EXPLORING METHOD EFFECTS IN THE SIX-FACTOR STRUCTURE OF THE SURVEY OF ATTITUDES TOWARD STATISTICS (SATS-36)

CHAO XU
University of Houston-Clear Lake
hzxuchao@gmail.com

CANDACE SCHAU
CS Consultants, LLC
Professor Emerita, University of New Mexico
cschau@comcast.net

ABSTRACT

We use ordinal confirmatory factor analysis techniques to investigate the six-factor structure of the Survey of Attitudes Toward Statistics (SATS-36) and to estimate method effects associated with its items and factors. We extend previous confirmatory research to include posttest, as well as pretest, item-level data. We also investigate method effects by adding a common method factor to the original six-factor model. Interestingly, results reveal noticeable proportions of common variance associated with the Difficulty construct only at pretest and with the Value and Interest constructs only at posttest. We examine the characteristics of the SATS items as a possible source for these method effects.

Keywords: *Statistics education research; Statistics attitudes; Ordinal confirmatory factor analysis*

1. INTRODUCTION

The number of undergraduate students enrolling in introductory statistics courses is growing and will continue to grow because conceptual understanding of statistics is critical in life and data-analytical skills are indispensable in many professions. Thus, it is important to accurately assess these student outcomes when they finish statistics courses. Students' attitudes toward statistics have been used to predict course achievement but, more importantly, are a vital course outcome (Carlson & Winquist, 2011; Chiesi & Primi, 2010; Garfield, Hogg, Schau, & Whittinghill, 2002; Hood, Creed, & Neumann, 2012; Paul & Cunnington, 2017; Ramirez, Schau, & Emmíoğlu, 2012; Schau, 2003; Tempelaar, Schim van der Loeff, & Gijselaers, 2007). The Survey of Attitudes Toward Statistics (SATS) is a widely used measure for assessing statistics attitudes, in research as well as in course and instructional evaluation. As such, it is important that the items and components in the SATS accurately assess students' attitudes.

The original version of the SATS, developed in the early 1990s, was designed to measure four interrelated constructs: *Affect*, *Cognitive Competence*, *Value*, and *Difficulty* (Schau, Stevens, Dauphinee, & Vecchio, 1995). The SATS-28 was revised into the SATS-36 by adding eight more items that measure two additional constructs: *Effort* and *Interest* (https://www.evaluationandstatistics.com). See Ramirez et al. (2012) for a summary of the process used in developing the SATS, as well as for definitions of the six attitude components and an example item from each.

Nolan, Beran, and Hecker (2012) review an abundance of studies that support the factorial structure validity of the SATS-28 and SATS-36 pretest and posttest component scores as used with a variety of student populations. Since their article was published, additional studies continue to confirm six interrelated distinct constructs in the SATS-36 (e.g., Stanisavljevic et al., 2014). All of these studies use item-parceled confirmatory factor analysis (CFA) techniques.

Two recent studies use ordinal CFA techniques. Vanhoof, Kuppens, Sotos, Verschaffel, and Onghena (2011) use these techniques to test the six-factor structure of SATS pretest item responses obtained from students enrolled in a Belgium undergraduate introductory statistics course. Their results suggest possibly merging the *Affect*, *Cognitive Competence*, and *Difficulty* constructs, yielding a four-factor model. They note, however, that the original six-factor model fits slightly better than this four-factor model. In addition, they test both a four-factor and a six-factor model modified primarily by deleting three of the *Difficulty* items with low factor loadings. Both modified models exhibit better fit than the original models, but again differ little in comparison to each other. They examine the three deleted *Difficulty* items and suggest possible reasons for their poor fit.

Persson, Kraus, Hansson, and Wallentin (2019) also use item-level analysis techniques on pretest responses from students enrolled in a Swedish undergraduate introductory statistics course. They conclude that their results support the six-factor structure. Like Vanhoof et al. (2011), they modified their model primarily by excluding three *Difficulty* items, two of which also were identified as problematic by Vanhoof et al.. They conclude that their results do not show improved model fit by combining *Affect*, *Cognitive Competence*, and *Difficulty* into one factor.

With two published studies using ordinal CFA techniques only on pretest item responses from two student populations, more research is needed. Using data collected from a diverse population of introductory statistics students, this study furthers our understanding of the structure of the SATS-36 by using ordinal CFA on both pretest and posttest item responses. The continued exploration of factor structure using item-level CFA results allows identification of potentially problematic items. Parcel-based techniques do not allow this fine grain evaluation.

In addition, whereas the usual CFA models are the standard choice for factorial validation, expanding these models to examine systematic common variance provides even more information to guide survey evaluation. Spector, Rosen, Richardson, Williams, and Johnson (2019) gather and synthesize a series of common factor CFA models where response variance is partitioned into three additive parts: these include variance due to a set of construct factors, a general factor, and random error. Similar variance decomposition models also can be found in the test theory literature as the general factor model (McDonald, 1999). The usual CFA model, appropriately used by Vanhoof et al. (2011) and Persson et al. (2019) to test factor structure, partitions item response variance into two parts: construct factors (that include variance from the SATS component constructs and the hypothesized general factor) and random error.

The general factor includes all sources of variance except those associated with the construct factors and random error. In applied contexts, the general factor is commonly referred to as the common method factor and interpreted as the result of method effects (e.g., Podsakoff, MacKenzie, & Podsakoff, 2012). A measurement process produces method effects when scores from this process assess more than the construct of interest (e.g., Maul, 2013). In self-report measures, including surveys such as the SATS, method effects can be construed as variance resulting from sources including, for example, item characteristics beyond what is due to the attitude components being assessed (Podsakoff, MacKenzie, Lee, & Podsakoff, 2003) and the measurement process itself (e.g., an electronically administered survey). In our data set, we can only examine item characteristics; we do not have a multimethod-multitrait design that would allow us to examine other sources of common variances.

An examination of the SATS-36 items suggests that method effects possibly could be the result, for example, of an item's directionality of wording (positive or negative), specific or general wording, and key word repetition across items. Wording directionality was an issue in the development of the SATS. The focus group that played an integral part in the development of the SATS-28 generated over 90 unique words and phrases they believed expressed their own and other students' attitudes toward statistics. About 80% of these expressed negatively-worded attitudes. They often contained emotionally charged words/phrases (e.g., Item 28: "I am scared by statistics") whereas the positively-worded attitudes tended to be less charged (e.g., Item 3: "I will like statistics."). Even though an approximately equal number of negatively- and positively-worded items were included in the pilot test item set, many more negatively-worded items were retained based on their stronger contribution to construct internal consistency. When the two additional constructs were added to create the SATS-36, the opposite problem occurred; it was difficult to create negatively-worded items. All of the items in both *Interest* and *Effort* contain only positively-worded items.

