IJRES International Journal of Research in Education and Science

www.ijres.net

Confirmatory Factor Analysis of Two Versions of the Statistical Anxiety Scale in a Sample of American Students

Keston G. Lindsay [1] *, Jessica B. Kirby [2], Brynn C. Adamson [3]

[1] University of Colorado Colorado Springs, USA, 0000-0002-2039-7477
[2] University of Colorado Colorado Springs, USA, 0000-0003-0063-6546
[3] University of Colorado Colorado Springs, USA, 0000-0002-1796-4047
* Corresponding author: Keston G. Lindsay (klindsay@uccs.edu)

Article Info

Article History

Received:
7 July 2025

Revised:
18 October 2025

Accepted:
21 December 2025

Published:
1 January 2026

Keywords

Statistics anxiety
Statistics Anxiety Scale
Statistics education
Undergraduate statistics students

Abstract

The objective of this study was to confirm the factor structure for two versions of the Statistics Anxiety Scale (SAS) created for American undergraduate statistics students. To address the inconsistent factor structure of SAS across different populations, Lorenzo-Seva et al. (2022) revised the SAS. This version reduced the number of questions and added a 4th factor called social desirability. They reported a good fit without the need for introducing correlations between error terms, which were outcomes found in different versions of the SAS including this study. This revised SAS considers a 4th dimension called social desirability, suggesting that this dimension could account for much of the error variance found in previous studies. That dimension was not included in this study. A modified three factor model explained construct validity in the original version of the SAS. A modified five factor model explained construct validity in a version of the SAS with additional items. Both versions of the SAS and their factors displayed acceptable levels of fit, were found to have reliability coefficients above .70, and generally shared moderate negative relationships with Wise's Attitude Toward Statistics scale.

Citation: Lindsay, K. G., Kirby, J. B., & Adamson, B. C. (2026). Confirmatory factor analysis of two versions of the statistical anxiety scale in a sample of American students. *International Journal of Research in Education and Science (IJRES), 12*(1), 105-115. https://doi.org/10.46328/ijres.5270

Introduction

Statistical competence is an important part of the curriculum for many fields, including the health sciences, (Aggarwal, 2018; Woltenberg, 2021), the medical professions (Manzar et al., 2023; Schmidt et al., 2021), sociology (Nousak et al., 2024), psychology (Brooks et al., 2024), and others (Berndt et al., 2021). Beyond academia, general statistical literacy is increasingly important as society becomes more data-driven (Johannssen et al., 2021).

Statistics Anxiety

Statistics anxiety is a form of state anxiety characterized by apprehension and fear when engaging with statistics-related content (Onwuegbuzie & Wilson, 2003). This fear has been shown to delay student enrollment in statistics courses, as well as affect successful completion of these classes. Students in applied human sciences and social sciences tend to avoid statistics due to low perceived relevance to their primary content areas (Aggarwal, 2018; Kaufmann et al., 2022; Rajecki et al., 2005), suggesting poor student attitudes toward statistics. It has also been shown that positive attitudes toward statistics share weak to moderate negative relationships with anxiety in several populations (Chiesi et al., 2011; Chew & Dillon, 2014; Lindsay et al., 2024). This phenomenon is so widespread that there have been several instruments created for its measurement (Nolan et al., 2012). Examples of these include Wise's Attitudes Toward Statistics Scale (ATS; Wise, 1985), Survey of Attitudes Toward Statistics (SATS; Schau et al., 1995), the Statistics Anxiety Rating Scale (STARS; Cruise et al., 1985), and the Statistics Anxiety Scale (SAS; Vigil-Colet et al., 2008).

