

SURVEYS ASSESSING STUDENTS' ATTITUDES TOWARD STATISTICS: A SYSTEMATIC REVIEW OF VALIDITY AND RELIABILITY

MEAGHAN M. NOLAN
University of Calgary - Canada
mmnolan@ucalgary.ca

TANYA BERAN
University of Calgary - Canada
tnaberan@ucalgary.ca

KENT G. HECKER
University of Calgary - Canada
kghecker@ucalgary.ca

ABSTRACT

Students with positive attitudes toward statistics are likely to show strong academic performance in statistics courses. Multiple surveys measuring students' attitudes toward statistics exist; however, a comparison of the validity and reliability of interpretations based on their scores is needed. A systematic review of relevant electronic databases yielded 532 citations, 78 of which were reviewed, and 35 included in a final analysis. Fifteen instruments were identified; however, evidence of validity and reliability has only accumulated for the Statistics Attitude Scale, Attitudes Toward Statistics Scale, and Survey of Attitudes Toward Statistics (two versions). In conclusion, a number of surveys exist, but there is a paucity of peer-reviewed validity and reliability evidence.

Keywords: *Statistics education research; Statistics attitudes*

1. INTRODUCTION

1.1. MEASURING STUDENTS' ATTITUDES TOWARD STATISTICS

Students' attitudes toward statistics have long been identified as a "special problem" for statistics educators (Bendig & Hughes, 1954). Negative attitudes are perceived to be widespread and likely to decrease students' academic performance in statistics courses, prevent them from acquiring statistical thinking skills, and leave them uncertain about solving statistical problems in the 'real world' (Gal & Ginsburg, 1994; Gal, Ginsburg, & Schau, 1997; Garfield, 1995). As a result, there has been much interest in assessing students' attitudes to predict academic performance in statistics courses and monitor attitudinal changes resulting from educational practices.

The construct of attitudes has been broadly defined as "not directly observable, inferred aspects, consisting of beliefs, feelings, and behavioural predispositions towards the object to which they are directed." (Auzmendi, 1992, p. 17, cited in Mondejar-Jimenez & Vargas-Vargas, 2010). The development of surveys measuring students' attitudes toward statistics began in the 1950s (Bendig & Hughes, 1954), and many instruments are now available. Although each of these surveys claims to measure students' attitudes toward statistics, the dimensionality, items, and results vary among surveys, suggesting that this construct is not yet clearly defined. Currently, a summary and comparison of the validity and reliability evidence for these various interpretations is absent from the literature, making it difficult for statistics' educators and researchers to make evidence-based decisions when selecting a survey or deciding where additional research and development are needed.

1.2. EVIDENCE OF VALIDITY AND RELIABILITY

Messick (1989) defines construct validity as "...an overall evaluative judgment of the degree to which empirical evidence and theoretical rationales support the adequacy and appropriateness of interpretations and actions on the basis of test scores or other modes of assessment" (p. 13). Clearly, construct validity must be established before the responses to any survey of students' attitudes toward statistics can be said to represent students' latent attitudes. The process of validation can involve the accumulation of six different forms of validity evidence: content, substantive, structural, generalizability, external, and consequential. However, inclusion of all six types of evidence is not absolutely required as long as a "compelling argument" for validity can still be made (Messick, 1995). With limited evidence of consequential and generalizability validity available to date, the present study addresses the aspects of content, substantive, structural, and external validity.

Content validity refers to whether or not the survey's items adequately sample the content domain of the construct being assessed without content contamination from other construct domains. This type of validity is indicated by the rigor of the process used to develop the survey's items. Substantive validity considers the strength of the theoretical basis for interpreting survey scores. Structural validity examines whether the intended dimensionality of the construct interpretation is reflected in the survey's scale, subscales, and items. Finally, external validity refers to the comparison of survey scores to external measures, revealing convergent, discriminant, or predictive relationships. Patterns of reasonably strong relationships between attitude scores and other measures of the same attitudes indicate convergent validity. Weak relationships with measures that do not assess these same attitudes indicate discriminant validity. Patterns of reasonably strong relationships with measures theorized to be criterion variables (e.g., student achievement scores) provide evidence of predictive validity (Messick, 1995).