The second possible source of common variance concerns specific or general wording (e.g., own versus others' attitudes). Most of the items in the constructs other than *Difficulty* clearly ask for the student's own attitudes by using the word "I." The "non-I" items may assess students' stereotypes (e.g., what everyone believes) rather than their own attitudes. These "non-I" stereotypical items initially were placed in the SATS because of the focus group's recommendations and remained in it because they behaved well in the pilot test analyses. Vanhoof et al. (2011) allude to this issue when they suggest that two of the three problematic *Difficulty* items they identify refer to "people" rather than to the student; however, none of the *Difficulty* items explicitly refer to the student through the use of "I."

The third potential source involves repetition of a key word. Each of the four *Interest* items contains the word "interest" because it was difficult to write items that appear to clearly assess students' interest without using that term.

The present study has two purposes. Building on the work of Vanhoof et al. (2011) and Persson et al. (2019), the first purpose is to extend the previous confirmatory work on the original six-factor structure to include SATS posttest, as well as pretest, item-level data. The second and main purpose utilizes the common factor model to quantitatively investigate method effects. This statistical procedure makes it possible to examine the common variance associated with the SATS-36 six-component structure and its items. To our knowledge, this approach has not been used previously with SATS data. We hope this study represents the first step towards a better understanding of the method effects associated with assessing statistics attitudes using the SATS-36.

2. METHODS

2.1. DATA SOURCE

The data used in this study are from students' responses obtained from the SATS Project data set. Survey Monkey, a web-based data collection software program, was used to collect the data across three academic years from the 2007 fall term through the spring term of 2010. Instructors teaching U.S. statistics courses volunteered to ask their students to take the SATS-36. Students responded to the survey during or outside of class within two weeks of the beginning and of the end of their classes. Each year, the SATS Project was approved by a Human Subjects Review Board (See Schau & Emmíoğlu, 2012, for more information).

Specific to this study, we are interested in students who took introductory statistics courses with either no mathematics prerequisite or with an algebra-only prerequisite that were taught in a U.S. mathematics or statistics department (i.e., "service" courses). Applying these criteria to the data set yielded 1865 students who completed at least one pretest item and 1562 students at the posttest. Of these, 1685 students (90%) answered all of the pretest items while 1387 (89%) completed all posttest items. These students were from 62 different course sections taught by 22 instructors in 12 postsecondary institutions. Using the on-line Carnegie Classifications (The Carnegie Classification of Institutions of Higher Education, n.d.), these institutions included one Community College, five Baccalaureate Colleges, four Master's Colleges and Universities, and two Doctoral/Research Universities.

Two common methods used to deal with missing survey data are subject deletion and item imputation. Both methods were applied to this data set and the results compared. Participants who responded to approximately 90% (32) or more of the items were retained for imputation. At pretest, 33 students (2%) were missing 4 or more items while 19 (1%) were identified at posttest by the same criterion. These small amounts of data were omitted from the data sets used for imputation, yielding 1832 students from the pretest and 1543 students from the posttest.

These students' demographics varied little between pretest and posttest. Their median age at both testing times was 19.6 years, with a minimum age of 16.9 and a maximum of 46.4 (2% missing at pretest and also at posttest). About two-thirds of the students self-reported as female (65% pretest, 64% posttest) with about one-third as male (34% pretest, 35% posttest) with 1% omitting this item. The great majority (96%) indicated that they were U.S. citizens at both pretest and posttest; 2% self-classified as foreign at both administration times and 1% as other with 1% leaving this item blank.

Using the VIM package in R (Kowarik & Templ, 2016), missing data for these two groups of students were estimated using hot-deck imputation, a technique commonly used for handling survey non-responses (Andridge & Little, 2010). To check the plausibility of the imputed data, simple summary statistics (i.e., sample means and standard deviations) were compared between the imputed data set and the data set from which participants with any number of missing values were omitted.

2.2. MEASURE

The SATS-36 contains 36 seven-point Likert scale items where 1 means strongly disagree, 4 means neutral/no opinion, and 7 means strongly agree. Each of these items belongs to one of six attitude component subscales: *Affect*, *Cognitive Competence*, *Value*, *Difficulty*, *Interest*, or *Effort*. The pretest and posttest versions contain identical items except for changes in tense, when needed. The responses to negatively-worded items are reversed before scoring. Students receive a mean score on each component. See https://www.evaluationandstatistics.com/.

The SATS-36 also contains additional items designed to assess a variety of other constructs (e.g., students' demographic characteristics, educational backgrounds, global attitudes toward statistics). In general, these constructs are studied less often than the six attitude components.

Ramirez et al. (2012) constructed a broad conceptual model (the SATS-M) relating the six attitude components, as well as two additional constructs assessed by the SATS-36, to statistics course outcomes (e.g., end-of-course attitudes, achievement). They explicitly show the congruence of these constructs to components found in Eccles' Expectancy Value Theory (EVT, e.g., Muenks, Wigfield, & Eccles, 2018; Wigfield & Eccles, 2002). This congruence is interesting because the SATS-28 was developed without a theoretical basis; the two components added to create the SATS-36 were adapted from Eccles' EVT.

2.3. STATISTICAL ANALYSIS

Model identification For both the original six-factor and the common method factor models to be identifiable, one nonzero loading for each factor is fixed to one. In addition, the common method factor must be set to be uncorrelated with the substantive factors. The hypothesized common factor model for the SATS project data is diagrammed in Figure 1. It is created by adding a common method factor to the original six-factor CFA model.