The Statistics Anxiety Scale

The original SAS was originally designed as a 24-item scale, developed by Vigil-Colet and colleagues (2008). The SAS contains three subscales: asking for help anxiety (AHA), examination anxiety (EA), and interpretation anxiety (IA). Asking for help anxiety is the experience of anxiety connected to requesting help from an authority figure (e.g., teacher, tutor). It consists of items 3, 5, 7, 12, 17, 21, 23, & 24. Examination anxiety is anxiety experienced during statistics examinations. It consists of items 1, 4, 9, 11, 13, 14, 15, & 20. Finally, IA is anxiety experienced during the use of formulae and interpretation of data in the statistical context. The survey was face and content validated by instructors with ≥ 10 years of experience, and several of these items (1, 2, 3, 6, 9, 10, 11, 14, 17, 18, 22) were derived from the STARS. The three-factor structure was found in a cohort of 159 psychology students enrolled in a statistics course. The factors and the SAS were shown to be reliable, with Cronbach α coefficients reported at .92, .87, .82, and .91 respectively for AHA, EA, IA, and SAS in unity. Each factor of the SAS has potential ranges between 8-40. Confirmation of this three-factor structure was reported in several later studies. Chiesi et al. (2011) reported acceptable fit in modified three-factor models with Spanish [$\chi^2(246)$ = 556.296, $p < .01$; Comparative Fit Index (CFI) = .926, Root Mean Square Error of Approximation (RMSEA) = .061] and Italian [$\chi^2(248)$ = 668.361, $p < .01$; CFI = .926, RMSEA = .058]. A modified three-factor structure was also confirmed (Chew & Dillon, 2014) in a sample of students from Australia and Singapore [$\chi^2(240]$ = 532.73, CFI = .92, Parsimony Goodness of Fit Index (PGFI) = .66, RMSEA = .08 (90 % CI: .07, .09)]. However, O'Bryant et al. (2021) reported a poor fit with the three-factor model [$\chi^2(71.1) = 153.46$, $p < .001$, CFI = .838, RMSEA = .106, Standardized Root Mean Square Residual (SMSR) = .073, Tucker-Lewis Index (TLI) = .82]. A modified two-factor model was considered a better fit in this study, though it was not statistically significant from the three-factor model [$\chi^2(38.1) = 49.37$, $p = .105$, CFI = .959, RMSEA = .076, Standardized Root Mean Square Residual (SMSR) = .035, TLI = .948]. Given this finding, Lindsay et al. (2024) found a three-factor structure using exploratory factor analysis, which was consistent with the original three-factor model (SAS-O; Lindsay et al., 2024). They also modified SAS-O by adding 6 items (SAS-M): four items that pertained to the use of technology (25-28), and two items that pertained to anxiety experienced when asking a peer a question (29, 30). Items 25-27 loaded onto a new factor, application anxiety (AA), while items 29 & 30 loaded onto another new factor, peer anxiety (PA). Item 28 loaded onto IA.

Wise's Attitude Toward Statistics Scale

The ATS (Wise, 1985) has been shown to be moderately negatively correlated with the SAS in several studies (Chew & Dillon, 2014; Lindsay et al., 2024), thus providing evidence for discriminant validity. The instrument contains 29 items, of which 14 are reverse scored. The items may be summed for totals on two factors (Attitude Toward Course and Attitude toward Field) or a total score in unity. Higher scores indicate better attitudes toward statistics, and the total score was used to establish discriminant validity of SAS in this study.

Hypotheses

The objective of this study was to confirm the factor structure of these versions of SAS. We propose four hypotheses. *H1:* A three-factor model will best describe construct validity of SAS-O. *H2*: A five-factor model will best describe construct validity of SAS-M. *H3*: Both versions of the SAS and their factors will have reliability coefficients above 7.0. *H4*: Both versions of the SAS and their factors will share weak to moderate correlations with the ATS.

Method

This study was approved by the IRB 2021-078-OI at the University of Colorado Colorado Springs. The nominal group technique was used to attain consensus on any items that needed to be reworded or adjusted, as well as the solicitation of new items. The first version of SAS (Vigil-Colet et al., 2008) was used for this process. Face and content validity were addressed at two levels: using instructors, and former statistics students. The instructors were experienced at one or more of the following: teaching undergraduate statistics courses, graduate statistics courses, or were experienced quantitative methodologists (e.g., supervision of quantitative theses). The SAS was independently reviewed by each faculty member, with all potential modifications sent to the moderator, who is an expert at qualitative methodology. All instructors then met with the moderator to discuss all adjusted and new survey items until consensus was reached. New items were added to reflect two general concepts: anxiety surrounding the use of statistical programs (Items 25-28), and anxiety around asking a classmate (i.e., peer) for help. Other changes were made to represent terminology more consistent with American English, as well as multiple platforms. These included online, hybrid, and face-to-face formats. The modified questions have been previously published (Lindsay et al., 2024).