Survey scores cannot be valid if they are not reliable, and similar to validity, different forms of reliability evidence exist. Internal reliability, also known as internal consistency, is commonly reported and is the degree of consistency among responses to items on a single survey that are intended to measure the same dimension. Internal consistency is often reported as Cronbach's α (Cronbach, 1951). External reliability, or stability, is the consistency of scores between administration times (test-retest reliability) and/or raters (inter-rater reliability). The external reliability of surveys of students' attitudes toward statistics has not been widely reported, however; therefore, only internal consistency evidence was summarized in this study.

1.3. PURPOSE

The purpose of this study was to identify all peer-reviewed and non-peer-reviewed surveys developed to assess students' attitudes toward statistics, and to systematically review the evidence of the construct validity and internal consistency of their scores. To our knowledge, such a study has not been conducted, despite the fact that such information would be of value to statistics educators and researchers.

2. METHOD

2.1. IDENTIFICATION OF SURVEYS

The first goal of this study was to identify and characterize surveys developed to assess post-secondary students' attitudes toward statistics. Surveys were identified from personal communications, dissertations, conference proceedings, book chapters, and peer-reviewed journal articles. Characteristics, including year of publication, number and descriptions of subscales, number of items, response format, and availability were extracted from publications describing and/or using these surveys, or from follow-up contact with the instruments' authors when necessary. Results from this phase of the study are summarized in Table 1, providing readers with an overview of the surveys that have been developed to assess students' attitudes toward statistics, which is, to our knowledge, complete and up-to-date.

Table 1. Instruments to assess students' attitudes toward statistics and their characteristics

Survey	Year	Pub.	Survey Subscales	Items	Scale	Availability
<i>Statistics Course Attitude Scale (SCAS)</i> ⁴	1954	1	Likely a single dimension.	30 -ve items	5-point	Not readily available
<i>Attitudes Toward Quantitative Concepts (ATQC)</i> ¹⁸	1978	1	Likely a single dimension.	14 +/-ve items	4-point	Available in publication ¹⁵
<i>Statistics Attitude Survey (SAS)</i> ³⁶	1980	1	Single dimension	33 +/-ve items	5-point	Currently available from Dr. Roberts
<i>Attitudes Toward Statistics Scale (ATS)</i> ⁵²	1985	1	1. Field (20 items) 2. Course (9items)	29 +/-ve items	5-point	Available online from http://bit.ly/T13ATj
<i>Multi-factorial Scale of Attitudes Toward Statistics (MSAS)</i> ²	1991	3	1. Motivation (4 items) 2. Enjoyment (5 items) 3. Confidence (4 items) 4. Anxiety (6 items) 5. Usefulness (6 items)	25 +/-ve items	5-point	Available in publication ²
<i>Statistics Attitude Scale (SASc)</i> ²⁵	1991	3	Likely a single dimension.	20 +/-ve items	5-point	Available in publication ²⁷
<i>Students' Attitudes Toward Statistics A (STATS-A)</i> ⁴²	1992	3	1. Interest and perceived usefulness (6 items) 2. Relationship to and impact of the instructor (4 items) 3. Attitude toward statistical tools (3 items) 4. Self-confidence (3 items) 5. Parental influence (2 items) 6. Initiative and extra effort (3 items)	21 +/-ve items	10-point	Available in publication ⁴⁴
<i>Student Attitude Toward Statistics B (STATS-B)</i> ³⁹	1993	4	1. Attitude toward mathematics (10 items) 2. Attitude toward computers (10 items) 3. Attitude toward statistics (9 items) 4. Attitude toward tests (5 items) 5. Attitude toward instructors (6 items)	40 items	7-point	Available in dissertation ³⁹

Table 1. Instruments to assess students' attitudes toward statistics and their characteristics