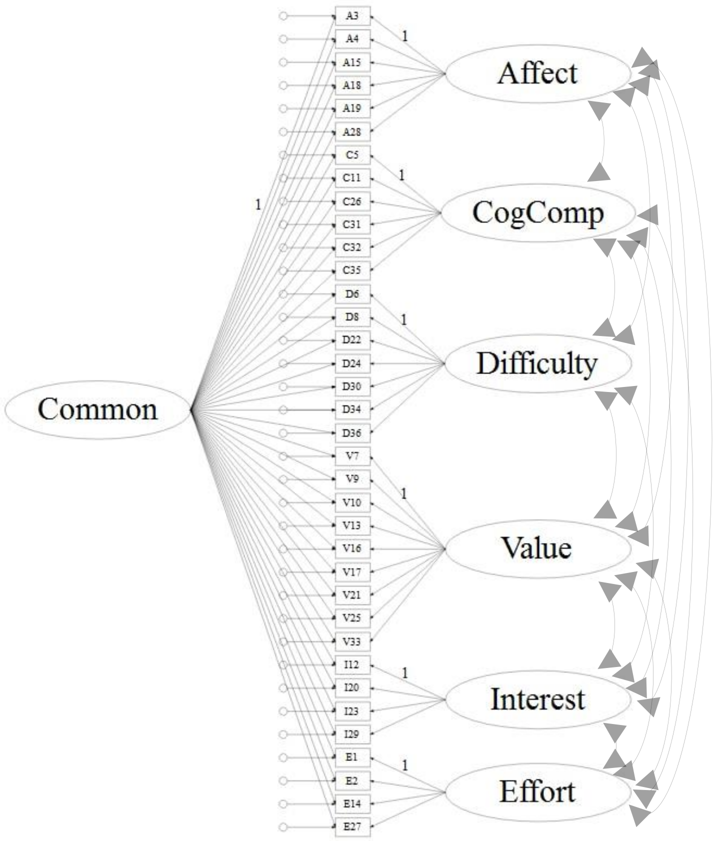

Figure 1. The common factor model applied to the SATS data

Model parameter estimation Because of the Likert-scale nature of the SATS-36 item responses, the data are ordinal in nature and so are fitted to the estimated polychoric correlation matrix (e.g., Yang-Wallentin, Jöreskog, & Luo, 2010). This technique assumes a bivariate normal distribution underlying each pair of ordinal variables. For both the pretest and posttest SATS data, this hypothesis is checked by using the test of close fit (Jöreskog, 2005). Because of the small Root Mean Square Error of Approximation (RMSEA) values for each pair of variables (RMSEA < 0.1), there is no evidence against the null hypothesis so we conclude the assumption was not violated.

The unweighted least squares estimator is used for parameter estimation due to many advantages (Forero, Maydeu-Olivares, & Gallardo-Pujol, 2009; Li, 2016). All models were estimated using the lavaan package in R (Rosseel, 2012).

Model fit evaluation The goodness-of-fit indices used for model fit evaluation include the unscaled chi-square statistic, Comparative Fit Index (CFI), RMSEA, Standardized Root Mean Square Residual (SRMR), and Non-normed Fit Index (NNFI) (e.g., Hu & Bentler, 1999). Chi-square statistics, scaled or not, are known to be very sensitive to large sample sizes and frequently reject good models. A model has a good fit if CFI is 0.95 or greater and RMSEA is 0.05 or less, with values between 0.05 and 0.08 indicating a reasonable fit. For SRMR, a value less than 0.08 is considered a good fit. A value greater than 0.95 is generally considered a good fit for NNFI. Hu and Bentler (1999) provide comprehensive guidelines for the cutoff values of model-fitting indices.

Variance decomposition The quantification of allocated variance relates directly to the decomposition of the polychoric correlation matrix with this form

$$\mathbf{y} = \boldsymbol{\Lambda}^{(1)}\boldsymbol{\eta}_s + \boldsymbol{\Lambda}^{(2)}\boldsymbol{\eta}_c + \boldsymbol{\varepsilon}$$

where $\mathbf{y}$ is a 36×1 vector of the continuous standard normal variables underlying the ordinal observed variables. The 6×1 vector $\boldsymbol{\eta}_s$ and the univariate $\boldsymbol{\eta}_c$ represent substantive and common method factors, respectively. The 36×6 matrix $\boldsymbol{\Lambda}^{(1)}$ is congeneric (i.e., each indicator loads on only one latent theoretical construct) and contains factor loadings for the substantive factors while $\boldsymbol{\Lambda}^{(2)}$ is a 36×1 vector that contains factor loadings for the common method factor. The 36×1 vector $\boldsymbol{\varepsilon}$ represents random errors. The decomposition of variance into three parts can then be carried out in the following manner:

$$\mathbf{I} = \mathrm{diag}(\boldsymbol{\Lambda}^{(1)}\,\boldsymbol{\Phi}\boldsymbol{\Lambda}^{(1)\prime}) + \mathrm{diag}(\boldsymbol{\Lambda}^{(2)}\,\boldsymbol{\varphi}\boldsymbol{\Lambda}^{(2)\prime}) + \boldsymbol{\Theta}$$

where $\mathbf{I}$ is an identity matrix, $\boldsymbol{\Phi}$ is the covariance matrix for $\boldsymbol{\eta}_s$ and $\boldsymbol{\varphi}$ is the variance for $\boldsymbol{\eta}_c$. The following analysis is made straightforward with standardization (Rodriguez, Reise, & Haviland, 2016). Given the congeneric property of $\boldsymbol{\Lambda}^{(1)}$ as well as prior standardization that yields unit variances for all latent factors, the $\mathrm{diag}(\boldsymbol{\Lambda}^{(1)}\boldsymbol{\Lambda}^{(1)\prime})$ and $\mathrm{diag}(\boldsymbol{\Lambda}^{(2)}\boldsymbol{\Lambda}^{(2)\prime})$ matrices have squared loadings for substantive factors and the common method factor on the main diagonal, respectively. $\boldsymbol{\Theta}$ is a matrix with residual variances on its diagonal. As a result, the proportion of variance attributable to the common method factor (i.e., $\mathrm{diag}(\boldsymbol{\Lambda}^{(2)}\boldsymbol{\Lambda}^{(2)\prime})$) can be computed for each of the six substantive factors and for each of the 36 items (Osman et al., 2009; Stucky & Edelen, 2014). The regular six-factor model cannot identify method variance as it specifies only $\boldsymbol{\Lambda\eta}$ and $\boldsymbol{\varepsilon}$.

3. RESULTS

3.1. MISSING DATA

Based on the summary statistics presented in Table 1, hot-deck imputation and deletion of missing values yielded almost identical summary statistics in both the pre- and post-settings. Imputation did not reduce data variability nor cause obvious biases. The imputed data sets were used for the CFAs, resulting in final sample sizes of 1832 and 1543 participants for the pre- and post-SATS data, respectively.

Table 1. Sample means and standard deviations for missing value-imputed and -deleted SATS project data

	Pretest imputed (n = 1832)		Pretest deleted (n = 1685)		Posttest imputed (n = 1543)		Posttest deleted (n = 1387)	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Effort	6.39	0.93	6.39	0.92	5.96	1.06	5.95	1.06
Cognitive Competence	4.85	1.03	4.86	1.03	4.94	1.13	4.95	1.13
Affect	4.17	1.10	4.18	1.10	4.28	1.33	4.28	1.34
Difficulty	3.62	0.77	3.62	0.76	3.75	0.93	3.75	0.94
Value	5.09	0.99	5.10	0.98	4.83	1.11	4.83	1.11
Interest	4.69	1.22	4.69	1.21	4.18	1.42	4.19	1.42

3.2. SIX-FACTOR STRUCTURES FOR PRETEST AND POSTTEST DATA

The first purpose of this study is to extend previous confirmatory results on the original six-factor structure to include SATS posttest, as well as pretest, item data. The model-fitting indices for data from both administration times are shown in Table 2. Neither of the two original models satisfy the empirical cutoff values for good fit although they are reasonably close; the six-factor model is a better fit to the posttest data then to the pretest data.

