PSYCHOMETRIC EVALUATION OF THE STUDENTS' ATTITUDES TOWARD STATISTICS AND TECHNOLOGY SCALE (SASTSc)

ALYSSA COUNSELL
Toronto Metropolitan University
a.counsell@ryerson.ca

JOSEPH ROVETTI
Western University
jrovetti@uwo.ca

ERIN M. BUCHANAN
Harrisburg University
EBuchanan@harrisburgu.edu

ABSTRACT

The current study sought to evaluate the Students' Attitudes toward Statistics and Technology Scale (SASTSc) in two samples of students taking a statistics course that incorporates statistical software. The SASTSc was given at two time points, once at the beginning of the semester and then again at the end of the semester. Our evaluation included examining competing factor analytic models, examining correlations with related measures, test-retest reliability, and assessing internal consistency. Our results in both samples replicate the scale's proposed factor structure; however, not all items were useful and we propose some changes to the wording of items to potentially improve the scale. Data, analysis scripts, and results are publicly available at https://osf.io/rv64m/.

Keywords: *Statistics education research; Statistics attitudes; Assessment; Scale development; Attitudes toward statistical technology; Statistical software*

1. INTRODUCTION

1.1. ATTITUDES TOWARD STATISTICS AND SOFTWARE

Attitudes toward statistics have received an abundant amount of research attention (Bond et al., 2012; Griffith et al., 2012; Ramirez et al., 2012; Walker & Brakke, 2017). The majority of this research has focused on how attitudes are related to students' course performance (Dempster & McCorry, 2009; Emmioğlu & Capa-Aydin, 2012; Gal & Ginsburg, 1994; Lavidas et al., 2020) or statistics anxiety (Onwuegbuzie, 2004; Onwuegbuzie & Wilson, 2003). Substantially less research has focused on how attitudes toward statistics relate to software usage or examined attitudes toward using technology in statistics courses. The use of statistical software and technology in statistics courses is an important topic because it is becoming more intrinsically tied to statistics courses (e.g., Davidson et al., 2019). In fact, statistical software has become a necessary skill for professionals involved in research and scientific inquiry more broadly (Nolan & Temple Lang, 2010). Modern data analysis in a variety of workplaces requires at least basic proficiency with software because decision making relies on the interpretation of software's end products (i.e., summary statistics and graphics). The Guidelines for Assessment and Instruction in Statistics Education (GAISE) put forth by the American Statistical Association (GAISE College Report ASA Revision Committee, 2016) noted the biggest change in statistics education over the past several decades was the increased prevalence of statistical software. Nolan and Temple Lang (2010) argued that statistical computing is now inextricably linked to modern statistical analysis and should be reflected in curriculum changes. Chance et al. (2007) further argued that technology has had an impact on statistics teaching and education, perhaps more than any other discipline. Hernández (2006) argued that software in social sciences statistics courses is especially important so that students gain a practical skill and avoid complex and sometimes anxiety-provoking mathematical computations. Because statistics courses' coverage varies, researchers and instructors

Statistics Education Research Journal, 21(3). https://doi.org/10.52041/serj.v21i3.29

who assess attitudes toward statistics broadly, may be measuring a mix of attitudes toward software, mathematical computations, or conceptual topics.

Using statistical software in courses has become increasingly common in psychology undergraduate courses (Davidson et al., 2019), but few studies have examined students' attitudes toward statistical software nor the impact of statistical software on students' course performance, learning, or general statistics attitudes. One study investigated using the point-and-click software *SPSS* and found mixed results about its effect on students' grades and attitudes toward statistics in an undergraduate psychology statistics course in an Indonesian university (Jatnika, 2015). Specifically, the author found that after using *SPSS* for the duration of the course, students felt their knowledge and skills in statistics were higher, but their course grades actually decreased compared to previous statistics courses. Brezavšček et al. (2016) examined the use of *SPSS* in undergraduate introductory psychology statistics courses at a Slovenian university. They examined how statistical attitudes influenced actual use of *SPSS* and future intentions to use *SPSS*. In other words, their focus was on predicting intentions to use software rather than how statistical software relates to statistical learning and attitudes. Counsell and Cribbie (2020) examined attitudes toward statistics and statistical software and course performance in students taking a psychology statistics course that uses the software, *R*. They found that students with more positive attitudes toward *R* held more positive general statistics attitudes and had higher grades than students with relatively more negative attitudes. One limitation of these findings is that software attitudes were defined differently because of the software used and lack of psychometric evidence for measures of statistics software attitudes.

1.2. STUDENTS' ATTITUDES TOWARD STATISTICS AND TECHNOLOGY SCALE

In an attempt to capture attitudes toward statistics that tap into the technology component within statistics courses, Anastasiadou (2011) created a measure called the Students' Attitudes toward Statistics and Technology Scale (SASTSc). The author noted the items were created based on student input with the purpose of using as a pre- or post-test measure. Similar to the popular Survey on Attitudes toward Statistics (SATS) by Schau and colleagues (2003; 1995), the SASTSc measures several different domains for attitudes toward statistics. Notably, it includes technology- and software-specific items the SATS does not. Specifically, the SASTSc includes five constructs: a) Statistics Cognitive Competence, positive and negative attitudes concerning a student's knowledge and skills as applied to statistics; b) Technology Cognitive Competence, positive and negative attitudes concerning a student's knowledge and skills as applied to technology; c) Attitudes toward Learning Statistics with Technology, positive and negative attitudes concerning a student's attitudes to learning statistics with technology; d) Value, positive and negative attitudes to the worth and usefulness of statistics in students' personal and professional life; and e) Affect, positive and negative emotions concerning statistics. These attitudinal domains provide useful information to statistics instructors about how their students view learning statistics with technology and whether these attitudes shift as they use the technology throughout the course. Attitudes toward technology generally (subscale b above), could also be used by instructors at the beginning of the course to identify whether additional technological training separate from statistics would be beneficial. Although not assessed, the original author of the SASTSc also noted it may be useful to examine whether more positive attitudes toward learning statistics with technology predict higher acheivement in statistics courses that use software. Counsell and Cribbie (2020) provided some evidence for this relationship.

An independent psychometric evaluation of the SASTSc is important for a number of reasons. First, Anastasiadou (2011) used principal component analysis (PCA) instead of exploratory factor analysis (EFA). PCA is not recommended as a method for examining latent variables; it assumes no measurement error and simply breaks down the sample correlation matrix into a linear combination of components—not factors that mathematically comprise the particular correlation structure seen in the sample. Although PCA results may coincide with EFA results, the analyses have different purposes and are mistakenly used interchangeably (Costello & Osborne, 2005; Preacher & MacCallum, 2003). Second, Anastasiadou (2011) noted that a total score is not appropriate because the different subscales comprise unique attitudinal domains. Anastasiadou also provided reliability evidence by reporting a coefficient alpha (α) of 0.84, which was inappropriate because it measured internal consistency for the entire scale. Further, a number of influential researchers have argued that coefficient α is over-used, not measuring what people think it measures, and not a very good measure of a scale's reliability because of its strict assumptions (Flora, 2020; McNeish, 2018; Sijtsma, 2009; Yang & Green, 2011). At the time of writing this manuscript, the SASTSc had been cited just over 100 times but the only external psychometric evaluation paper involved translating the SASTSc into Persian for use with medical students in Iran (Saki et al., 2016). They found satsifactory reliability across the five subscales (α ranged from 0.78 to 0.88), but their confirmatory factor analysis (CFA) provided somewhat mixed results. For example, Saki et al. (2016) argued their model fit was good, but their reported statistics (NNFI, CFI, RMSEA) did not meet recommended cut-offs (e.g., Hu & Bentler, 1999). Given the language differences, small sample for a CFA (N = 192), and issues described above, more work supporting the SASTSc's measurement properties is crucial for researchers seeking to examine its five constructs.