Participants

The data were collected during Fall 2021, Spring 2022, and Fall 2022 semesters. Participants (n = 453) were sampled from statistics classes in universities in the states of Illinois, Texas, and Colorado. Participants reported a mean age of 19.4 ± 2.6 years (n = 450), ranging between 17 to 48 years old. Participants were enrolled in face-to-face (n = 199), online only (n = 212), and hybrid (n = 42) courses. There were 308 students who identified as female, 134 males, five gender non-conforming, and six students who preferred to not answer. There were 101 students who identified as Asian/Asian American, 45 students who identified as Black/African American, 76 students who identified as Hispanic/Latinx, 1 one student who identified as Pacific Islander, and 213 students who identified as White/Caucasian. The most common majors were psychology (n = 31), political science (n = 24), nursing (n = 22), kinesiology (n = 21), elementary education (n = 18), and business (n = 17). There were many other majors across a wide range of areas including family sciences, animal science, biology, earth science, engineering, social work, and other education subdisciplines. There were 59 undecided majors. The statistics courses emphasized the application of statistical tools and methods. Topics included variables, measures of central tendency, dispersion, probability, basic research design, correlation, regression, z-tests, t-tests and analysis of variance.

Data Analysis

Data were analyzed using descriptive analysis, reliability analyses, as well as, confirmatory factor analysis in the structural equation modeling module in JASP 19.2 (JASP Team, 2024). Two confirmatory factor analyses were used to analyze the data, with one being used to analyze the factor structure of the original 24-item scale, and the other for the modified 30-item scale. Both models were evaluated using fit indices for Model Comparison, Model Fit, and Model Parsimony. For SAS-O, a one-factor model was specified, followed by a three-factor model and a modified three-factor model. Fit indices showed a better fit for the three-factor model compared to the one-factor model. A third model showed a better fit when errors were allowed to correlate, when compared to the second model. For this process, the error with the highest modification index (MI) was allowed to correlate. If this resulted in a significant model improvement, the next highest MI was allowed to correlate. This process was continued until there were no statistically significant improvements in the model. Errors were only used if they were between items that were hypothesized to load upon the same factor. A total of ten errors significantly improved the model. For SAS-M, a one-factor model was specified, followed by a five-factor model and a modified five-factor model. Similarly to SAS-O, a five-factor model was better than the one factor model, with an improvement in model fit where errors were allowed to correlate. Similarly to SAS-O, errors were iteratively allowed to correlate in descending order of magnitude until no model improvement was seen. Only errors hypothesized to load upon the same factor were used. For this model, 18 errors significantly improved the model. Pearson's correlations were used to examine relationships between both versions of the SAS, their factors, and ATS. Cronbach's α analyses were used to determine the internal consistency of each factor/subscale.

Results

Table 1 shows descriptive statistics (mean ± standard deviation), for each factor. Reliability analyses are also presented as McDonald's ω (where applicable), and Cronbach's α analyses for each factor. Cronbach's α analysis showed internal reliability for each subscale of O-SAS and M-SAS.

Table 1. Factor Descriptive Statistics and Reliability Measures

	AHA	EA	IA (O)	IA (N)
Cronbach's α	.94	.91	.87	.88
McDonald's ω	.94	.91	.87	.88
Mean ± S.D.	20.0 ± 8.1	30.0 ± 7.3	18.7 ± 6.2	21.0 ± 6.9
Present Range	8-40	9-40	9-42	9-45
Potential range	8-40	8-40	8-40	9-45
	AA	PA	SAS-O	SAS-M
Cronbach's α	.86	.94	.94	.96
McDonald's ω	.86	-	.93	.95
Mean ± S.D.	9.8 ± 3.3	4.7 ± 2.2	68.3 ± 17.5	85.0 ± 21.5
Present Range	3-15	2-10	24-120	30-150
Potential range	3-15	2-10	24-120	30-150

Table 2 shows correlation matrices for the SAS, its factors, and ATS. All subscales shared moderate to strong relationships with one another (Cohen, 1988), and generally moderate relationships with ATS. The models for SAS-O and SAS-M are shown in Figure 1 and Figure 2 respectively.

