interpretation Anxiety that the learners encounter when taking a statistics class or interpreting statistical data; the FA is similar to the Asking for Help Anxiety, defined as the anxiety the learners experienced when asking an instructor for help. The OS is defined as the anxiety the learners face when using an online computer system or statistical software and when there is a lack of personal contact during the statistics learning process. It has the Computer Usefulness and Experience Anxiety component (Zanakis & Valenzi, 1997), but it also takes the unique characteristics of online learning such as the lack of personal interactions into consideration. The PC is defined as the negative mindset or preconception that the learners associated with taking online statistics courses such as "I worry about how much of the content I have to learn on my own" which matches with the experts' recommendations and the past literature (Schulze, 2009; Tichavsky et al., 2015).

The results of the EFA and CFA revealed that the four-factor SASOH model represents an adequate description of statistics anxiety in an online or hybrid setting. Additionally, the predictive validity and the discriminant validity of the instrument were confirmed. Moreover, multiple-groups CFA affirmed that the resulting model achieved at least partial measurement and structural invariance across gender and program. Therefore, males and females respond to items in SASOH in a similar manner, so do undergraduates and graduates. Considering that the current statistics anxiety instruments were mostly designed and validated for undergraduate education, the new instrument also provides the users a tool to measure statistics anxiety in graduate programs. Similarly, the gender effect on statistics anxiety has been studied in the past literature (Hsiao & Chiang, 2011), and such effect should be further studied in an online or hybrid setting using SASOH.

The newly developed instrument could be beneficial to instructors in several ways: As a diagnostic tool, it can be used to diagnose the level of statistics anxiety of students among the four dimensions. Based on the anxiety scores, the students could be roughly classified into a low or high anxiety group. If statistics anxiety can be measured longitudinally, the progress of increased or decreased anxiety can be tracked for the duration of a course overtime. Finally, with knowledge in the above areas, instructors can be better informed of their students' level of anxiety. Thus, they can modify their instructional plans or design intervention strategies to cope with their students' anxiety. In addition, researchers and practitioners can use SASOH to assess statistics anxiety and test its role as a covariate, predictor, or outcome variable in relationship to other variables in the hybrid or online setting.

Based on the results and limitations of the current study, there are several recommendations for future studies. Because the SASOH is a new instrument, replication studies providing further evidence of the validity and reliability of the instrument are needed. Although Qualtrics aimed at collecting a national representative sample for the study, the sample was still convenient in nature. Therefore, further studies are recommended to test the instrument using a large nationally and/or internationally representative sample. In the replication studies, it is important to examine the generalizability of the four-factor model in different demographics groups such as gender groups, ethnicity groups, and age groups, and in groups with different levels of online learning and statistics experiences.

In this study, the measurement invariance and population heterogeneity were tested between males and females and between undergraduates and graduates using the multiple-group CFA. It is important to conduct the test on other variables as well. For example, researchers could test the measurement invariance and population heterogeneity in the following areas: (1) between different types of programs such as part-time vs. full-time, (2) between different settings of the statistics course such as hybrid vs. online, or (3) among different levels of experiences with online learning or statistics courses.

Even though this study tested the first-order multidimensional construct based on the past literature, the CFA test revealed that the four factors were moderately to highly related (range of rs = 0.507 to 0.856). Therefore, a second-order CFA model could be tested in which there is a single broader dimension of statistics anxiety (SA) and there are four subdimensions of statistics anxiety (CI, FA, OS, and PC). This could allow researchers to see if the second-factor model provides a more parsimonious structure to explain the correlations among the subdimensions (Brown, 2015). The researchers could use the rescaled likelihood ratio tests to test the nested models for the superiority of the first-order or the second-order SASOH model.

Although statistics anxiety is a widely studied topic and several existing instruments have been developed to measure the construct, the cutoff scores of the measurement were not provided (Chew & Dillon, 2014). If the SASOH instrument reveals good results in the replication studies, then I would recommend that future studies focus on separating the statistics anxiety scale into meaningful low, medium, and high ranges because such ranges will be helpful for diagnostic, classification, progress, and modification-of-instruction purposes (Angoff, 1984).