1.3. STUDY AIMS

The aim of the current study is to examine the application of the SASTSc with two samples of students enrolled in statistics courses that incorporated statistical software. Specifically, we re-examined the scale's factor structure in Sample 1 using EFA instead of PCA, and then applied CFA to Sample 2 to support the results obtained in Sample 1. In both samples, we sought to uncover reliability evidence through coefficients alpha (α) and omega (ω) and test-retest reliability. We also examined the correlations of the SASTSc domains with the related SATS subscale scores.

2. METHOD

2.1. PARTICIPANTS

In total, we had 1060 participants from three academic institutions (two large urban Canadian universities in the same city and one medium sized STEM-focused university in the United States). Of the total participants, 132 were excluded from the sample for one of the following reasons: not taking a statistics course, their course was not using software, did not consent to participate, already participated in the study in a previous semester, dropped out after completing less than 40% of the survey, or explicitly stated their data should not be used in the analyses. For 816 of the participants, students were sampled from both undergraduate and graduate level courses in psychology, sociology, kinesiology, or analytics across three academic semesters (spanning 18 months). The remaining 244 participants data came from a previous project reported in Counsell and Cribbie (2020). Data from the university in the United States were used in the exploratory factor analysis (EFA) while data from the two Canadian universities were pooled into one sample for use in a confirmatory factor analysis (CFA).

EFA sample. Sample 1 included 411 participants from 22 unique courses. All participants came from the same analytics masters program but varied considerably in their undergraduate degrees and statistics experiences. All students were enrolled in the first semester Introduction to Analytics course that covered basic statistical concepts (t-tests, ANOVA, regression). Of the total participants, 171 (41.6%) provided data at both time points, whereas 81 (19.7%) provided data at Time 1only and 159 (38.7%) provided data at Time 2 only. The majority of participants identified as East or South Asian (43.0% and 45.8% respectively). Just over 3% of the participants identified as White, 3.5% as Black, with the remaining racial categories accounting for about 1% each (Latinx, Middle Eastern, Mixed or Other racial identity). The median age in the sample was 28 years (M = 28.44, SD = 3.35). Of the 400 participants who provided gender identity data, 52.5% identified as women, 46.5% as men, and 1.0% as nonbinary or gender non-conforming. All participants used *R* as the primary software in their course.

CFA sample. Sample 2 included 517 participants from 33 unique courses ranging from introductory undergraduate courses to graduate level statistics courses in psychology, sociology, or kinesiology. Of the total participants, 164 (31.7%) provided data at both time points, whereas 200 (38.7%) provided data at Time 1 only and 153 (29.6%) provided data at Time 2 only. The survey used in the Counsell and Cribbie (2020) project did not include racial or ethnic identity as a demographic variable, resulting in 47% missing data on this variable for Sample 2. For the remaining participants (n = 227), their ethnic/racial identity was reported as 59.0% White, 4.8% Black, 3.5% Latinx, 12.8% East Asian, 5.7%

South Asian, 7.9% Middle Eastern, 2.2% Indigenous, and 4.0% with a mixed or other ethnic identity. The median age in the sample was 21 years (M = 22.32, SD = 4.57). Of the 411 participants with gender data, 77.9% identified as women, 20.9% as men, and 1.2% as nonbinary or gender non-conforming. The majority of participants reported using *R* in their course (57.8%); 31.6% used *SPSS*, however, 6.4% used *Excel*, and 4.2% used another software package (e.g., *SAS*, *MATLAB*).

2.2. PROCEDURE

Participants in the Counsell and Cribbie (2020) study filled out a short 10-minute paper and pencil survey during class time at two points in the semester: the first day of class and the last day of class. See their paper for the full procedure details. The EFA data and remaining CFA data were collected from participants who completed an online survey that included the same questions used in the Counsell and Cribbie (2020) study, along with additional measures not relevant to this paper. Instructors of students eligible to participate mentioned the study in class and provided the survey link in the location where course files were housed. Instructors were provided with a script stating they were not involved in the study and student participation was anonymous. A link to the online survey was posted twice: once at the beginning of their statistics course (for completion within the first two weeks of starting) and once again at the end of the term (within the last two weeks of their course finishing). The online survey included quantitative measures as well as qualitative questions whereby participants were able to provide further information about their experiences with learning statistical software in their course (at Time 2 only). Only, the quantitative data relevant to SASTSc and related measures are reported in this paper.

Ethics approval was obtained from all three institutional Research Ethics Boards, and student participation was voluntary. Responses to the survey questions were anonymous. To link a participant's surveys across the two time points, unique numeric ID variables were generated based on answers to a series of non-identifying questions (e.g., name of first pet, number of siblings, birth month). Optional email addresses were collected in a separate survey for incentives. Students in Sample 1 were awarded partial course credit for participation. Participants in Sample 2 were entered into a draw to win one of six $150 gift cards (two per semester).

2.3. MEASURES

Students' Attitudes toward Statistics and Technology Scale (SASTSc). The SASTSc (Anastasiadou, 2011) is a 28-item scale measuring attitudes toward statistics and statistics technology. The scale includes five domains: i) Statistics Cognitive Competence, ii) Technology Cognitive Competence, iii) Attitudes toward Learning Statistics with Technology, iv) Value, and v) Affect. Each of the items is measured with a Likert scale ranging from 1 (*strongly disagree*) to 7 (*strongly agree*). Items are averaged into a composite score for each subscale, whereby higher scores are more positive. An example question from subscale iii) is "I prefer to use technology to evaluate statistical problems." Although the original Item 16 asked about *SPSS* specifically, we adapted the item for broader applicability to students by substituting the word "statistical software" for *SPSS*. All SASTSc items used in the current study are available in the Appendix.

The original author reported satisfactory reliability scores for the five domains (α ranging from 0.74 to 0.90) in their Greek sample. Saki et al. (2016) reported similar reliability scores using their Persian translated version. The factor structure of the English language SASTSc has not been examined with confirmatory factor analysis, and the measure generally has little research investigating its measurement properties (see Nolan et al., 2012).

Survey on Attitudes toward Statistics (SATS). Given some overlap in the subscale constructs, we also measured attitudes toward statistics generally using the SATS (Schau, 2003). The SATS is a 36-item scale with Likert items assessing attitudes toward statistics among students currently taking a statistics course. The original 28-item scale (Schau et al., 1995) measured attitudes across four domains, whereas the revised revision includes six domains. Although correlated with one another, the subscales measure different domains and therefore a single total score for the scale is not appropriate. The subscales are i) Affect, feelings concerning statistics; ii) Cognitive Competence, perceptions of students' own intellectual knowledge and skills in statistics; iii) Value, the value or worth attributed to statistics in students' personal and professional lives; iv) Difficulty, the perceived difficulty of statistics; v) Interest, students' level of individual interest in statistics; and vi) Effort, the amount of work students expend to learn statistics (Schau, 2003). Each item is rated on a 7-point scale expressing the amount with which the student agrees or disagrees with the item. Half of the items are negatively worded and must be reverse coded to compute subscale scores. An example of an item from the Affect subscale includes "I will like statistics" and an example of a negatively worded (i.e., reverse coded) item from the Cognitive Competence subscale is "I will have trouble understanding statistics because of how I think." Higher scores on the subscales reflect a more positive attitude toward that domain with the exception of Difficulty, where higher scores reflect perceiving statistics as easier. The SATS does not have any items related to the use of statistical software.

The SATS domains demonstrate convergent evidence with other related scales, good internal consistency scores, and several research papers support the factor structure proposed [see Nolan et al. (2012) for a review]. In our samples, we found sufficiently high internal consistency for ($\alpha > 0.80$) for the Affect, Cognitive Competence, Value, and Interest domains across both samples and time points. The Difficulty domain had values ranging from 0.59 to 0.76 by sample and time point. In Sample 1, the Effort subscale also had values > 0.80 but in Sample 2, α was 0.65 at Time 1 and 0.71 at Time 2.