Table 2. Fitting indices for the models with and without the common method factor

Model	χ^2	*df*	CFI	RMSEA	NNFI	SRMR
Pre-original	11873.959	579	0.903	0.103	0.895	0.099
Pre-common	2386.597	543	0.982	0.043	0.982	0.044
Post-original	7292.534	579	0.948	0.087	0.943	0.084
Post-common	2578.165	543	0.986	0.049	0.982	0.050

All but three of the pretest items meet the 0.4 criterion used by Vanhoof et al. (2011) (see the Pre-original column in Table 3). These items are from the *Difficulty* scale: D22 ("Statistics is a subject quickly learned by most people."), D24 ("Learning statistics requires a great deal of discipline."), D30 ("Statistics involves massive computations."). None of the posttest items failed this criterion although the loadings from these same three *Difficulty* items are relatively weak (see Post-original column in Table 3).

Cognitive Competence and *Affect* show the strongest estimated correlation between pairs of latent factors at both administration times (see Table 4): both pairs correlated over 0.9. Three other pretest factor pairs also correlated strongly (r > 0.7) although not as strongly as *Cognitive Competence* and *Affect*. These factor pairs included *Cognitive Competence* and *Difficulty*, *Affect* and *Difficulty*, and *Interest* and *Value*. These four pairs, as well as *Interest* and *Affect*, were correlated strongly at posttest.

3.3. COMMON METHOD FACTOR AND VARIANCE DECOMPOSITION

The second and main purpose of this study utilizes the general factor model to investigate method effects at the item and factor levels. To examine the common variance associated with the structure of the SATS-36, one common method factor was added to the original six-factor model (see Figure 1). As shown in Table 2, both pretest and posttest data fit the common factor model; unlike the fit of the original model, the fit is equally good at both administration times.

The proportion of common variance at both the individual item and the construct levels is particularly useful. These proportions are informative because they can suggest individual items and factors that need more evaluation. Three different types of variances were estimated and analyzed through variance decomposition (see Table 5). All of these variances are calculated from the estimated factor loadings from the common factor models (Table 3). For example, the substantive variance for the first item E1 for the pretest method model (0.723) is obtained by squaring its factor loading (0.850)

Table 3. Factor loading estimates from models with or without the Common factor

Item	Pre-original	Pre-common		Post-original	Post-common	
		Substantive factors	Common factor		Substantive factors	Common factor
E1	0.878	0.850	0.355	0.842	0.630	0.434
E2	0.868	0.722	0.473	0.871	0.813	0.381
E14	0.771	0.578	0.521	0.552	0.727	0.190
E27	0.757	0.644	0.409	0.595	0.579	0.228
C5	0.647	0.618	-0.356	0.632	0.700	0.173
C11	0.713	0.697	-0.145	0.814	0.601	0.497
C26	0.592	0.564	-0.333	0.586	0.625	0.178
C31	0.740	0.758	0.170	0.759	0.515	0.506
C32	0.727	0.725	0.026	0.765	0.590	0.440
C35	0.754	0.721	-0.476	0.844	0.819	0.337
A3	0.659	0.696	0.224	0.774	0.318	0.756
A4	0.590	0.559	-0.354	0.607	0.690	0.192
A15	0.659	0.639	-0.234	0.695	0.657	0.334
A18	0.615	0.577	-0.451	0.735	0.749	0.308
A19	0.724	0.763	0.228	0.795	0.325	0.779
A28	0.726	0.688	-0.481	0.779	0.796	0.326
D6	0.801	0.735	-0.142	0.825	0.567	0.400
D8	0.698	0.496	-0.484	0.756	0.772	0.102
D22	**0.386**	0.309	-0.158	0.414	0.322	0.166
D24	**0.304**	0.035	-0.650	0.423	0.676	-0.209
D30	**0.366**	0.181	-0.468	0.438	0.577	-0.076
D34	0.438	0.256	-0.456	0.522	0.698	0.010
D36	0.491	0.334	-0.408	0.578	0.602	0.079
V7	0.715	0.698	0.152	0.765	0.506	0.554
V9	0.653	0.622	0.270	0.723	0.371	0.597
V10	0.660	0.616	0.363	0.704	0.297	0.629
V13	0.742	0.725	0.155	0.702	0.530	0.468
V16	0.789	0.796	0.031	0.823	0.659	0.524
V17	0.583	0.584	0.063	0.644	0.412	0.479
V21	0.592	0.595	0.022	0.651	0.619	0.350
V25	0.744	0.733	0.121	0.788	0.661	0.481
V33	0.740	0.738	0.059	0.764	0.644	0.464
I12	0.699	0.619	0.361	0.767	0.325	0.695
I20	0.898	0.825	0.295	0.932	0.485	0.800
I23	0.864	0.777	0.371	0.887	0.355	0.822
I29	0.884	0.792	0.437	0.900	0.266	0.874

Note: Factor loadings less than 0.4 are bold.

associated with the *Effort* construct while the common variance (0.126) is obtained by squaring the common method factor loading (0.355). The total variance of E1 is obtained by summing the substantive and common variances; it is the variance that is not due to random error.

The desired pattern in variance decomposition for both items and constructs includes a large total variance that is constructed of a large substantive variance and a small common variance. Table 5 shows the variance decomposition at the individual item level.

As Table 5 shows, the partitioning of item variance at pretest yields estimated proportions of substantive variance that range from 0.001 (D24) to 0.723 (E1) with a median proportion of 0.444. Only five pretest items operate poorly, all in the *Difficulty* construct. The total variances for D24 ("Learning statistics requires a great deal of discipline."), D30 ("Statistics involves massive computations."), D34 ("Statistics is highly technical.") and D36 ("Most people have to learn a new way of thinking to do sta-

Table 4. Latent factor correlation estimates for the six-factor models

Pre-original	*Effort*	*Cog.Comp*	*Affect*	*Difficulty*	*Value*
Cog.Comp	0.259				
Affect	0.132	0.942			
Difficulty	-0.225	0.717	0.759		
Value	0.379	0.573	0.504	0.196	
Interest	0.384	0.531	0.613	0.179	0.728

Post-original	*Effort*	*Cog.Comp*	*Affect*	*Difficulty*	*Value*
Cog.Comp	0.280				
Affect	0.213	0.936			
Difficulty	-0.101	0.732	0.737		
Value	0.278	0.620	0.652	0.323	
Interest	0.253	0.555	0.707	0.299	0.798

Note: All correlation coefficient estimates are statistically significant.

tistics.") are small and mostly are accounted for by the common method factor, not the substantive factor. With a small total variance at pretest, Item D22 ("Statistics is a subject quickly learned by most people.") also operates poorly, consisting primarily of random error.