ACKNOWLEDGEMENTS

I would like to thank the LaVerne Academy for funding this research study.

REFERENCES

American Educational Research Association, American Psychological Association, & National Council on Measurement in Education (Eds.). (2014). *Standards for educational and psychological testing*. American Educational Research Association.

Angoff, W. H. (1984). *Scales, norms, and equivalent scores.* Educational Testing Service.

Baloğlu, M. (2002). Psychometric properties of the statistics anxiety rating scale. *Psychological Reports, 90*(1), 315–325. https://doi.org/10.2466/pr0.2002.90.1.315

Baloğlu, M., Abbassi, A., & Kesici, S. (2017). Multivariate relationships between statistics anxiety and motivational beliefs. *Education, 137*(4), 430–444.

Bentler, P., & Chou, C. (1987). Practical issues in structural modeling. *Sociological Methods & Research, 16*(1), 78–117. https://doi.org/10.1177/0049124187016001004

Betz, N. E. (1978). Prevalence, distribution, and correlates of math anxiety in college students. *Journal of Consulting Psychology, 25*(5), 441–448. https://doi.org/10.1037/0022-0167.25.5.441

Boas, T. C., Christenson, D. P., & Glick, D. M. (2020). Recruiting large online samples in the United States and India: Facebook, Mechanical Turk, and Qualtrics. *Political Science Research and Methods, 8*(2), 232–250. https://doi.org/10.1017/psrm.2018.28

Boettcher, J. V., & Conrad, R. (2010). *The online teaching survival guide: Simple and practical pedagogical tips*. Jossey-Bass.

Bowen, N., & Masa, R. (2015). Conducting measurement invariance tests with ordinal data: A guide for social work researchers. *Journal of the Society for Social Work and Research, 6*(2), 229–249. https://doi.org/10.1086/681607

Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Publications.

Campbell, D. T., & Fiske , D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81–105. https://doi.org/10.1037/h0046016

Chiou, C., Wang, Y., & Lee, L. (2014). Reducing statistics anxiety and enhancing statistics learning achievement: Effectiveness of a one-minute strategy. *Psychological Reports, 115*(1), 297–310. https://doi.org/10.2466/11.04.pr0.115c12z3

Chew, P. K., & 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*(2), 196–208. https://doi.org/10.1177/1745691613518077

Chew, P. K. H., Swinbourne, A., & Dillon, D. B. (2017). An absence of attentional bias: Statistics anxiety is unique among anxieties. *Journal of Articles in Support of the Null Hypothesis*, *13*(2), 91–112.

Chiesi, F., & Primi, C. (2010). Cognitive and non-cognitive factors related to students' statistics achievement. *Statistics Education Research Journal*, *9*(1), 6–26. https://iase-web.org/documents/SERJ/SERJ9(1)_Chiesi_Primi.pdf?1402525009

Clark, L., & Watson, D. (1995). Constructing validity: Basic issues in objective scale development. *Psychological Assessment*, *7*(3), 309–319. https://doi.org/10.1037/1040-3590.7.3.309

Cruise, R. J., Cash, R. W., & Bolton, D. L. (1985). Development and validation of an instrument to measure statistical anxiety. *American Statistical Association Proceedings of the Section on Statistical Education*, Nevada, August 5–8 (pp. 92–97).

DeVaney, T. A. (2010). Anxiety and attitude of graduate students in on-campus vs. online statistics courses. *Journal of Statistics Education*, *18*(1), Article 2. https://doi.org/10.1080/10691898.2010.11889472

DeVaney, T. A. (2016). Confirmatory factor analysis of the statistical anxiety rating scale with online graduate students. *Psychological Reports*, *118*(2), 565–586. https://doi.org/10.1177/0033294116644093

Dimitrov, D. (2010). Testing for factorial invariance in the context of construct validation. *Measurement and Evaluation in Counseling and Development*, *43*(2), 121–149. https://doi.org/10.1177/0748175610373459

Earp, M. (2007). *Development and validation of the statistics anxiety measure.* [Doctoral dissertation] Univeristy of Denver.