3. RESULTS

3.1. SAMPLE 1

Data quality and missingness. Missing data were generally not a problem within each time point. Few participants were missing item level data, which was treated as missing at random. At Time 1, all items were missing less than 1.0% and, at Time 2, item level missing data ranged from 0 to 1.88%. Accordingly, correlation matrices and factor analytic procedures were calculated using pairwise deletion.

We further examined the data for signs of inattention and random responding based on survey completion times, straightlining, i.e., selecting the same value for a large number of items in a row (Johnson, 2005), and extreme inter-item (within participant) response variability (Marjanovic et al., 2015). Data were considered of questionable quality if a participant completed the full survey (which included several other scales not discussed in this paper) in less than 600 seconds, scored greater than 2 *SD* above the mean for individual response variability, or had a straightlining score of more than 10 items on the SATS or 20 items on the SASTSc. Note that the number is lower for the SATS because it has reverse coded items, but the SASTSc does not. These measures resulted in flagging 68 participants at Time 1 as questionable and 92 participants at Time 2 as questionable. Twenty-seven of those participants had questionable data at both time points. We conducted a sensitivity analysis with the factor analysis to see whether including those with questionable data quality changed the results. Overall, the pattern of results did not drastically change; the results in the sample with questionable data, however, produced EFAs with more cross-loadings, and generally, the correlations were stronger. We decided to present the results with the questionable data removed but the results from the full sample can be seen at https://osf.io/rv64m/.

Exploratory factor analyses. The EFA analyses were conducted at each time point using the *psych* package (Revelle, 2020) in *R* version 4.0.2 (https://www.R-project.org/). For all analyses, the unweighted least squares (ULS) factor extraction method was used with an oblimin (i.e., oblique) rotation to allow for correlations between factors. Although items are ordinal (seven categories), reasonable variability and no drastic deviations from normality led to the decision to use regular Pearson correlations in the analyses. We examined scatterplots for obvious nonlinearity patterns and were satisfied that linear relationships adequately described the data. At Time 1, a parallel analysis (Horn, 1965) and the minimum average partial (MAP) test (Velicer, 1976) both suggested a 5-factor solution, while there were six eigenvalues greater than 1.00. At Time 2, the parallel analysis suggested four factors, and the MAP test and eigenvalues > 1.00 criteria suggested a 5-factor solution. We examined 4-, 5-, and 6-factor models at each time point, but, due to space constraints, we present only the 5-factor models for each time point because we believe that the 5-factor solution provides the best balance of interpretational and statistical evidence. All additional analyses, item level correlation matrices, and tables can be found at https://osf.io/rv64m/.

Time 1 EFA results. Table 1 includes the factor loadings, communalities (proportion of variance in each item explained by their respective factor), and proportion of variance accounted for by each factor in the 5-factor solution. Across the five factors, there were minimal cross-loadings and the items loaded reasonably strongly onto the factor that matches the original structure reported in Anastasiadou (2011). Communalities were all greater than 0.40 with the exception of Item 18, which was only 0.15 and did not demonstrate sufficient loadings on any factor. The item, "Statistics makes me overqualified," appears to be a poorly worded item with a distribution that was almost uniform in nature. The five factors accounted for 65% of the variance in the items. Factor correlations are included in the lower diagonal of the correlation matrix in Table 2. The factors are all moderately correlated with one another with the exception of Technology Cognitive Competence and Value (r = 0.10). Of note is the strong correlation between Affect and statistical Cognitive Competence (r = 0.57), which will be elaborated further in the discussion section.

Table 1. Five factor EFA solution for Time 1 data (N = 262)

	Cog Comp	Affect	Stats Tech	Value	Tech Cog Comp	h^2
Item 1	**.811**	.089	.021	-.051	-.010	.720
Item 2	**.884**	-.098	.008	.129	.011	.784
Item 3	**.901**	-.100	.023	.114	.015	.813
Item 4	**.687**	.172	-.022	-.007	.102	.699
Item 5	**.818**	.074	-.007	-.140	.069	.726
Item 6	**.716**	.189	-.041	-.055	.073	.703
Item 7	.174	.039	.003	.018	**.533**	.422
Item 8	-.060	.020	.135	.064	**.724**	.616
Item 9	.147	.080	.050	.003	**.721**	.728
Item 10	.151	.082	.074	.065	**.597**	.589
Item 11	.056	-.100	**.789**	-.001	.185	.740
Item 12	-.013	.024	**.881**	-.063	.045	.787
Item 13	-.142	-.007	**.803**	.053	.104	.696
Item 14	-.040	.118	**.646**	.053	.071	.560
Item 15	.175	.240	**.546**	.065	-.115	.581
Item 16	.190	.203	**.615**	.054	-.198	.595
Item 17	.209	.058	**.423**	.285	-.176	.456
Item 18	.015	.239	.055	.174	-.069	.148
Item 19	.161	.027	.232	**.566**	-.031	.591
Item 20	.014	.085	.081	**.712**	.100	.680
Item 21	-.003	-.005	-.032	**.829**	.038	.669
Item 22	-.012	.190	-.130	**.689**	-.022	.573
Item 23	.016	**.856**	-.012	.050	-.001	.789
Item 24	.021	**.887**	.039	.030	-.045	.845
Item 25	-.033	**.698**	.179	.085	.002	.680
Item 26	.023	**.749**	-.045	.044	.050	.615
Item 27	.005	**.792**	-.014	.087	.069	.745
Item 28	.095	**.698**	.022	-.095	.157	.621
Variance Explained	17%	17%	14%	9%	8%	

Note. EFA was estimated using unweighted least squares (ULS) with oblimin factor rotation; RMR = 0.03; h^2 is the communality; italicized item did not load strongly on any factor; Cog Comp is the Statistical Cognitive Competence domain; Stats Tech is the Learning Statistics with Technology Attitudes domain; Tech Cog Comp is the Technological Cognitive Competence domain.

Table 2. Factor correlations for Time 1 and Time 2 in the EFA sample

	Cog Comp	Affect	Stats Tech	Value	Tech Cog Comp
Cog Comp	----	.61	.47	.33	.47
Affect	*.57*	----	.50	.49	.47
Stats Tech	*.28*	*.43*	----	.37	.50
Value	*.33*	*.54*	*.36*	----	.16
Tech Cog Comp	*.43*	*.31*	*.42*	*.10*	----

Note. The factor correlations at Time 1 are italicized in the lower triangle while Time 2 factor correlations are in plain text in the upper triangle. Cog Comp is the Statistical Cognitive Competence domain; Stats Tech is the Learning Statistics with Technology Attitudes domain; Tech Cog Comp is the Technological Cognitive Competence domain.

Time 2 EFA results. Table 3 includes the factor loadings, communalities, and proportion of variance accounted for by each factor in the 5-factor solution. Generally, the results replicated the factor structure at Time 1 and the original structure reported in Anastasiadou (2011). Again, communalities were all greater than 0.40 with the exception of Item 18, which was only 0.09 and did not demonstrate sufficient loadings on any factor. This item was the same problematic item from the Time 1 EFA results. Although having a moderate communality, Item 17, "Statistics is valuable," did not demonstrate a reasonably sized loading on any single factor, but had a standardized loading around 0.20 on four of the five factors. The five factors accounted for 66% of the variance in the items. Factor correlations are included in the upper diagonal of the correlation matrix in Table 2. Correlations at Time 2 were generally stronger than at Time 1 (especially between Cognitive Competence and Learning Statistics with Technology attitudes, which increased from $r = 0.28$ to $r = 0.47$).