Table 6 presents the variance decomposition results at the construct level. Total variances at pretest range from 0.347 (*Difficulty*) to 0.710 (*Interest*). The estimated proportion of substantive variance at pretest ranges from 0.159 (*Difficulty*) to 0.570 (*Interest*). Although the rest of the constructs function well, over half of the total variance in *Difficulty* is associated with the common method factor; the great majority of the total variances in the other five factors is substantive. *Difficulty* as a construct operates poorly at pretest.

As Table 5 shows, the estimated proportions of substantive variance at posttest range from 0.071 (I29) to 0.671 (C35) with a median proportion of 0.359. Item D22 still shows a small total variance as it did at pretest. Now, however, each of the other *Difficulty* items show larger substantive than common variances, although their total variances are not large.

At pretest, the four *Interest* items exhibit large total and substantive variances (and so small common variances). At posttest, the *Interest* construct still exhibits a large amount of total variance, but now most of it is common variance (over 80%). All of the posttest *Interest* items are affected by the common method factor: I12 ("I am interested in being able to communicate statistical information to others."), I20 ("I am interested in using statistics).", I23 ("I am interested in understanding statistical information."), I29 ("I am interested in learning statistics."). A large amount of each item's large total variance is attributed to the common method factor.

At posttest, four of the nine *Value* items (V7, V9, V10, V17) show larger common variances than substantive variances. V7 ("Statistics is worthless.") is a negatively-worded item that does not specifically ask about the student's own attitudes; that is, it does not contain the word "I." The other three are positively-worded, with two asking about the use of statistics in professional life: V9 ("Statistics should be a required part of my professional training."), V10 ("Statistical skills will make me more employable."), and one asking about personal life, V17 ("I use statistics in my everyday life."). Additionally, two of the six *Affect* items are affected by method effects: A3 ("I will like statistics.") and A19 ("I will enjoy taking statistics courses."); these are the only two positively-worded items in this construct.

As Table 6 shows, at posttest, total construct variances range from 0.400 (*Difficulty*) to 0.778 (*Interest*). As occurred at pretest, total variances are smallest for the *Difficulty* construct and largest for the *Interest* construct. The estimated proportion of substantive variance at posttest ranges from 0.131 (*Interest*) to 0.480 (*Effort*). As expected from the item-level results, the pretest problem with *Difficulty*, where over 50% of the total variance is associated with method effects, disappears at posttest with only 10% due to method. Again, mirroring the item-level analysis, over 80% of the posttest total variance in *Interest* is accounted for by the common method factor while only 25% at pretest is associated with method. In addition, the *Value* construct exhibits a possible common variance issue at posttest (almost 50%) but not at pretest (about 20%).

Table 5. Variance decomposition (proportion) at the individual item level

Items	Pre-method			Post-method		
	Total Variance	Substantive Variance	Method Variance	Total Variance	Substantive Variance	Method Variance
E1	0.849	0.723	0.126	0.586	0.397	0.189
E2	0.745	0.522	0.223	0.806	0.660	0.146
E14	0.606	0.334	0.272	0.565	0.529	0.036
E27	0.582	0.414	0.168	0.387	0.335	0.052
C5	0.509	0.382	0.127	0.520	0.490	0.030
C11	0.508	0.486	0.022	0.607	0.361	0.246
C26	0.429	0.318	0.111	0.423	0.391	0.032
C31	0.603	0.574	0.029	0.522	0.266	0.256
C32	0.527	0.526	0.001	0.542	0.349	0.193
C35	0.747	0.520	0.227	0.784	0.671	0.113
A3	0.535	0.485	0.050	**0.674**	**0.101**	**0.573**
A4	0.438	0.313	0.125	0.512	0.476	0.036
A15	0.463	0.408	0.055	0.543	0.432	0.111
A18	0.536	0.333	0.203	0.656	0.562	0.094
A19	0.634	0.582	0.052	**0.713**	**0.016**	**0.607**
A28	0.705	0.474	0.231	0.741	0.634	0.107
D6	0.561	0.541	0.020	0.482	0.322	0.160
D8	0.480	0.250	0.230	0.607	0.597	0.010
D22	**0.120**	0.095	0.025	**0.131**	0.104	0.027
D24	**0.424**	**0.001**	**0.423**	0.501	0.457	0.044
D30	**0.252**	**0.033**	**0.219**	0.339	0.333	0.006
D34	**0.274**	**0.066**	**0.208**	0.357	0.356	0.001
D36	**0.278**	**0.112**	**0.166**	0.368	0.362	0.006
V7	0.510	0.487	0.023	**0.564**	**0.256**	**0.307**
V9	0.459	0.387	0.072	**0.495**	**0.138**	**0.357**
V10	0.511	0.379	0.132	**0.484**	**0.088**	**0.395**
V13	0.550	0.526	0.024	0.499	0.280	0.219
V16	0.635	0.634	0.001	0.708	0.434	0.274
V17	0.345	0.341	0.004	**0.399**	**0.170**	**0.229**
V21	0.354	0.353	0.001	0.506	0.383	0.123
V25	0.552	0.537	0.015	0.669	0.437	0.231
V33	0.549	0.546	0.003	0.629	0.414	0.215
I12	0.513	0.384	0.130	**0.589**	**0.106**	**0.483**
I20	0.767	0.680	0.087	**0.876**	**0.236**	**0.640**
I23	0.741	0.603	0.137	**0.802**	**0.126**	**0.676**
I29	0.819	0.628	0.191	**0.835**	**0.071**	**0.764**

Note: Bolded values indicate items with a larger proportion of method than of substantive variance or with a small proportion of total variance (D22 only).

Table 6. Variance decomposition at the individual construct level

Constructs	Pre-common			Post-common		
	Total Variance	Substantive Variance	Common variance	Total Variance	Substantive Variance	Common variance
Effort	0.700	0.500	0.200	0.590	0.480	0.110
CogComp	0.554	0.465	0.089	0.570	0.420	0.150
Affect	0.550	0.430	0.120	0.640	0.390	0.250
Difficulty	0.347	0.159	0.188	0.400	0.360	0.040
Value	0.500	0.470	0.030	0.550	0.290	0.260
Interest	0.710	0.570	0.140	0.778	0.131	0.646

4. DISCUSSION, LIMITATIONS, AND FUTURE DIRECTIONS

We discuss our findings in four areas. The first explores the potential impact of our sample on the findings. The second examines what missing data suggest about SATS item quality. The third area considers findings from the CFAs that evaluate the original six-factor structure for SATS pretest and posttest data. The fourth examines the impact and meaning of common variance on constructs and items at both pretest and posttest.