Table 2. SAS Correlation Matrices

SAS-O	AHA	EA	IAO	SAS-O
Asking for Help Anxiety				
Examination Anxiety	.43			
Interpretation Anxiety (original; IAO)	.59	.43		
SAS-O	.85	.77	.81	
Attitude Toward Statistics Scale (ATS)	-.37	-.39	-.45	-.49

SAS-M	AHA	EA	IAN	AA	PA	SAS-M
EA	.43					
Interpretation Anxiety (new; IAN)	.60	.44				
Application Anxiety	.38	.50	.56			
Peer Anxiety	.58	.30	.53	.30		
SAS-M	.83	.75	.83	.67	.64	
ATS	-.37	-.39	-.46	-.30	-.27	-.49

Note: all relationships significant at the $p < .001$ level.

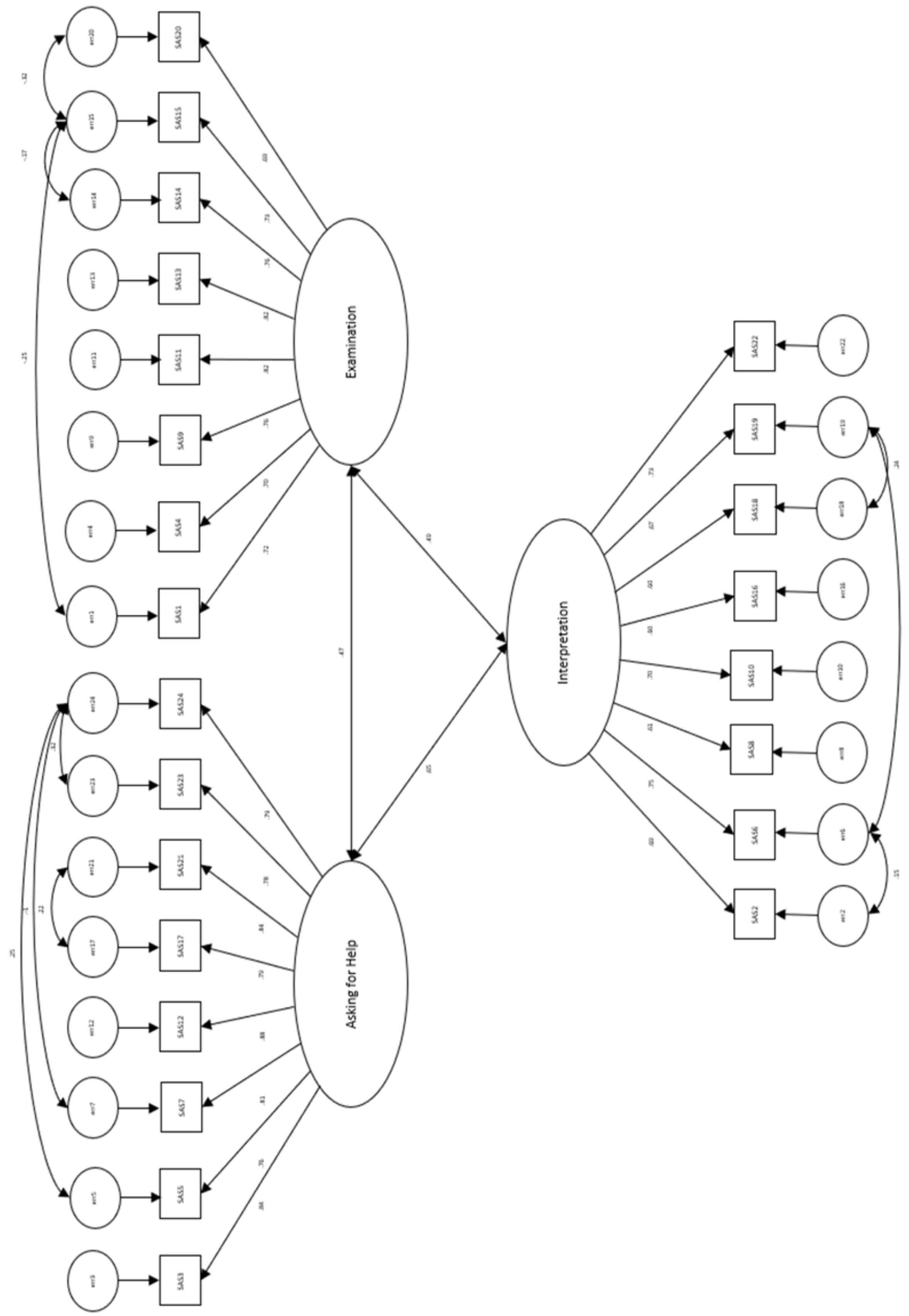

Figure 1. Standardized Estimates for Modified SAS-O Three-Factor Model

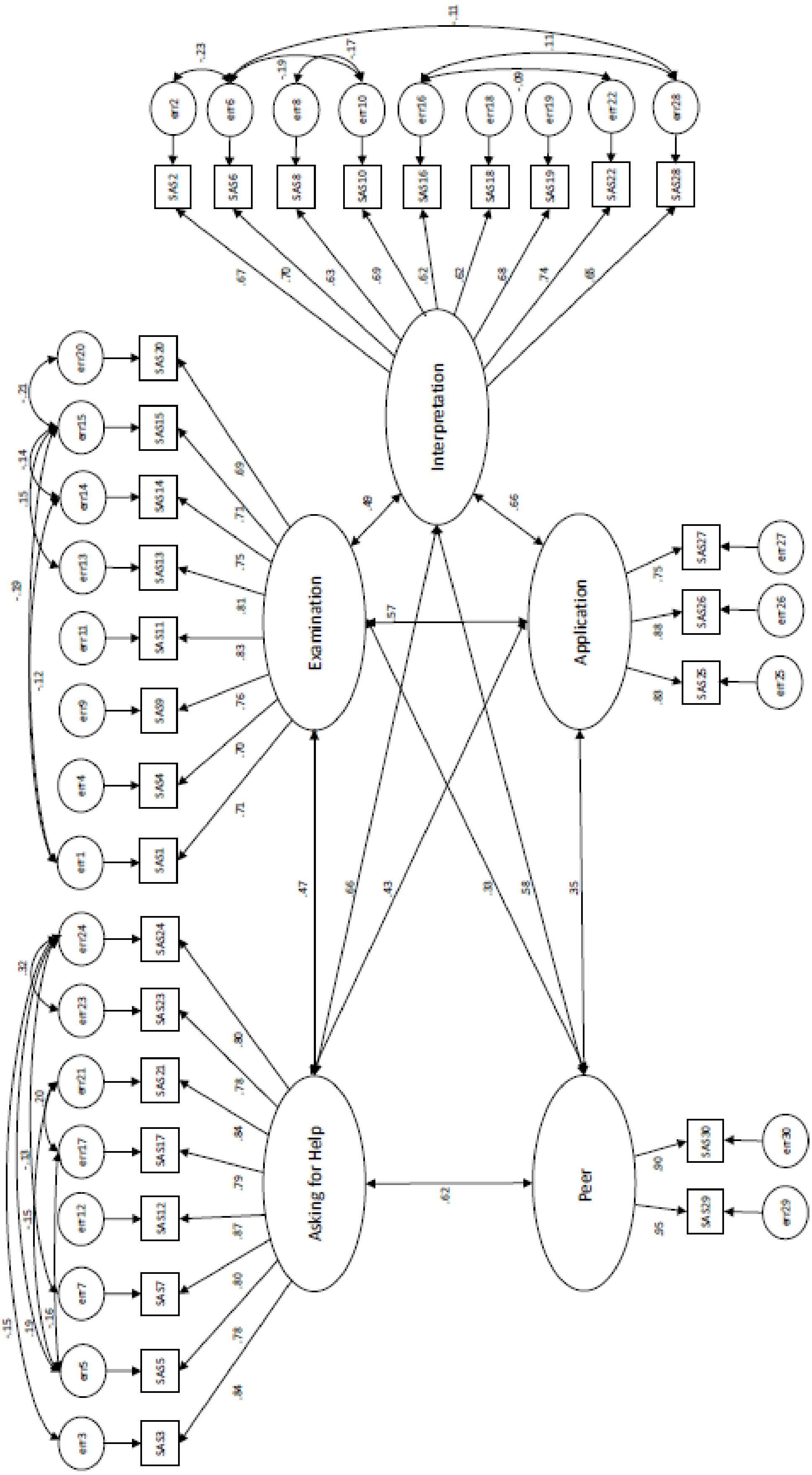

Figure 2. Standardized Estimates for Modified SAS-M Five-Factor Model

Table 3 contains the fit statistics for both confirmatory factor analyses. For both analyses, chi-squared divided by the degrees of freedom test (χ^2 /d. f.) was acceptable, (using ≤ 2 threshold), and the Goodness of Fit Index (GFI) was marginally poor (.95 is good, .90 to .95 is acceptable). Root Mean Square Error of Approximation (RMSEA), Tucker-Lewis Index (TLI), and Comparative Fit Index (CFI) were good for both models (using the ≤ .6 criterion).

Table 3. Fit Statistics of the SAS-O and SAS-M

Model	Fit				Comparison		Parsimony	
	χ^2 (d. f.)	χ^2 /d. f.	GFI	RMSEA	CFI	TLI	PNFI	AIC