Reliability and convergent evidence. Internal consistency reliability was examined using both coefficients alpha (α; Cronbach, 1951) and omega (ω; McDonald, 1970, 1999). Test-retest reliability was calculated by correlating the Time 1 domain with its Time 2 counterpart. Reliability results can be found in Table 4. All subscales had α or $\omega > 0.80$ with most > 0.90 except for Value at Time 2 ($\alpha = 0.77$). All test-retest correlations were > 0.60, which we believe is indicative of reasonable test-retest reliability because attitudes toward statistics may shift somewhat over the duration of a statistics class. We further examined the correlations of the SASTSc subscales with the SATS subscales within each time point to test for convergent evidence (see Table 5). All correlations of comparable subscales (e.g., Value on SATS and Value on SASTSc) were in the expected directions and moderate to large in magnitude.

Table 3. Five factor EFA solution for Time 2 data (N = 160)

	Cog Comp	Affect	Stats Tech	Value	Tech Cog Comp	h^2
Item 1	**.857**	.039	.065	-.081	-.049	.749
Item 2	**.874**	.007	.030	.074	-.007	.841
Item 3	**.844**	-.044	.094	.104	-.024	.799
Item 4	**.770**	.177	.045	-.073	.039	.823
Item 5	**.798**	.071	-.154	.002	.153	.733
Item 6	**.759**	.026	-.020	.051	.124	.721
Item 7	.234	-.011	.297	-.062	**.426**	.571
Item 8	-.038	.037	-.042	.041	**.741**	.531
Item 9	.086	.048	.129	.028	**.769**	.848
Item 10	.168	.007	.212	.065	**.530**	.622
Item 11	-.002	-.014	**.775**	.003	.105	.682
Item 12	-.020	.044	**.875**	-.015	.050	.825
Item 13	.096	.015	**.709**	-.020	.023	.596
Item 14	-.038	-.011	**.955**	.001	-.032	.843

Item 15	.170	.137	**.434**	.197	.052	.587
Item 16	.089	.179	**.490**	.160	.029	.571
Item 17	.193	.204	.268	.191	-.039	.412
Item 18	-.130	.200	.031	.191	-.008	.089
Item 19	.317	.020	.089	**.509**	-.088	.511
Item 20	.082	.005	.026	**.794**	.071	.734
Item 21	-.083	.015	-.048	**.714**	-.003	.467
Item 22	-.166	.387	.078	**.480**	.076	.549
Item 23	.043	**.895**	.011	.027	-.035	.856
Item 24	.084	**.865**	.034	.033	-.072	.847
Item 25	.121	**.796**	.056	.097	-.055	.867
Item 26	-.097	**.627**	-.030	.017	.181	.442
Item 27	-.099	**.811**	.055	-.055	.094	.649
Item 28	.150	**.653**	-.052	-.062	.225	.695
Variance Explained	18%	17%	15%	8%	8%	

Note. Model was estimated using unweighted least squares (ULS), with oblimin factor rotation; RMR = 0.03; h^2 is the communality; italicized items did not load strongly on any factor; Cog Comp is the Statistical Cognitive Competence domain; Stats Tech is the Learning Statistics with Technology Attitudes domain; Tech Cog Comp is the Technological Cognitive Competence domain.

Table 4. Reliability estimates for composite scores in each sample

	Time 1		Time 2		Test-retest Correlation
	α	ω	α	ω	[95% *CI*]
Sample 1					
Cog Comp	.94	.96	.95	.96	.65 [.52, .75]
Affect	.93	.96	.93	.95	.65 [.53, .75]
Stats Tech	.90	.94	.92	.97	.63 [.49, .73]
Value	.82	.92	.77	.85	.60 [.46, .71]
Tech Cog Comp	.84	.89	.86	.88	.64 [.51, .74]
Sample 2					
Cog Comp	.93	.95	.93	.95	.47 [.34, .59]
Affect	.91	.94	.91	.95	.59 [.47, .69]
Stats Tech	.92	.94	.93	.95	.57 [.45, .67]
Value	.75	.85	.77	.85	.53 [.40, .64]
Tech Cog Comp	.88	.92	.90	.93	.63 [.52, .72]

Note. Sample 1: *N* at Time 1 = 262; *N* at Time 2 = 160; Sample 2: *N* at Time 1 = 335; *N* at Time 2 = 291; Cog Comp is the Statistical Cognitive Competence domain; Stats Tech is the Learning Statistics with Technology Attitudes domain; Tech Cog Comp is the Technological Cognitive Competence domain.

3.2. SAMPLE 2

Data quality and missingness. As described above, missing data were also not a problem within a time point in the CFA sample. In the Time 1 data, item level missingness ranged from 2.09% to 2.99%. At Time 2, each item was missing between 0.34% to 1.37%. Missing data were treated as missing at random. We used pairwise deletion in the correlation analyses and full information maximum likelihood in the CFAs.

The same techniques for data quality described above were used in the CFA sample. Namely, data were examined for signs of inattention and random responding based on survey completion times, straightlining, and extreme inter-item response variability. Data were considered of questionable quality if a participant completed the survey in less than 600 seconds, scored greater than two standard deviations above the mean for individual response variability, or had a straightlining score of more than 10 items on the SATS or 20 items on the SASTSc. These measures flagged 29 participants at Time 1 as questionable and 26 participants at Time 2 as questionable. Only one participant had questionable data at both time points.

We conducted a sensitivity analysis with the CFA models to see whether including those with questionable data quality changed the results. For consistency with the approaches taken in the EFA section, we decided to remove the poor-quality data. We examined scatterplots for obvious nonlinearity patterns and were satisfied that linear relationships adequately described the data and examined histograms for each item, which were approximately normal. We examined multivariate normality with Mardia's test (Mardia, 1970), which was violated. Accordingly, we employed a robust maximum likelihood estimator to correct for multivariate kurtosis.

Confirmatory factor analysis. We sought to replicate the results from the EFA and previous sample. Within each time point, we tested the five-factor model proposed in Anastasiadou (2011) as well as a five-factor model removing Items 17 and 18 from the Value factor. CFA models were conducted with the *lavaan* package version 0.6-6 (Rosseel, 2012) using maximum likelihood estimation with robust standard errors (MLR estimator) and full information maximum likelihood to handle the small amounts of missing data within each time point. Model fit was considered satisfactory if the CFI > 0.95 and SRMR < 0.08 (Hu & Bentler, 1999) and RMSEA < 0.08 (MacCallum et al., 1996). Common convention in psychology is to use various RMSEA cut-offs (Yuan et al., 2016) to indicate the degree of fit such that 0.01 is excellent, 0.05 is close, 0.08 is fair, 0.10 is mediocre, and values greater than 0.10 are indicative of poor fit. Due to space constraints, only standardized factor loadings and factor correlations are presented on the path diagram. Full results can be obtained using the data and analysis scripts at https://osf.io/rv64m/.