4.1. SAMPLE

Our data set has many strengths. It is large and includes a varied group of students from diverse colleges and universities. Class sizes range from small to large. This diversity allows broad generalization but also potentially can cause issues. For example, some of our data are multilevel while some are not. That is, we have students from smaller stand-alone classes, as well as students from larger lecture sections with accompanying smaller laboratory sections. We have instructors who participated only once and others who participated several times across the three years of data collection. The instructional methods varied across the classes. These differences may be one additional cause of the common variances we found. Future research is needed to explore this possibility.

This sample consists of students enrolled in U.S. classes taught by instructors who volunteered to participate. Thus, it is likely that these classes either were taught or supervised by instructors who value good instruction, want to evaluate and improve their own teaching, and believe that students' attitudes are important course outcomes. As is true of the students in both the Vanhoof et al. (2011) and Persson et al. (2019) studies, these students volunteered. To the best of our knowledge, there is no information available that can be used to evaluate the representativeness of this sample. The volunteer aspect of most surveys is especially problematic when statistical significance tests are used to evaluate the impact of the results. That is not the case in our study. Research with other samples, both within the United States and in other countries, is needed to see whether the Common Factor Model results replicate.

4.2. MISSING DATA

An excessive amount of missing responses to an item may mean that the item is of poor quality. In this study, the highest non-response rate to any item at pretest or posttest is low (< 1%), indicating that students likely believed they understood the items well enough to respond to them. This finding is important for both researchers and instructors using the SATS. To obtain useful component scores, complete or mostly complete sets of responses are needed.

4.3. SATS PRETEST AND POSTTEST ORIGINAL SIX-FACTOR STRUCTURES

The first purpose of this study is to evaluate the fit of the SATS original six-factor model using pretest and posttest data. As was the case for both Vanhoof et al. (2011) and Persson et al. (2019), the pretest data do not fit the original six-factor model as well as desired, although the values of the fit indices often are close to the desired criteria values. Thus, the six-factor model is not completely adequate in accounting for pretest item variability.

The estimated factor loadings of the original six-factor model at pretest share some similar patterns with those from Vanhoof et al. (2011) and Persson et al. (2019). Some factor loadings, however, show discrepancies across the three studies. In the three studies, all but three pretest items exhibit loadings that were greater than 0.4, the criterion used in both previous studies. Only the loading for D22, however, was below this value in all three studies. The loadings for items D30 and D34 fall below the cutoff in two of the studies, while those for D24 and D36 fail to meet the criterion in only one study.

In this study, the original six-factor model fits better at posttest than at pretest, although it still does not meet the criteria for good fit. None of the posttest item factor loadings, including those for all of the *Difficulty* items, fall below the 0.4 cutoff criterion.

These differences across student samples and across administration times indicate that caution must be exercised when considering item revision or deletion. Item modification or omission is needed when there is a ubiquitous issue regardless of, for example, population characteristics and time of administration. The differing results across studies do not show a consistent pattern supporting item elimination. They do suggest that the *Difficulty* items may need further scrutiny.

Like Vanhoof et al. (2011) and Persson et al. (2019), *Cognitive Competence* and *Affect* exhibit a very strong pretest correlation, and three other pretest item pairs also correlate strongly (*Cognitive Competence* and *Difficulty, Affect* and *Difficulty,* and *Interest* and *Value*). In addition, *Interest* and *Affect* were strongly correlated at the posttest administration.

These correlational patterns are similar to those found in most studies using SATS pretest or posttest data from a wide variety of student groups. Vanhoof et al. (2011) suggest possibly merging some of the strongly-related constructs (especially *Affect, Cognitive Competence,* and *Difficulty),* even though each construct is theoretically unique. Another possibility is that there may be a hierarchical structure associated with some of the constructs.

This possible hierarchical nature needs further exploration. Specifically, do *Affect* and *Cognitive Competence* (with or without *Difficulty*) form a super construct that adds additional information to that provided by these constructs separately? Does *Value* already act as a super construct because it contains items that generally sort into two categories: the value of statistics in professional life and its value in personal life?

In addition, research exploring these questions has direct bearing on the SATS-M and other uses of the EVT in statistics education. EVT includes four components under the super construct of Subjective Task Value (e.g., Muenks et al., 2018; Wigfield & Eccles, 2002). The SATS-M, and the SATS as it currently is used, do not explicitly include super components. The idea of attitude super constructs in statistics education needs consideration and expansion.

The answers to these questions have important educational implications. It may well be the case that different instructional approaches are needed to improve students' attitudes in these two possible *Value* subconstructs. Similarly, it does not seem likely that one specific type of intervention would improve *Affect*, *Cognitive Competence*, and *Difficulty*, although that again is an empirical question.

Another related area of importance for instruction, research, and theory involves expanded exploration of the potentially causal direction of the relationships among the SATS components (e.g., Paul & Cunnington, 2017) as is found among the EVT components. Future work in this area should help instructors better understand the developmental flow of students' attitudes toward statistics, as well as provide a framework for evaluating the effects of interventions designed to improve attitudes.

4.4. VARIANCE DECOMPOSITION

The second purpose of this study uses the general factor model to investigate method effects. When a common method factor is added to the original six-factor model, the new model now fits at both the pretest and posttest. That is, a common method factor accounts for an additional amount of item variability that is not accounted for by the substantive constructs alone. Our results corroborate many previous CFA investigations (e.g., Gignac, 2006; Keeping & Levy, 2000; Lahey et al., 2018), which find that common factor models typically fit better than do regular factor models. Decomposition of variance assists in identifying individual items and constructs that need further exploration.

In the presence of a common method factor, the pretest version of the SATS-36 exhibits good item and construct quality except for *Difficulty* and some of its items. Among the items, D22 (a positively-worded item) accounts for a small amount of total variance while D24, D30, D34, and D36 (all negatively-worded items) at pretest only are affected by the common method factor (i.e., larger common variances and smaller substantive variances). All *Difficulty* items are general items; that is, they do not ask explicitly for the students' attitudes about themselves.

In general, the SATS-36 posttest version also exhibits reasonable item and construct quality in the presence of the common method factor. The *Difficulty* construct continues to account for a relatively small amount of total variance, although most of it now is substantive. D22 shows the same pattern found at pretest: small total variance with most of it substantive. The variances from the other problematic *Difficulty* pretest items now consist mostly of substantive, not method, variance, although they are not strong items. Again, these results suggest that the *Difficulty* construct and its items may need additional scrutiny.

The *Interest* construct shows the largest total variance of any posttest construct, but about half of it is associated with method effects. Each item in the construct, all positively-worded and containing the word "interest," also shows larger method than substantive variances. This pattern is the opposite of that found with the pretest data where variances associated with the *Interest* construct and all of its items are mostly substantive.

The *Value* construct shows a reasonable total variance where about half of it is common variance. About half of its items exhibit larger method than substantive variances. Three of these are positively-worded while the fourth is a general negatively-worded item.