Survey	Year	Pub.	Survey Subscales	Items	Scale	Availability
<i>Attitudes Regarding Graduate Statistics Scale (ARGSS)</i> ¹⁴	1994	3	-	-	-	Not readily available
<i>Survey of Attitudes Toward Statistics - 28 (SATS-28)</i> ⁴⁰	1995	1	1. Affect (6 items) 2. Cognitive competence (6 items) 3. Value (9 items) 4. Difficulty (7 items)	28 +/-ve items	7-point	Available from Dr. Schau, see www.evaluationandstatistics.com
<i>Survey of Attitudes Toward Statistics - 36 (SATS-36)</i> ⁴⁶	2003	5	Dimensions of the SATS-28, and 5. Interest (4 items) 6. Effort (4 items)	36 +/-ve items	7-point	Available from Dr. Schau, see www.evaluationandstatistics.com
<i>Bad Attitudes Toward Statistics (BATS)</i> ⁵	1998	1	1. Usefulness (10 items) 2. Course (7 items)	17 +/-ve items	6-point	Not readily available
<i>Students' Attitudes and Conceptions in Statistics (STACS)</i> ¹⁶	2005	4	1. Attitudes (30 items) 2. Statistical Concepts (14 items)	44 +/-ve items	5-point	Currently available from Dr. Evans
<i>Students' Attitudes Toward Statistics Questionnaire (SATSQ)</i> ⁵	2005	2	1. Affective a. Interest b. Anxiety 2. Evaluative a. Current perceived utility b. Future perceived utility	27 +/-ve items	5-point	Available in publication ³⁰
<i>Students' Attitudes Toward Statistics and Technology Scale (SASTSc)</i> ¹	2011	1	1. Statistics cognitive competence (6 items) 2. Technology cognitive competence (4 items) 3. Learning statistics with technology (6 items) 4. Value of the discipline (6 items) 5. Affect (6 items)	28 +ve items	7-point	Available in publication ¹

Note. Year = Survey's initial publication year; Pub. = Publication level (1 = Peer-reviewed journal article, 2 = Book chapter, 3 = Conference proceeding, 4 = Dissertation, 5 = Personal communication); - = Not reported; Not readily available = Author could not be contacted, or survey is no longer available.

2.2. SYSTEMATIC REVIEW GUIDELINES

Once surveys assessing students' attitudes toward statistics had been identified, an analysis of the construct validity and internal consistency evidence was conducted. The systematic review was conducted according to Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Moher, Liberati, Teztlaff, Altman, & Group, 2009). The PRISMA statement was developed to standardize the information reported in systematic reviews, and additional information may be found at <http://www.prisma-statement.org/>.

2.3. LITERATURE SEARCH

The electronic databases PsycINFO, Education Research Information Center (ERIC), Web of Science and Scopus were searched, and included articles from 1806 to the week of October 2, 2011. For PsycINFO and ERIC, four major search concepts and their synonyms were developed and used: "students' attitudes," "statistics," "attitude measures," and "test validity." For Web of Science and Scopus, author and survey names were searched. Finally, article references were manually searched for additional citations. A complete search strategy is available upon request from the primary author. Due to the conflicting evidence regarding the interaction between attitudes and anxiety in statistics education, statistics anxiety was not included in this systematic review. Please see Onwuegbuzie and Wilson (2003) for a review of statistics anxiety.

2.4. INCLUSION AND EXCLUSION CRITERIA AND RESULTS

To be included in the systematic review of validity and reliability evidence, articles needed to: (1) be written in English; (2) be published in a peer-reviewed journal; and (3) present extractable validity and/or reliability data for instruments assessing students' attitudes toward statistics. Articles were excluded if (1) a tool was used to measure students' attitudes toward statistics, but no validity or reliability data were presented; (2) a tool was used to measure statistics anxiety or another construct besides students' attitudes toward statistics; or (3) a full text copy of the article could not be located.

A total of 532 citations were identified by our search strategy, 486 of which were unique. The title and abstract of each citation was reviewed, and included or excluded based on the aforementioned selection criteria. Review articles were not included. Seventy-eight citations required full text review by two of the authors, and of these, 43 were excluded, leaving 35 peer-reviewed articles eligible for analysis of validity and reliability evidence. In the case of a disagreement between reviewers, a full text copy of the article was jointly reviewed.

2.5. DATA EXTRACTION AND ANALYSIS

Evidence of content, substantive, structural, and external validity, and of internal consistency in the form of Cronbach's α was extracted from 35 peer-reviewed articles. We first examined content and substantive validity, which often were presented in the form of written descriptions. Structural validity evidence was examined next and typically consisted of exploratory (EFA) or confirmatory factor analysis (CFA) results, although it should be noted that CFA results are considered stronger evidence than EFA results. For EFA, the number of factors, and, where available, their names and the percentage of variance accounted for by each factor as well as the total variance (%) are reported. For CFA, the identified factors are reported. In addition, we have reported correlations between factors where available. External validity evidence is reported in the form of correlation, regression, or structural equation model (SEM) results. Correlations consist of Pearson product moment correlations (r), or Spearman's rank order correlation coefficients (r_s). Regression and SEM results are presented as partial correlations (r) and variance explained (%). Internal consistency estimates in the form of Cronbach's α were extracted and are presented as unweighted means and ranges where three or more estimates were available. The heterogeneous group of outcome variables precluded a full meta-analysis. Instead, the construct validity and internal consistency data were organized into summary tables.