Time 1 CFA results. Model 1 includes the five-factor model described in Anastasiadou (2011) with no cross-loadings or residual covariances. This model demonstrated sub-optimal global fit statistics, $\chi^2(340) = 922.31$, $p < 0.001$ (Satorra-Bentler scaling factor = 1.14), CFI = 0.903, RMSEA = 0.077, 90% CI = [0.071, 0.083], SRMR = 0.079. Model 2 followed the same structure as Model 1, but removed Items 17 and 18. Removing these two items slightly improved global model fit, $\chi^2(289) = 732.02$, $p < 0.001$ (Satorra-Bentler scaling factor = 1.13), CFI = 0.924, RMSEA = 0.073, 90% CI = [0.066, 0.079], SRMR = 0.067; the CFI, however, remained below the recommended cut-off for a satisfactory model. Next, we examined local misfit of each model based on the standardized residual matrix. In the first model, items 17 and 18 demonstrated high z-score residuals (> 3.00) with many of the other items, supporting their removal in Model 2. We also found that item 28 ("I am not afraid of statistics") had high residuals with eight other items. Despite being part of the Affect subscale, the item appears to tap more into statistics anxiety (see Chew & Dillon, 2014), which is not captured by any of the domains. Item 19 ("Statistics is a part of our daily life") also demonstrated some high residual covariances with a few of the items but was not as problematic as Item 28. Accordingly, we decided to re-specify the model after removing Item 28 (Model 3). Model 3 demonstrated good model fit: $\chi^2(265) = 587.49$, $p < 0.001$ (Satorra-Bentler scaling factor = 1.13), CFI = 0.942, RMSEA = 0.065, 90% CI = [0.058, 0.072], SRMR = 0.061. Although the CFI is just under the 0.95 cut-off, we believe that this model fits sufficiently well and do not want to arbitrarily add cross-loadings or error covariances to improve model fit past the cut-off because these modifications are less likely to generalize (i.e., may be sample specific). The path diagram for our final model (Model 3) with standardized factor loadings and factor correlations is presented in Figure 1. The factor loadings are all strong (> 0.60) with most in the 0.80s range. The communalities ranged from 0.37 (item 26) to 0.94 (item 24). Factor correlations ranged from 0.25 (Value and Technology Cognitive Competence) to 0.75 (Affect and Cognitive Competence), with most falling in the medium to large range (0.3 to 0.6). These results generally support the factor structure obtained in Sample 1 and by the original author (Anastasiadou, 2011), despite some problematic items.

Table 5. Convergent evidence: Correlations between the SATS and SASTSc domains in Sample 1

	1	2	3	4	5	6	7	8	9	10	11
1. SATS Affect	----	.80	.54	.37	.59	.24	.60	.54	.51	.32	.64
2. SATS CogComp	*.85*	----	.61	.28	.49	.32	.49	.39	.37	.28	.46
3. SATS Value	*.55*	*.58*	----	.09	.60	.29	.41	.37	.41	.47	.49
4. SATS Difficulty	*.34*	*.32*	*.05*	----	.02	-.05	.28	.29	.20	.01	.26
5. SATS Interest	*.52*	*.52*	*.67*	*.05*	----	.37	.59	.41	.51	.49	.67
6. SATS Effort	*.35*	*.36*	*.45*	*-.19*	*.52*	----	.13	.15	.35	.17	.16
7. SASTSc CogComp	*.55*	*.48*	*.33*	*.25*	*.44*	*.25*	----	.60	.56	.44	.65
8. SASTSc TechComp	*.51*	*.40*	*.25*	*.26*	*.35*	*.17*	*.55*	----	.65	.36	.59
9. SASTSc StatTech	*.37*	*.32*	*.35*	*-.03*	*.35*	*.35*	*.38*	*.52*	----	.52	.59
10. SASTSc Value	*.29*	*.22*	*.43*	*-.05*	*.36*	*.28*	*.43*	*.30*	*.51*	----	.60
11. SASTSc Affect	*.63*	*.48*	*.47*	*.17*	*.55*	*.40*	*.60*	*.48*	*.54*	*.63*	----
T1 *M* (*SD*)	5.16	5.45	5.64	3.34	5.93	6.22	5.12	5.27	5.83	5.57	5.63
	(1.14)	(.91)	(.82)	(.74)	(.74)	(.75)	(1.08)	(1.06)	(.82)	(.83)	(.98)
T2 *M (SD)*	4.69	5.17	5.26	3.28	5.61	6.18	5.13	5.26	5.73	5.49	5.42
	(1.19)	(.90)	(.98)	(.69)	(1.00)	(.72)	(.92)	(.94)	(.78)	(.75)	(1.02)

Note. The correlations at Time 1 (N = 262) are italicized and in the lower triangle while Time 2 correlations (N = 160) is in the upper triangle. SATS = Survey of Attitudes toward Statistics (Schau, 2003); SASTSc = Students Attitudes toward Statistics and Technology Scale (Anastasiadou, 2011); CogComp is the Statistical Cognitive Competence domain; StatsTech is the Learning Statistics with Technology Attitudes domain; TechComp is the Technological Cognitive Competence domain.

Time 2 CFA results. The same models from Time 1 were fitted to the Time 2 data. The pattern of global fit results for the same three models examined above was almost identical to their Time 1 counterpart. Namely, Model 1: $\chi^2(340) = 947.60$, $p < 0.001$ (Satorra-Bentler scaling factor = 1.14), CFI = 0.898, RMSEA = 0.084, 90% CI = [0.077, 0.090], SRMR = 0.092; Model 2: $\chi^2(289) = 746.24$, $p < 0.001$ (Satorra-Bentler scaling factor = 1.16), CFI = 0.919, RMSEA = 0.079, 90% CI = [0.072, 0.086], SRMR = 0.078; Model 3: $\chi^2(265) = 601.90$, $p < 0.001$ (Satorra-Bentler scaling factor = 1.17), CFI = 0.937, RMSEA = 0.071, 90% CI = [0.064, 0.079], SRMR = 0.069. The strength of the loadings and magnitude of communalities was also comparable to what was observed in the Time 1 CFA models. Item 26 demonstrated the weakest relationship with its factor (Affect); it had a factor loading of 0.57 and communality of 0.32. All of the remaining factor loadings were > 0.60, with most in the 0.80s range, and all communalities were > 0.45. The factor correlations were all > 0.3, ranging from 0.30 (Value and Technology Cognitive Competence) to 0.69 (Affect and Cognitive Competence). The same factor correlation pattern was observed as in Time 1 but the correlations on average were slightly stronger. We do not present the path diagram for this model again to avoid redundancy, but the full details are available at https://osf.io/rv64m/.

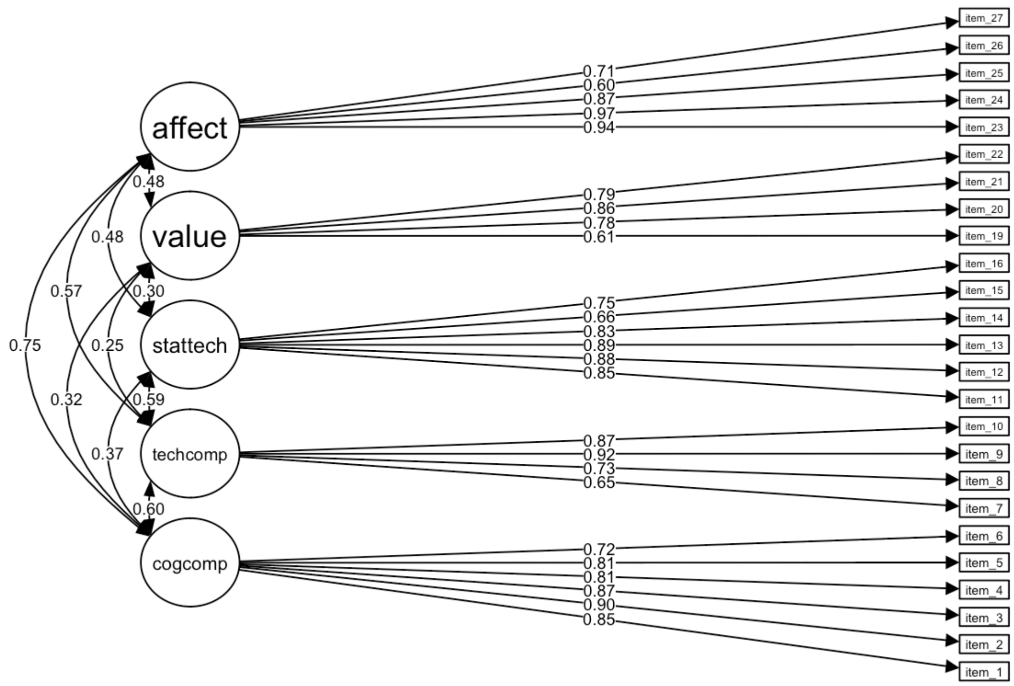

Figure 1. Time 1 CFA model with standardized factor loadings and correlations (N = 335)