1	2892.7 (252)	11.48	0.5	0.15	.61	.58	.54	29673.6
2	917.1(249)	3.68	0.85	0.08	.90	.89	.79	27704
2a	682.9(239)	2.86	0.89	0.058	.94	.93	.78	27489.8
3	4232.8 (405)	10.45	0.48	0.14	.57	.53	.51	37394.5
4	1205.9 (395)	3.05	0.85	0.07	.91	.90	.79	34387.7
4a	910.9(377)	2.42	0.89	0.056	.94	.93	.78	34128.6

Discussion

The objective of this study was to confirm construct validity and reliability of a modified version of the SAS in a sample of American students. As aforementioned, there were four hypotheses: *H1:* A three-factor model will best describe construct validity of SAS-O. *H2***:** A five-factor model will best describe construct validity of SAS-M. *H3***:** Both versions of the SAS and their factors will have reliability coefficients above 7.0. *H4:* Both versions of the SAS and their factors will share weak to moderate correlations with the ATS. There was partial support for *H1*. A three-factor model described SAS-O better than a one factor model. However, a model with 10 error correlations described SAS-O better than the three-factor model. This general outcome is similar to the findings of other studies that have confirmed acceptable models in the three-factor structure of the SAS (Chew & Dillon, 2014).

There was partial support for *H2*. A five-factor model described SAS-M better than the one-factor model. A model with 18 error correlations described SAS-M better than the original five-factor model. The criteria used to describe these findings are fit indices (χ^2 /d. f., GFI, RMSEA), comparison indices (CFI, TLI), and parsimony indices (PNFI, AIC). A nonsignificant χ^2 /d. f. is optimal, but this test has been indicated to be too powerful for research studies. Values ≤ 2 indicate a good fit, and values between 2 and 5 are acceptable. Model s 2, 2a, 4 and 4a fit the latter criterion. The RMSEA values ≤ .06 represent a good fit, .07-.08 represent a moderate fit, .09-.10 represent marginal fit, and ≥ .10 represent a poor fit. Based on this criterion, models 2a and 4a are a good fit, with models 2 and 4 being moderate. In contrast, GFI should be ≥ .95, with values ≥ .90 being of acceptable fit. Based on this criterion, all models were a poor fit. The best models using this criterion were 2a and 4a (GFI = .89) indicating a marginally poor fit. For the CFI and TLI values ≥ .95 represent a good fit, with values ≥ .90 representing an acceptable fit. Based on these criteria, models 2a and 4a were an acceptable fit. Finally, parsimony was achieved in all models using PNFI, as they either met or exceeded the ≥.50 criterion. In summary, models 2a and 4a were a good fit using RMSEA, a marginally poor fit using GFI, and acceptable using all other criteria.