Einbinder, S. D. (2014). Reducing research anxiety among MSW students. *Journal of Teaching in Social Work*, *34*(1), 2–16. https://doi.org/10.1080/08841233.2013.863263

Gorsuch, R. L. (1997). Exploratory factor analysis: Its role in item analysis. *Journal of Personality Assessment*, *68*(3), 532–560. https://doi.org/10.1207/s15327752jpa6803_5

Gravetter, F., & Wallnau, L. (2014). *Essentials of statistics for the behavioral sciences* (8th ed.). Wadsworth, Cengage Learning.

Hair, J., Anderson, R. E., Tatham, R., & Black, W. (1995). *Multivariate data analysis with readings* (4th ed.). Prentice Hall.

Hanna, D., & Dempster, M. (2009). The effect of statistics anxiety on students' predicted and actual test scores. *The Irish Journal of Psychology*, *30*(3-4), 201–209. https://doi.org/10.1080/03033910.2009.10446310

Hanna, D., Shevlin, M., & Dempster, M. (2008). The structure of the statistics anxiety rating scale: A confirmatory factor analysis using UK psychology students. *Personality and Individual Differences*, *45*(1), 68–74. https://doi.org/10.1016/j.paid.2008.02.021

Heen, M. S. J., Lieberman J. D., & Miethe, T. D. (2014). A comparison of different online sampling approaches for generating national samples. [Monograph]. *State Data Brief Center for Crime and Justice Policy, Las Vegas, September 2014*. [CCJP 2014-01].

Hopko, D. R., Mahadevan, R., Bare, R. L., & Hunt, M. K. (2003). The abbreviated math anxiety scale (AMAS): Construction, validity, and reliability. *Assessment*, *10*(2), 178–182. https://doi.org/10.1177/1073191103010002008

Hsiao, T. (2010). The statistical anxiety rating scale: Further evidence for multidimensionality. *Psychological Reports*, *107*(3), 977–982. https://doi.org/10.2466/07.11.pr0.107.6.977-982

Hsiao, T., & Chiang, S. (2011). Gender differences in statistics anxiety among graduate students learning English as a foreign language. *Social Behavior and Personality: An International Journal*, *39*(1), 41–42. https://doi.org/10.2224/sbp.2011.39.1.41

Hsu, M. K., Wang, S. W., & Chiu, K. K. (2009). Computer attitude, statistics anxiety and self-efficacy on statistical software adoption behavior: An empirical study of online mba learners. *Computers in Human Behavior*, *25*(2), 412–420. https://doi.org/10.1016/j.chb.2008.10.003

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Kinkead, K. J., Miller, H., & Hammett, R. (2016). Adult perceptions of in-class collaborative problem solving as mitigation for statistics anxiety. *The Journal of Continuing Higher Education*, *64*(2), 101-111. https://doi.org/10.1080/07377363.2016.1178057

Kline, P. (1994). *An easy guide to factor analysis*. Routledge.

Kline, P. (2000). *The handbook of psychological testing* (2nd ed.). Routledge.

Lin, Y., Durbin, J. M., & Rancer, A. S. (2016). Math anxiety, need for cognition, and learning strategies in quantitative communication research methods courses. *Communication Quarterly*, *64*(4), 390–409. https://doi.org/10.1080/01463373.2015.1103294

Liu, L., & Haque, M. (2017). Age difference in research course satisfaction in a blended Ed.D. program: A moderated mediation model of the effects of internet self-efficacy and statistics anxiety. *Online Journal of Distance Learning Administration*, *20*(2), Article 5. https://www.westga.edu/~distance/ojdla/summer202/liu_haque202.html

Malik, S. (2015). Undergraduatess statistics anxiety: A phenomenological study. *The Qualitative Report*, *20*(2), 120–133. https://nsuworks.nova.edu/tqr/vol20/iss2/11

Mji, A. (2009). Differences in university students' attitudes and anxiety about statistics. *Psychological Reports*, *104*(3), 737–44. https://doi.org/10.2466/pr0.104.3.737-744

Muthén, L. K., & Muthén, B. O. (1998/2017). *Mplus user's guide* (8th ed.). Muthén and Muthén.