***Reliability and convergent evidence**.* We used the same methods for examining reliability and convergent evidence in Sample 2 as used in Sample 1. Reliability results for both time points can be found in Table 4. All subscales had α or $\omega > 0.85$ with most > 0.90, except for Value at ($\alpha = 0.75$ at Time 1 and $\alpha = 0.77$ at Time 2). Test-retest correlations were > 0.50, except for Cognitive Competence ($r = 0.47$). These correlations were slightly lower on average, compared to Sample 1. That said, we still believe the magnitude is indicative of adequate levels of test-retest reliability because attitudes may shift somewhat over the course of taking a statistics class. We further examined the correlations of the SASTSc subscales with the SATS subscales within each time point to test for convergent evidence. All correlations of comparable subscales (e.g., Value on SATS and Value on SASTSc) were in the expected directions and moderate to large in magnitude. Of note is that the correlational pattern was somewhat different from Sample 1. Notable differences on the SASTSc subscales include Cognitive Competence and Affect were more strongly related in Sample 2, especially at Time 1; Affect and Value were more strongly related in Sample 1 at Time 1 only; and Value and Learning Statistics with Technology were more strongly related in Sample 1 at Time 1 only. See Table 6 for the Sample 2 SATS and SASTSc correlations.

Table 6. Convergent evidence: Pearson correlations between the SATS and SASTSc domains in Sample 2

	1	2	3	4	5	6	7	8	9	10	11
1. SATS Affect	----	.85	.58	.62	.62	.06	.74	.42	.38	.37	.78
2. SATS CogComp	*.80*	----	.57	.58	.58	.13	.78	.48	.41	.38	.72
3. SATS Value	*.45*	*.52*	----	.27	.81	.17	.52	.35	.54	.63	.65
4. SATS Difficulty	*.61*	*.59*	*.26*	----	.23	-.13	.47	.29	.11	.18	.46
5. SATS Interest	*.57*	*.51*	*.73*	*.26*	----	.18	.58	.38	.56	.58	.77
6. SATS Effort	*-.01*	*.08*	*.13*	*-.04*	*.11*	----	.22	.08	.14	.07	.21
7. SASTSc CogComp	*.72*	*.75*	*.47*	*.48*	*.56*	*.05*	----	.58	.44	.38	.75
8. SASTSc TechComp	*.43*	*.37*	*.29*	*.35*	*.36*	*.02*	*.51*	----	.56	.30	.50
9. SASTSc StatTech	*.33*	*.26*	*.35*	*.19*	*.40*	*.07*	*.37*	*.55*	----	.53	.50
10. SASTSc Value	*.27*	*.34*	*.59*	*.24*	*.51*	*.06*	*.40*	*.29*	*.34*	----	.54
11. SASTSc Affect	*.78*	*.68*	*.56*	*.50*	*.72*	*.06*	*.78*	*.54*	*.46*	*.49*	----
T1 *M* (*SD*)	3.84	4.68	5.09	3.26	5.00	6.60	4.09	4.21	4.77	4.78	4.11
	(1.24)	(1.14)	(1.04)	(.75)	(1.35)	(.55)	(1.32)	(1.30)	(1.20)	(.93)	(1.37)
T2 *M (SD)*	4.13	4.96	5.21	3.41	4.97	6.24	4.59	4.51	4.98	4.96	4.67
	(1.36)	(1.12)	(1.19)	(.88)	(1.49)	(.82)	(1.29)	(1.38)	(1.43)	(1.02)	(1.37)

Note. The correlations at Time 1 (*N* = 327 with pairwise deletion) are italicized and in the lower triangle while Time 2 correlations (*N* = 290 with pairwise deletion) is in the upper triangle. SATS = Survey of Attitudes toward Statistics (Schau, 2003); SASTSc = Students Attitudes toward Statistics and Technology Scale (Anastasiadou, 2011); CogComp is the Statistical Cognitive Competence domain; StatsTech is the Learning Statistics with Technology Attitudes domain; TechComp is the Technological Cognitive Competence domain.

4. DISCUSSION AND CONCLUSION

4.1. SUMMARY OF KEY FINDINGS

The current study sought to examine the psychometric properties of the Students' Attitudes toward Statistics and Technology Scale (SASTSc) in a diverse sample of students taking a statistics course. We found the factor structure mostly replicated Anastasiadou's (2011) results (i.e., a five factor structure with the same items loading on the same factors), but there are some problematic items that warrant revision or removal. In both samples, Item 18 ("Statistics makes me overqualified") was problematic (low factor loadings and communalities). This item lacks specificity, is difficult to interpret, and with the importance of statistics skills for a variety of jobs, it is unlikely that statistical skills would make one overqualified for most positions that students would consider post-degree. Further, in Anastasiadou's (2011) paper, it was not clear whether the original scale was administered in Greek or in English (the sample consisted of Greek students). Issues with this item, therefore, may arise due to differences in translation or between native and non-native English speakers. A potential revision to the item that is more consistent with the other Value items is "Statistics skills make me more qualified for jobs." Item 17 also produced issues in the factor analytic solutions (at Time 2) with low loadings and communalities. This item ("Statistics is valuable") may be too broad. It is not clear whether the item is asking about the value of statistics for the student or whether generally speaking statistics is valuable. One potential revision could be "Statistics skills are valuable for me." Items 19 and 28 demonstrated sufficiently high loadings and communalities but were also somewhat problematic in that they had higher cross-loadings in the EFA or high residual correlations with other items in the CFA. We recommend removing Item 28 ("I am not afraid of statistics") because the item appears to tap more into statistics anxiety than statistics affect. As discussed in both Chew and Dillon (2014) and Nasser (2004), much of the attitudes literature fails to distinguish between statistics "attitudes" and "anxiety" and may use the terms interchangeably. Lastly, we recommend revising Item 19 ("Statistics is a part of our daily life") to "Statistics is relevant to my daily life" to tap into each participant's view of statistics as relevant to that person rather than a broad and uncertain population (i.e., "our" daily life). All of these proposed changes are summarized in the Appendix.

We found good internal consistency in our sample for each of the five domains of the SASTSc and convergent evidence with the domains in the SATS. Of note is that the Value domain had the lowest reliability scores. The problematic items—17, 18, and 19—are all part of this domain. Although the α and ω scores are considered adequate support of internal consistency, improving the specificity of the items would likely improve these scores because the content of the revised items should be more consistent with the other items on the same domain. The remaining domains had scores in the 0.80s or 0.90s range providing strong evidence for the consistency of items on similar domains. Correlations between the SASTSc domains and relevant SATS domains were generally strong in magnitude (e.g., $r = 0.48$ to 0.67) supporting the convergent evidence of the SASTSc domains. The SASTSc Value domain, though had generally lower correlations with the other domains (e.g., only $r = 0.43$ with the SATS Value domain) and small to medium correlations with other domains, providing further support for the necessity of revising some of the items on this subscale.

Despite the psychometric evidence in favour of using the SASTSc in a diverse sample of students, there remain some additional topics that warrant discussion. Anastasiadou (2011) discusses how the measure can be used as pre-post assessment for statistics courses, but our results found a few differences in correlational patterns at the beginning vs. end of a statistics course. At the end of the course, Item 17 had moderate cross loadings across the domains instead of one prominent one on Value. The broadness of the original item may have appealed to students before they had a lot of experiences with statistics and software (i.e., at Time 1), but produced more ambiguity at Time 2 where they may be thinking about whether statistics is valuable to them personally, or on a societal level. Because Value and Affect are moderately correlated, the students' feelings toward their statistics course may be intertwined with their perceived value of statistics. We were somewhat surprised there were not more psychometric differences across the time points, as atititudes have the potential to shift, but the relationships between the SASTSc and SATS domains were stronger or weaker depending on the time point and sample.