There was support for *H3*, with internal reliability coefficients between .86 and .94 across both versions of the SAS. These are also similar to the findings of previous studies that found internal consistency ranges between .82 and .92 (Vigil-Colet et al., 2008), and .88 to .95 (Chew & Dillon, 2014). There was also support for *H4* with moderate negative relationships between the SAS-O factors and ATS (-.37 to -.45). Similarly, there were moderate negative relationships (-.37 to -.46) between the SAS-M factors and ATS. Both the SAS-O and SAS-M shared a moderate (bordering on high) negative relationship with ATS. Overall, the results of this study support the use of SAS-O and SAS-M to measure statistical anxiety in American undergraduate college students.

Conclusion

In the present study, the SAS-O and SAS-M were shown to have acceptable levels of fit, when errors were allowed to correlate. Their subscales also showed acceptable levels of reliability and discriminant validity. Hence, these versions of the SAS may be used to measure statistics anxiety in American undergraduate student populations.

For both models, even though RMSEA, TLI, CFI, and χ^2 /d.f. were all at least acceptable, the GFI was marginally poor. These results may not be generalized to undergraduate students outside of the United States. These results should also not be generalized to graduate students. SAS-O and SAS-M had similar outcomes. One previous study has shown that SAS-M only explains 4% more variance than SAS-O (Lindsay et al, 2024). Additionally, it should be noted that during the development and analysis of this version of the SAS, the original version was in the process of being revised.

To address the inconsistent factor structure of SAS across different populations, Lorenzo-Seva et al. (2022) revised the SAS. This version reduced the number of questions and added a 4th factor called social desirability. They reported a good fit without the need for introducing correlations between error terms, which were outcomes found in different versions of the SAS including this study. This revised SAS considers a 4th dimension called social desirability, suggesting that this dimension could account for much of the error variance found in previous studies. That dimension was not included in this study. Similarly to the previous SAS, such studies should be replicated to confirm reliability and validity in different populations. However, studies using both exploratory and confirmatory approaches should be repeated in order to compare findings. We also do not expect replicability of the error correlations in both models, as these may be due to data characteristics.

References

Aggarwal, R. (2018). Statistical literacy for healthcare professionals: Why is it important? *Annals of Cardiac Anaesthesia, 21*(4), 349-350. https://doi.org/10.4103/aca.ACA_177_18

Berndt, M., Schmidt, F. M., Sailer, M., Fischer, F., Fischer, M. R., & Zottmann, J. M. (2021). Investigating statistical literacy and scientific reasoning & argumentation in medical-, social sciences-, and economics students. *Learning and Individual Differences*, *86*, 101963. https://doi.org/10.1016/j.lindif.2020.101963

Brooks, P. J., Gravelle, C. D., Zapparrata, N. M., Che, E. S., Lodhi, A. K., Roberts, R., & Brodsky, J. E. (2024). Redesigning the introductory psychology course to support statistical literacy at an open-admissions college. *Scholarship of Teaching and Learning in Psychology*. https://doi.org/10.1037/stl0000407

Chew, P. K. H., & Dillon, D. B. (2014). Reliability and validity of the Statistical Anxiety Scale among students in Singapore and Australia. *Journal of Tropical Psychology, 4*(7), 1-7. https://doi.org/10.1017/jtp.2014.7

Chiesi, F., Primi, C., & Carmona, J. (2011). Measuring statistics anxiety: Cross-country validity of the Statistics Anxiety Scale (SAS). *Journal of Psychoeducational Assessment, 29*, 559-569. https://doi.org/10.1177/0734282911404985

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Cruise, R. J., Cash, R. W., & Bolton, D. L. (1985, August). Development and validation of an instrument to measure statistical anxiety. In *American Statistical Association Proceedings of the Section on Statistical Education* (Vol. 4, No. 3, pp. 92-97). American Statistical Association.

JASP Team (2024). *JASP (Version 19.2)[Computer software]*.