Onwuegbuzie, A. J. (2004). Academic procrastination and statistics anxiety. *Assessment and Evaluation in Higher Education*, *29*(1), 3–19. https://doi.org/10.1080/0260293042000160384

Onwuegbuzie, A. J., & Seaman, M. A. (1995). The effect of time constraints and statistics test anxiety on test performance in a statistics course. *Journal of Experimental Education*, *63*(2), 115–124. https://doi.org/10.1080/00220973.1995.9943816

Pett, M., Lackey, N., & Sullivan, J. (2003). *Making sense of factor analysis: The use of factor analysis for instrument development in health care research*. SAGE Publications.

Pretorius, T. B., & Norman, A. M. (1992). Psychometric data on the statistics anxiety scale for a sample of South African students. *Educational and Psychological Measurement*, *52*(4), 933–937. https://doi.org/10.1177/0013164492052004015

Rapp-McCall, L. A., & Anyikwa, V. (2016). Active learning strategies and instructor presence in an online research methods course: Can we decrease anxiety and enhance knowledge? *Advances in Social Work*, *17*(1), 1–14. https://doi.org/10.18060/20871

Razavi, S. A., Shahrabi, A., & Siamian, H. (2017). The relationship between research anxiety and self-efficacy. *Materia Socio-Medica*, *29*(4), 247–250. https://doi.org/10.5455/msm.2017.29.247-250

Rodriguez, M., & Albano, A. (2017). *The college instructor's guide to writing test items: Measuring student learning*. Routledge.

Schulze, S. (2009). Teaching research methods in a distance education context: Concerns and challenges. *South African Journal of Higher Education*, *23*(5), 992–1008. https://doi.org/10.4314/sajhe.v23i5.48812

Tichavsky, L. P., Hunt, N., Driscoll, A., & Jicha, K. (2015). "It's just nice having a real teacher": Student perceptions of online versus face-to-face instruction. *International Journal for the Scholarship of Teaching and Learning*, *9*(2), Article 2. https://doi.org/10.20429/ijsotl.2015.090202

U.S. Department of Education. (2018). National Center for Education Statistics, Integrated Postsecondary Education Data System (IPEDS), Spring 2016 and Spring 2017, Fall Enrollment component. [NCES 2017136]

Vigil-Colet, A., Lorenzo-Seva, U., & Condon, L. (2008). Development and validation of the statistical anxiety scale. *Psicothema*, *20*(1), 174–180.

Whitcome, J. A. (2004). Measuring statistics anxiety using a stage theory. *Academic Exchange Quarterly*, *8*(3), 140–147. https://www.thefreelibrary.com/Measuring+statistics+anxiety+using+a+stage+theory.-a0126683358

Williams, A. S. (2013). Worry, intolerance of uncertainty, and statistics anxiety. *Statistics Education Research Journal*, *12*(1), 48–59. http://iase-web.org/documents/SERJ/SERJ12(1)_Williams.pdf?1402525003

Xu, D., & Jaggars, S. (2014). Performance gaps between online and face-to-face courses: Differences across types of students and academic subject areas, *The Journal of Higher Education*, *85*(5), 633–659. https://doi.org/10.1353/jhe.2014.0028

Zanakis, S., & Valenzi, E. (1997). Student anxiety and attitudes in business statistics. *Journal of Education for Business*, *73*(1), 10–16. https://doi.org/10.1080/08832329709601608

Zeidner, M. (1991). Statistics and mathematics anxiety in social science students: Some interesting parallels. *British Journal of Educational Psychology*, *61*(3), 319–332. https://doi.org/10.1111/j.2044-8279.1991.tb00989.x

LU LIU
Leo Hall 101
1950 Third Street
La Verne, CA, 91750

APPENDIX A

Item Response Frequencies, and Classical Item Statistics of the 27-item SASOH (*n* = 680)