Another issue was that some of the domains were so highly correlated with one another, introducing some skepticism about them being separate theoretical constructs. Specifically, the Affect and

Cognitive Competence domains were highly related (e.g., correlations ranging from 0.57 to 0.75 across time points and samples). Because the higher correlations tended to occur at Time 1, this result supports previous research about how attitudes toward statistics are strongly tied to students' perceptions of their statistical ability and their preconceived ideas about statistics beforetaking a course (Bond et al., 2012; Dempster & McCorry, 2009; Evans, 2007). Correlations closer to 0.5 or 0.6 at Time 2 may suggest some attitudinal shifts after taking a statistics course. Lastly, we had some concerns about the wording or language of some of the items, even though the factor analytic results were not problematic. The Technology Cognitive Competence subscale and Learning Statistics with Technology subscale include terms like *technology* and *software*, which may be interpreted differently by participants. It is not clear whether technology relates to topics such as cell phones, computers, tablets, applications, statistical software, and so on. Accordingly, it may be advantageous to increase the specificity of these items to clarify what technology means, perhaps changing *technology* to *statistical software* in Items 11–15, as is used in Item 16.

4.2. LIMITATIONS

It is important to note the limitations of the current study. First, because we changed the item that referenced *SPSS* to reference "statistical software" instead, our results are not perfectly comparable to those of Anastasiadou (2011). Due to potential translational differences and the need for adapting the measure for software other than *SPSS*, however, we believe the modifications were justifiable. Second, differences in psychometric properties observed in our student sample depending on the software used is possible. Although we do not discuss it in the paper, we did some exploratory analyses, separating Sample 2 into two sub-samples based on whether the software used was syntax based (e.g., *R*, *SAS*) or point-and-click (i.e., *Excel*, *SPSS*) and re-examined the five factor solution. The two subsamples demonstrated similar factor loading patterns, but the loadings in the point-and-click software group tended to be slightly lower, on average. Future studies should seek to examine potential differences further. Lastly, to obtain large enough samples for EFA and CFA, we needed to use samples from different classrooms and universities, which necessarily introduced heterogeneity into the sample without explicitly modeling the sources of that heterogeneity. The samples differed by ethnic and gender identity as well as by study major. Despite these differences, we were able to replicate the same factor structure pattern and reliability results, which confirms the robustness of the findings.

4.3. FUTURE DIRECTIONS

Given the issues described above for Item 17, 18, and 19, we recommend re-examining the factor structure for the Value subscale. Additionally, we believe more work is needed to examine the qualitative differences between the Affect and Cognitive Competence domains. One suggestion is to use the response process evaluation method outlined in Wolf et al. (2019). In that procedure, participants respond to smaller batches of items and provide detailed information about how they intepret and understand those items. This information helps to ensure participants respond to items in a way consistent with their intended use (i.e., provide construct validity evidence). Another avenue is to collect qualitative data through interviews or open ended questions about statistics and software attitudes to further refine the items. Lastly, many of the items could be adapted and evaluated for individuals not taking a statistics course at the time of administration.

4.4. CONCLUDING REMARKS FOR STATISTICS INSTRUCTORS

Statistical software and technology are some of the most useful aspects of statistics courses (Nolan & Temple Lang, 2010), particularly for students in mathematics adverse fields like the social sciences (Hernández, 2006). In fact, learning theory research (e.g., Eccles & Wigfield, 2002) stipulates that students learn and perform best in courses in which they feel the course material has value and they are able to succeed. Accordingly, assessing how students build software competence and self-efficacy and their attitudes toward statistical software is another important step in predicting students' course performance and general attitudes toward statistics. Building positive experiences (including competence) with statistical software is also important for preparing students for the reality of data analysis in professional or graduate programs or non-academic careers (industry, government, etc.).

The results of our study support using the SASTSc for assessing both students' attitudes toward technology and statistics software as well as more traditional statistics attitudes (i.e., affect, statistics cognitive competence). The Value subscale, however, may not be interpretable in its current form due to the issues with numerous items. For all other domains, with a minor modification of Item 16, instructors can easily incorporate the SASTSc into their courses to examine how attitudes toward statistics, technology, and learning statistics with software change over the duration of a statistics course. Although the original scale had one item particular to *SPSS*, we have modified it to apply more generally to *statistical software*, but instructors could further adapt it to the specific software they use in their class (*Excel*, *SPSS*, *R*, *SAS*, etc.). Importantly, the measure is brief enough that it only takes a couple of minutes to complete, allowing for its use in the classroom. Lastly, collecting student data with the SASTSc may be useful for decision-making about the particular software or course content if the majority of student attitudes in the (technology) cognitive competence and learning statistics with technology domains are overwhelmingly negative and unchanging by the end of the course.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the financial support provided to Alyssa Counsell from the Social Sciences and Humanities Research Council (Insight Development Grant 2019-2021). We would also like to thank the course instructors from the three institutions for posting the study information to their students and Linda Farmus for her role in helping with data collection for the project.

REFERENCES

Anastasiadou, S. D. (2011). Reliability and validity testing of a new scale for measuring attitudes toward learning statistics with technology. *Acta Didactica Napocensia*, *4*(1), 1–10.

Bond, M., Perkins, S., & Ramirez, C. (2012). Students' perceptions of statistics: An exploration of attitudes, conceptualizations, and content knowledge of statistics. *Statistics Education Research Journal*, *11*(2), 6–25. https://doi.org/10.52041/serj.v11i2.325

Brezavšček, A., Šparl, P., & Žnidaršič, A. (2016). Factors influencing the behavioural intention to use statistical software: The perspective of the Slovenian students of social sciences. *EURASIA Journal of Mathematics, Science and Technology Education*, *13*(3), 953–986. https://doi.org/10.12973/eurasia.2017.00652a

Chance, B., Ben-Zvi, D., Garfield, J., & Medina, E. (2007). The role of technology in improving student learning of statistics. *Technology Innovations in Statistics Education*, *1*(1), 1–26. https://doi.org/10.5070/T511000026

Chew, P. K. H., & Dillon, D. B. (2014). Statistics anxiety update: Refining the construct and recommendations for a new research agenda. *Perspectives on Psychological Science*, *9*(2), 196–208. https://doi.org/10.1177/1745691613518077

Counsell, A., & Cribbie, R. (2020). Student attitudes toward learning statistics with R. *Psychology Teaching Review*, *26*(2), 36–56. https://doi.org/10.31234/osf.io/76w2p

Costello, A. B., & Osborne, J. (2005). Best practices in exploratory factor analysis: Four recommendations for getting the most from your analysis. *Practical Assessment, Research, and Evaluation*, *10*(1), 1–9. https://doi.org/10.7275/jyj1-4868

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334. https://doi.org/10.1007/BF02310555

Davidson, H., Jabbari, Y., Patton, H., O'Hagan, F., Peters, K., & Cribbie, R. (2019). Statistical software use in Canadian university courses: Current trends and future directions. *Teaching of Psychology*, *46*(3), 246–250. https://doi.org/10.1177/0098628319853940

Dempster, M., & McCorry, N. K. (2009). The role of previous experience and attitudes toward statistics in statistics assessment outcomes among undergraduate psychology students. *Journal of Statistics Education*, *17*(2). https://doi.org/10.1080/10691898.2009.11889515

Eccles, J. S., & Wigfield, A. (2002). Motivational beliefs, values, and goals. *Annual Review of Psychology*, *53*(1), 109–132. https://doi.org/10.1146/annurev.psych.53.100901.135153

Emmioğlu, E., & Capa-Aydin, Y. (2012). Attitudes and achievement in statistics: A meta-analysis study. *Statistics Education Research Journal*, *11*(2). https://doi.org/10.52041/serj.v11i2.332

Evans, B. (2007). Student attitudes, conceptions, and achievement in introductory undergraduate college statistics. *The Mathematics Educator*, *17*(2), 24–30.