Johannssen, A., Chukhrova, N., Schmal, F., & Stabenow, K. (2021). Statistical Literacy—Misuse of Statistics and Its Consequences. *Journal of Statistics and Data Science Education*, *29*(1), 1–9. https://doi.org/10.1080/10691898.2020.1860727

Kaufmann, L., Ninaus, M., Weiss, E. M., Gruber, W., & Wood, G. (2022). Self-efficacy matters: Influence of students' perceived self-efficacy on statistics anxiety. *Annals of the New York Academy of Sciences*, *1514*(1). https://doi.org/10.1111/nyas.14797

Lindsay, K. G., Kirby, J. B., & Adamson, B. C. (2024). Re-Validation and Exploration of Modified Versions of the Statistics Anxiety Scale Developed for College Students in the United States. *International Journal of Research in Education and Science*, *10*(2), 426–445. https://doi.org/10.46328/ijres.3400

Lorenzo-Seva, U., Vigil-Colet, A. & Ferrando, P. J. (2022). Development of a Revised Version of the Statistical Anxiety Scale. *Psicothema, 34*(4), 562-570. https://doi.org/10.7334/psicothema2022.231

Manzar, S., El Koussaify, J., Garcia, V. C., Ozdag, Y., Akoon, A., Dwyer, C. L., Klena, J. C., & Grandizio, L. C. (2023). Statistical Literacy in Hand and Upper-Extremity Patients. *Journal of Hand Surgery Global Online*, *5*(6), 793–798. https://doi.org/10.1016/j.jhsg.2023.07.009

Nolan, M. M., Beran, t., & Hecker, K. G. (2012). Surveys assessing students' attitudes toward statistics: a systematic review of validity and reliability. *Statistics Education Research Journal*, *11*(2), 103–123. Https://doi.org/10.52041/serj.v11i2.333

Nousak, S., Barry, L., & Fisk, S. R. (2024). Deepening Learning and Addressing Inequalities: A Psychosocial Approach to Improving Statistical Literacy Throughout Sociology Curricula. *Teaching Sociology*, *52*(3), 276–287. https://doi.org/10.1177/0092055x231222280

O'Bryant, M., Batley, P.N., & Onwuegbuzie, A.J. (2021). Validation of an adapted version of the Statistical Anxiety Scale in English and its relationship to attitudes toward statistics. *SAGE Open*, *11*(1). https://doi.org/10.1177/21582440211001378

Onwuegbuzie, A. J., & Wilson, V. A. (2003). Statistics Anxiety: Nature, etiology, antecedents, effects, and treatments--a comprehensive review of the literature. *Teaching in higher education, 8*(2), 195-209. https://doi.org/10.1080/1356251032000052447

Rajecki, D. W., Appleby, D., Williams, C. C., Johnson, K., & Jeschke, M. P. (2005). Statistics can wait: Career plans activity and course preferences of American psychology undergraduates. *Psychology Learning &*

Teaching, 4(2), 83–89. https://doi.org/10.2304/plat.2004.4.2.83

Schau, C., Stevens, J., Dauphinee, T. L., & Vecchio, A. D. (1995). The Development and Validation of the Survey of Attitudes toward Statistics. *Educational and Psychological Measurement*, *55*(5), 868–875. https://doi.org/10.1177/0013164495055005022

Schmidt, F. M., Zottmann, J. M., Sailer, M., Fischer, M. R., & Berndt, M. (2021). Statistical literacy and scientific reasoning & argumentation in physicians. GMS Journal for Medical Education, 38(4), Doc77. https://doi.org/10.3205/zma001473

Vigil-Colet, A., Lorenzo-Seva, U., & Condon, L. (2008). Development and validation of the Statistical Anxiety Scale. *Psicotherma, 20*(1), 174-180. https://doi.org/10.1037/t62688-000

Wise, S. L. (1985). The development and validation of a scale measuring attitudes toward statistics. *Educational and Psychological Measurement, 45*(2),401-405. https://doi.org/10.1177/00131644850450022

Woltenberg, L. N. (2021). Cultivating Statistical Literacy Among Health Professions Students: a Curricular Model. *Medical Science Educator*, *31*(2), 417–422. https://doi.org/10.1007/s40670-021-01256-4