The *Affect* construct operates quite well. Its two problematic items, however, are the only two positively-worded items in the construct. Both Vanhoof et al. (2011) and Persson et al. (2019) allow these items to covary in their models.

These differential findings at pretest and posttest indicate that method effects are item- and construct-specific as well as time-dependent. They also are related differentially to item word-directionality (positive or negative), general or specific wording, and repetition of key words. At pretest, most of the potentially problematic items are negatively-worded items that do not explicitly refer to the student's own attitudes ("non-I" items); they are found in the *Difficulty* construct. At posttest, most of the problematic items are found in two constructs (*Interest* and *Value*). Most are positively-worded and all are explicitly specific to the student ("I" items).

The principal advantage of the common factor model is that this model allows researchers to determine the degree to which method variance is associated with a measure. This flexibility is particularly useful as a feature of an investigative tool. Like any other statistical model, however, there are also some limitations inherent in the common factor model. First, this model is restrictive in the sense that it does not allow the researchers to identify empirically the individual sources of method effects. Second, the model specifies an independent relationship, which is required for model identifiability, between the common method factor and each of the six substantive factors. This assumption may not be tenable in real-life situations. Lastly, studies employing simulation have shown that an addition of the common method factor may lead to biased estimation of correlations among substantive factors (Spector et al., 2019). As a result, caution must be exercised when estimating factor correlations is the main purpose of a study where the common factor model is used.

To ameliorate the foregoing problems, procedural remedies are needed at the stage of data collection. A multimethod-multitrait design or the CFA marker technique (see Williams, Hartman, & Cavazotte, 2010) is needed in order to more fully explore these method effects. It is unlikely that a data set with such a design exists, and it isn't yet clear how to create such a design or a marker variable in terms of students' attitudes toward statistics.

A better understanding of mean and variance changes in academic attitudes is important to effective instruction and to instructional innovation (Blazar & Kraft, 2017; Schau & Emmíoğlu, 2012). This understanding may eventually lead to targeted interventions designed to improve attitudes (Cohen, Garcia, & Goyer, 2017). Kerby and Wroughton (2017) report that temporal changes in statistics attitudes do not occur monotonically within individuals; in fact, the direction, and even the existence, of attitude changes differ across students. Table 2, as well as a great deal of other research, shows that introductory statistics students' mean attitudes, on average, either change little from pretest to posttest or, unfortunately, decrease. However, the causes of attitude change not only operate on central tendency but also on variability (Maul, 2013); in fact, Table 1 shows that the variance of every attitude component increased from pretest to posttest. Our observation of temporal changes in common variances along with increases in component variances, coupled with Kerby and Wroughton's finding of differences in attitude change trajectories as courses progress, may indicate important developmental processes impacting students' attitudes and so underlying the way in which students respond to the SATS items at different points in the semester. This hypothesis should be explored in future research.

4.5. CONCLUSIONS

Research is needed that explores models containing factors that may influence students' attitudes toward statistics. Ramirez et al. (2012) report research findings examining various student characteristics and previous achievement-related experiences, two global constructs from their SATS-M. Data sets such as the one used in this research, however, provide a broader, richer set of variables that can be used to explore these factors as part of models based on the wide variety of motivational theories that exist, including Eccles' EVT. Additional research examining causal models interrelating the components of the SATS-M model is needed. Continued development and evaluation of instructional interventions designed to improve students' attitudes toward statistics is needed, as is research into possible developmental processes affecting students' attitudes. All of these research areas require a measure of students' attitudes toward statistics that exhibits good psychometric properties and that is well understood. This research study advances our understanding of the SATS-36.

Both Vanhoof et al. (2011) and Persson et al. (2019) indicate that three of the pretest *Difficulty* items should be omitted due to low factor loadings. These problematic items differ across the three studies. In addition, these items do not exhibit loadings that fall below the criterion in our posttest results. Similarly, with the exception of strong correlations, there is little evidence to recommend combining the theoretically distinct constructs of *Affect, Cognitive Competence,* and *Difficulty* into one construct. These differences and lack of evidence make item deletion and construct combination problematical; they require future research before making these decisions.

The SATS has an extensive history of psychometric evaluation across varying student populations and courses. Even so, this study is the first to examine the possible impact of common variance on the structure of the SATS. To the best of our knowledge, the common variance model has not been used to evaluate method effects in any other attitude survey. Thus, even with the issues identified in this and previous studies, the SATS-36 still remains the best survey for assessing students' attitudes toward statistics, as Nolan et al. (2012) concluded.

ACKNOWLEDGEMENTS

We are grateful to the instructors and students who participated in the SATS Project, as well as to Marjorie Bond for her work in making the Project a reality.

REFERENCES

Andridge, R. R., & Little, R. J. (2010). A review of hot deck imputation for survey non-response. *International Statistical Review, 78*(1), 40–64.

Blazar, D., & Kraft, M. A. (2017). Teacher and teaching effects on students' attitudes and behaviors. *Educational Evaluation and Policy Analysis, 39*(1), 146–170.

Carlson, K. A., & Winquist, J. R. (2011). Evaluating an active learning approach to teaching introductory statistics: A classroom workbook approach. *Journal of Statistics Education, 19*(1). [Online: https://doi.org/10.1080/10691898.2011.11889596]

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

Cohen, G. L., Garcia, J., & Goyer, J. P. (2017). Turning point: Targeted, tailored, and timely psychological intervention. In A. J. Elliot, C. S. Dweck, & D. S. Yeager (Eds.), *Handbook of competence and motivation: Theory and application* (pp. 657–686). New York: Guilford Press.

Forero, C. G., Maydeu-Olivares, A., & Gallardo-Pujol, D. (2009). Factor analysis with ordinal indicators: A Monte Carlo study comparing DWLS and ULS estimation. *Structural Equation Modeling, 16*(4), 625–641.

Garfield, J., Hogg, B., Schau, C., & Whittinghill, D. (2002). First courses in statistical science: The status of educational reform efforts. *Journal of Statistics Education*, *10*(2). [Online: https://doi.org/10.1080/10691898.2002.11910665]

Gignac, G. E. (2006). A confirmatory examination of the factor structure of the Multidimensional Aptitude Battery: Contrasting oblique, higher order, and nested factor models. *Educational and Psychological Measurement, 66*(1), 136–145.

Hood, M., Creed, P. A., & Neumann, D. L. (2012). Using the expectancy value model of motivation to understand the relationship between student attitudes and achievement in statistics. *Statistics Education Research Journal, 11*(2), 72–85. [Online: https://iase-web.org/documents/SERJ/SERJ11(2)_Hood.pdf]

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1–55.