Flora, D. B. (2020). Your coefficient alpha is probably wrong, but which coefficient omega is right? A tutorial on using R to obtain better reliability estimates. *Advances in Methods and Practices in Psychological Science*, *3*(4), 484-501. https://doi.org/10.1177/2515245920951747

GAISE College Report ASA Revision Committee. (2016). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016*. American Statistical Association. http://www.amstat.org/education/gaise

Gal, I., & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework. *Journal of Statistics Education*, *2*(2). https://doi.org/10.1080/10691898.1994.11910471

Griffith, J. D., Adams, L. T., Gu, L. L., Hart, C. L., & Nichols-Whitehead, P. (2012). Students' attitudes toward statistics across the disciplines: A mixed-methods approach. *Statistics Education Research Journal*, *11*(2), 45–56. https://doi.org/10.52041/serj.v11i2.328

Hernández, O. (2006). Teaching statistics to undergraduate students of the social sciences. In A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education. Proceedings of the 7th International Conference on Teaching Statistics* (ICOTS7), Salvador, Bahia, Brazil, July. 2–7. International Statistical Institute.

Horn, J. L. (1965). A rationale and test for the number of factors in factor analysis. *Psychometrika*, *30*(2), 179–185. https://doi.org/10.1007/BF02289447

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

Jatnika, R. (2015). The effect of SPSS course to students attitudes toward statistics and achievement in statistics. *International Journal of Information and Education Technology*, *5*(11), 818–821. https://doi.org/10.7763/IJIET.2015.V5.618

Johnson, J. A. (2005). Ascertaining the validity of individual protocols from Web-based personality inventories. *Journal of Research in Personality*, *39*(1), 103–129. https://doi.org/10.1016/j.jrp.2004.09.009

Lavidas, K., Barkatsas, T., Manesis, D., & Gialamas, V. (2020). A structural equation model investigating the impact of tertiary students' attitudes toward statistics, perceived competence at mathematics, and engagement on statistics performance. *Statistics Education Research Journal*, *19*(2), 27–41. https://doi.org/10.52041/serj.v19i2.108

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149. https://doi.org/10.1037//1082-989X.1.2.130

Mardia, K. V. (1970). Measures of multivariate skewness and kurtosis with applications. *Biometrika*, *57*(3), 519–530. https://doi.org/10.1093/biomet/57.3.519

Marjanovic, Z., Holden, R., Struthers, W., Cribbie, R., & Greenglass, E. (2015). The inter-item standard deviation (ISD): An index that discriminates between conscientious and random responders. *Personality and Individual Differences*, *84*, 79–83. https://doi.org/10.1016/j.paid.2014.08.021

McDonald, R. P. (1970). The theoretical foundations of principal factor analysis, canonical factor analysis, and alpha factor analysis. *British Journal of Mathematical and Statistical Psychology*, *23*(1), 1–21. https://doi.org/10.1111/j.2044-8317.1970.tb00432.x

McDonald, R. P. (1999). *Test theory: A unified approach*. Lawrence Erlbaum Associations, Publishers.

McNeish, D. (2018). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*, *23*(3), 412–433. https://doi.org/10.1037/met0000144

Nasser, F. M. (2004). Structural model of the effects of cognitive and affective factors on the achievement of Arabic-speaking pre-service teachers in introductory statistics. *Journal of Statistics Education*, *12*(1). https://doi.org/10.1080/10691898.2004.11910717

Nolan, D., & Temple Lang, D. (2010). Computing in the statistics curricula. *The American Statistician*, *64*(2), 97–107. https://doi.org/10.1198/tast.2010.09132

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
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., & 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
Preacher, K. J., & MacCallum, R. C. (2003). Repairing Tom Swift's electric factor analysis machine. *Understanding Statistics: Statistical Issues in Psychology, Education, and the Social Sciences*, *2*(1), 13–43. https://doi.org/10.1207/S15328031US0201_02
Ramirez, C., Schau, C., & Emmioglu, E. (2012). The importance of attitudes in statistics education. *Statistics Education Research Journal*, *11*(2), 57–70. https://doi.org/10.52041/serj.v11i2.329
Revelle, W. (2020) psych: Procedures for personality and psychological research. Northwestern University. R package version 2.0.9. https://CRAN.R-project.org/package=psych
Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of Statistical Software*, *48*(2), 1–36.
Saki, A., Tabesh, H., Yousefi, R., & Khalili, S. (2016). Clustering the attitudes towards statistics and technology among medical post graduate students. *British Journal of Applied Science and Technology*, *18*(4), 1–11. https://doi.org/10.9734/BJAST/2016/29673
Schau, C. (2003). *Survey of Attitudes Toward Statistics—36*. CS Consultants, LLC. www.evaluationandstatistics.com
Schau, C., Stevens, J., Dauphinee, T. L., & Del Vecchio, A. (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
Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika*, *74*(1), 107–120. https://doi.org/10.1007/s11336-008-9101-0
Velicer, W. (1976). Determining the number of components from the matrix of partial correlations. *Psychometrika*, *41*, 321–327. https://doi.org/10.1007/BF02293557
Walker, E. R., & Brakke, K. E. (2017). Undergraduate psychology students' efficacy and attitudes across introductory and advanced statistics courses. *Scholarship of Teaching and Learning in Psychology*, *3*(2), 132–140. https://doi.org/10.1037/stl0000088
Wolf, M. G., Ihm, E., Maul, A., & Taves, A. (2019). *Survey item validation*. PsyArXiv Preprints. https://doi.org/10.31234/osf.io/k27w3
Yang, Y., & Green, S. B. (2011). Coefficient alpha: A reliability coefficient for the 21st century? *Journal of Psychoeducational Assessment*, *29*(4), 377–392. https://doi.org/10.1177/0734282911406668
Yuan, K.-H., Chan, W., Marcoulides, G. A., & Bentler, P. M. (2016). Assessing structural equation models by equivalence testing with adjusted fit indexes. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(3), 319–330. https://doi.org/10.1080/10705511.2015.1065414

ALYSSA COUNSELL
Toronto Metropolitan University
Department of Psychology
350 Victoria Street
Toronto, ON, M5B 2K3

Appendix

Students' Attitudes toward Statistics and Technology items used and proposed wording changes

Cognitive Competence

1. I am confident with statistics.
2. I can understand statistical reasoning easily.
3. I can understand statistical inference easily.
4. I can learn statistics easily.
5. I can solve difficult statistical test-hypothesis problems.
6. I get high marks in statistics.

Technology Cognitive Competence

7. I am very good with computers.
8. I don't have problems using software.
9. I can easily run statistical software.
10. I can fix problems that arise while using statistical software.

Learning Statistics with Technology

11. Technology makes the learning of statistics easier.
12. Technology makes the learning of statistics more interesting.
13. Technology helps me to understand statistics.
14. I prefer to use technology to evaluate statistical problems.
15. I like to use technology to make statistical graphs.
16. Statistical software helps me to discover many different statistical applications.

Value

17. Statistics is valuable. *
18. Statistics makes me overqualified.*
19. Statistics is a part of our daily life. *
20. Statistics helps me to understand our economy.
21. Statistics helps me to understand politics.
22. Statistics helps me to understand news reports.

Affect

23. Learning statistics is enjoyable.
24. I like learning statistics.
25. Statistics is interesting.
26. Statistics is not a frustrating discipline.
27. I find solving statistical problems satisfying.
28. I am not afraid of statistics. *

Proposed Changes to Starred Items

17. Statistics skills are valuable for me.
18. Statistics skills make me more qualified for jobs. [possibly remove item]
19. Statistics is relevant to my daily life.
28. Remove item due to construct inconsistency