		Response frequencies (%)[b]					5-response categories		
Item	**Subdomain[a]**	**1**	**2**	**3**	**4**	**5**	**M**	**SD**	**r[c]**
1. Doing the coursework for a statistics course	CI	11.14	22.57	29.76	23.13	13.26	3.05	1.20	0.65
2. Interpreting the meaning of a table in a journal article	CI	16.36	23.84	28.49	20.87	10.44	2.85	1.22	0.69
3. Making an objective decision based on empirical data	CI	17.07	29.48	27.50	18.19	7.76	2.70	1.18	0.68
4. Asking instructors for individual help with material I am having difficulty understanding	FA	20.31	23.27	25.11	17.91	13.40	2.81	1.31	0.61
5. Reading a journal article that includes some statistical analyses	CI	25.25	26.94	24.54	15.51	7.62	2.53	1.23	0.67
6. Trying to decide which statistical analysis is appropriate for my research project	CI	8.89	22.57	30.18	24.54	13.82	3.12	1.17	0.68
7. Interpreting statistical concepts in the discussion forums	CI	13.82	25.39	26.66	22.43	11.71	2.93	1.22	0.72
8. Asking instructors for help in understanding a statistical concept	FA	21.44	23.70	24.82	17.07	12.83	2.76	1.31	0.64
9. Not seeing instructors teach physically	OS	25.39	25.95	24.54	14.67	9.45	2.57	1.27	0.53
10. Critiquing a quantitative journal article or study	CI	16.08	27.93	28.63	18.34	9.03	2.76	1.19	0.68
11. Interpreting the meaning of a probability value once I have found it	CI	19.04	28.21	24.82	17.63	10.30	2.72	1.25	0.72
12. Importing a data file into the statistical software	OS	21.72	26.52	24.68	17.07	10.01	2.67	1.26	0.67
13. Not allowing personal contact in an online setting	OS	27.93	25.95	22.85	13.96	9.31	2.51	1.28	0.62
14. Determining whether to reject or retain the null hypothesis	CI	15.09	29.48	26.80	19.18	9.45	2.78	1.19	0.69
15. Posting a presentation on statistical results interpretation	CI	12.41	22.85	24.68	19.04	21.02	3.13	1.32	0.67
16. Asking instructors for help when trying to interpret a results table	FA	20.73	25.53	25.53	18.05	10.16	2.71	1.26	0.68

17. Using statistical software to run an analysis	OS	22.71	24.82	26.94	17.63	7.90	2.63	1.23	0.68
18. Having to deal primarily with computers during the statistics course	OS	29.76	27.79	21.58	14.10	6.63	2.40	1.23	0.66
19. Asking instructors to explain a topic that I do not understand	FA	22.85	23.13	23.55	16.93	13.54	2.75	1.34	0.65
20. Trying to understand the statistical analyses described in the abstract of a journal article	CI	14.95	28.77	28.91	19.61	7.76	2.76	1.16	0.74
21. Submitting a research project which requires statistical analysis	CI	10.86	23.27	27.93	21.30	16.64	3.10	1.24	0.68
22. Asking instructors about how to do an assignment	FA	23.98	24.96	22.85	16.50	11.71	2.67	1.32	0.64
23. I heard so many people had problems with statistics which makes me nervous	PC	8.74	16.50	22.43	35.40	16.93	3.35	1.19	0.63
24. I do NOT know what to expect in a statistics course	PC	15.09	26.66	27.64	19.46	11.14	2.85	1.22	0.62
25. The statistical formulas are all gibberish to me.	PC	13.96	27.36	24.82	23.41	10.44	2.89	1.21	0.58
26. I worry about how much of the content I have to learn on my own	PC	8.32	17.49	22.14	36.39	15.66	3.34	1.18	0.56
27. Many people can do statistics better than me	PC	6.63	17.07	30.89	30.75	14.53	3.30	1.11	0.54

Note. SASOH = The Statistical Anxiety Scale in an Online or Hybrid setting; *M* = mean; *SD* = standard deviation. [a]The 27 items measure 4 subdomains of statistics anxiety: Class and Interpretation anxiety (CI), Fear of Asking for help (FA), Online System anxiety (OS), and Pre-Conception (PC). [b]Response score category contain: 1 = No Anxiety, 2 = A Little Anxiety, 3 = Some Anxiety, 4=Moderate Anxiety, 5 = A lot of Anxiety for Item 1-22. Response score category contain: 1 = Strongly Disagree, 2 = Disagree, 3=Neutral, 4 = Agree, 5 = Strongly Agree for Item 23-27. [c]r = item-total correlation.

APPENDIX B

Figure 1. Final CFA Model