Jöreskog, K. G. (2005). *Structural equation modeling with ordinal variables using LISREL*. [Online: http://www.ssicentral.com/lisrel/techdocs/ordinal.pdf]

Keeping, L. M., & Levy, P. E. (2000). Performance appraisal reactions: Measurement, modeling, and method bias. *Journal of Applied Psychology, 85*(5), 708–723.

Kerby, A. T., & Wroughton, J. R. (2017). When do students' attitudes change? Investigating student attitudes at midterm. *Statistics Education Research Journal, 16*(2), 476–486.
[Online: https://iase-web.org/documents/SERJ/SERJ16(2)_Kerby.pdf]

Kowarik, A., & Templ, M. (2016). Imputation with the R Package VIM. *Journal of Statistical Software, 74*(7), 1–16. [Online: https://doi.org/10.18637/jss.v074.i07]

Lahey, B. B., Zald, D. H., Perkins, S. F., Villalta-Gil, V., Werts, K. B., Van Hulle, C. A., . . . Poore, H. E. (2018). Measuring the hierarchical general factor model of psychopathology in young adults. *International Journal of Methods in Psychiatric Research, 27*(1), e1593.
[Online: https://doi.org/10.1002/mpr.1593]

Li, C.-H. (2016). Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior Research Methods, 48*(3), 936–949.

Maul, A. (2013). Method effects and the meaning of measurement. *Frontiers in Psychology, 4*, 169.
[Online: https://doi.org/10.3389/fpsyg.2013.00169]

McDonald, R. P. (1999). *Test theory: A unified treatment*. Mahwah, NJ: Lawrence Erlbaum Associates.

Muenks, K., Wigfield, A., & Eccles, J. S. (2018). I can do this! The development and calibration of children's expectations for success and competence beliefs. *Developmental Review, 48*, 24–39.

Nolan, M. M., Beran, T., & Hecker, K. G. (2012). Surveys sssessing students' attitudes toward statistics: A systematic review of validity and reliability. *Statistics Education Research Journal, 11*(2), 103–123.
[Online: https://iase-web.org/documents/SERJ/SERJ11(2)_Nolan.pdf]

Osman, A., Williams, J. E., Espenschade, K., Gutiéhorrez, P. M., Bailey, J. R., & Chowdhry, O. (2009). Further evidence of the reliability and validity of the multidimensional anxiety. *Journal of Psychopathology and Behavioral Assessment, 31*(3), 202–214.

Paul, W., & Cunnington, R. C. (2017). An exploration of student attitudes and satisfaction in a GAISE-influenced introductory statistics course. *Statistics Education Research Journal, 16*(2), 487–510.
[Online: https://iase-web.org/documents/SERJ/SERJ16(2)_Paul.pdf]

Persson, I., Kraus, K., Hansson, L., & Wallentin, F. (2019). Confirming the structure of the Survey of Attitudes Toward Statistics (SATS-36) by Swedish students. *Statistics Education Research Journal, 18*(1), 83–93.
[Online: https://iase-web.org/documents/SERJ/SERJ18(1)_Persson.pdf]

Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903.

Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. *Annual Review of Psychology, 63*, 539–569.

Ramirez, C., Schau, C., & Emmíoğlu, E. (2012). The importance of attitudes in statistics education. *Statistics Education Research Journal, 11*(2), 57–71.
[Online: https://iase-web.org/documents/SERJ/SERJ11(2)_Ramirez.pdf]

Rodriguez, A., Reise, S. P., & Haviland, M. G. (2016). Evaluating bifactor models: Calculating and interpreting statistical indices. *Psychological Methods, 21*(2), 137–150.

Rosseel, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software, 48*(2).
[Online: https://doi.org/10.18637/jss.v048.i02]

Schau, C. (2003, August). *Students' attitudes: The* "other" *important outcome in statistics education.* Paper presented at the Joint Statistical Meetings, San Francisco, CA.
[Online: http://www.statlit.org/pdf/2003SchauASA.pdf]

Schau, C., & Emmíoğlu, E. (2012). Do introductory statistics courses in the United States improve students' attitudes? *Statistics Education Research Journal, 11*(2), 86–94.
[Online: https://iase-web.org/documents/SERJ/SERJ11(2)_Schau.pdf]

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.

Spector, P. E., Rosen, C. C., Richardson, H. A., Williams, L. J., & Johnson, R. E. (2019). A new perspective on method variance: A measure-centric approach. *Journal of Management*, *45*(3), 855–880.

Stanisavljevic, D., Trajkovic, G., Marinkovic, J., Bukumiric, Z., Cirkovic, A., & Milic, N. (2014). Assessing attitudes towards statistics among medical students: Psychometric properties of the Serbian version of the Survey of Attitudes Towards Statistics (SATS). *PloS one*, *9*(11). [Online: https://doi.org/10.1371/journal.pone.0112567]

Stucky, B. D., & Edelen, M. O. (2014). Using hierarchical IRT models to create unidimensional measures from multidimensional data. In S. P. Reise & D. A. Revicki (Eds.), *Handbook of item response theory modeling: Applications to typical performance assessment* (pp. 183–206). New York: Routledge.

Tempelaar, D. T., Schim van der Loeff, S., & Gijselaers, W. H. (2007). A structural equation model analyzing the relationship of students' attitudes toward statistics, prior reasoning abilities and course performance. *Statistics Education Research Journal*, *6*(2), 78–102. [Online: https://iase-web.org/documents/SERJ/SERJ6(2)_Tempelaar.pdf]

The Carnegie Classification of Institutions of Higher Education (n.d.). [Online: http://carnegieclassifications.iu.edu/downloads/cc2010_classification_data_file.xls]

Vanhoof, S., Kuppens, S., Sotos, A. E. C., Verschaffel, L., & Onghena, P. (2011). Measuring statistics attitudes: Structure of the Survey of Attitudes Toward Statistics (SATS-36). *Statistics Education Research Journal*, *10*(1), 35–52. [Online: https://iase-web.org/documents/SERJ/SERJ10(1)_Vanhoof.pdf]

Wigfield, A., & Eccles, J. S. (2002). The development of competence beliefs, expectancies for success, and achievement values from childhood through adolescence. In A. W. J. S. Eccles (Ed.), *Development of achievement motivation* (pp. 91–120). San Diego, CA: Academic Press.

Williams, L. J., Hartman, N., & Cavazotte, F. (2010). Method variance and marker variables: A review and comprehensive CFA marker technique. *Organizational Research Methods, 13*(3), 477–514.

Yang-Wallentin, F., Jöreskog, K. G., & Luo, H. (2010). Confirmatory factor analysis of ordinal variables with misspecified models. *Structural Equation Modeling, 17*(3), 392–423.

CANDACE SCHAU
CS Consultants, LLC
12812 Hugh Graham Road NE
Albuquerque, New Mexico 87111