3. RESULTS

3.1. SURVEYS FOR THE ASSESSMENT OF STUDENTS' ATTITUDES TOWARD STATISTICS

Fifteen survey instruments assessing students' attitudes toward statistics were identified from peer-reviewed and non-peer-reviewed sources (Table 1). The earliest was published in 1954 (*Statistics Course Attitude Scale*, SCAS), and the latest in 2011 (*Students' Attitudes Toward Statistics and Technology Scale*, SASTSc). Seven instruments were published in peer-reviewed journals: SCAS, *Attitudes Toward Quantitative Concepts* (ATQS), *Statistics Attitude Survey* (SAS), *Attitudes Toward Statistics* (ATS), *Survey of Attitudes Toward Statistics-28* (SATS-28), *Bad Attitudes Toward Statistics* (BATS), and SASTSc. One was published as a book chapter (*Students' Attitudes Toward Statistics Questionnaire*, SATSQ), four as conference papers (*Multi-factorial Scale of Attitudes Toward Statistics*, MSAS; *Statistics Attitudes Scale*, SASc; *Students' Attitudes Toward Statistics*, STATS-A; and *Attitudes Regarding Graduate Statistics Scale*, ARGSS), two as thesis chapters (*Student Attitudes Toward Statistics*, STATS-B; and *Students' Attitudes and Conceptions in Statistics*, STACS), and one as a personal communication (*Survey of Attitudes Toward Statistics-36*, SATS-36).

The construct of students' attitudes toward statistics was interpreted differently across surveys, as indicated by differences in the number of subscales (see Table 1). Instruments included one (SCAS, ATQC, SAS, and SASc), two (ATS, BATS, STACS), four (SATS-28, SATSQ), five (MSAS, STATS-B, SASTSc), or six underlying dimensions (STATS-A and SATS-36). Generally, newer instruments tend to have a larger number of subscales than older instruments. The dimensions were named and defined differently amongst surveys; however, three common themes were identified: students' feelings about statistics (MSAS, SATS-28, SATS-36, SATSQ, and SASTSc), perceived value or usefulness of statistics (ATS, MSAS, STATS-A, SATS-28, SATS-36, BATS, SATSQ, and SASTSc), and perceived ability to understand statistics (ATS, MSAS, STATS-A, SATS-28, SATS-36, BATS, SASTSc).

As indicated in Table 1, of the surveys that reported the number of items, the shortest was 14 (ATQC) and the longest was 44 (STACS). Most surveys were designed with positively and negatively worded items to prevent response acquiescence bias, with the SCAS and the SASTSc being the only two known exceptions. All surveys used Likert-type response formats, except the ARGSS, for which this information was not reported. Most surveys used 5- or 7-point scales, with the ATQC, BATS, STATS-A as the only surveys to use 4-, 6-, and 10-point response formats, respectively. Most surveys used scales anchored with "strongly disagree" and "strongly agree." The SATS-28 and SATS-36 are the only surveys known to have differently worded versions for pre-course and post-course administrations; however, many of the surveys in Table 1 have been administered at multiple time points to measure changes in students' attitudes toward statistics.

3.2. CONSTRUCT VALIDITY EVIDENCE

At this time, only four surveys have been used in more than one peer-reviewed, published article examining score validity: SAS, ATS, SATS-28, and SATS-36. An accumulation of content, structural, and external validity evidence was found for only these four questionnaires, and is summarized in Tables 2a-d. Validity and reliability evidence for the remaining instruments is summarized in Table 3.

Content and Substantive Validity Generally, the development of survey content was not well documented. As Tables 2a-d and Table 3 indicate, descriptions of this process were found for 11 surveys (SCAS, ATQC, SAS, ATS, SASc, SATS-28, SATS-36, BATS, STACS, SATSQ, and SASTSc). The content development process for three surveys (SCAS, SATS-28, and SASTSc) involved students, and two (ATS and SATS-28) involved experts (i.e., statistics educators). The most rigorous content development process appears to have been employed for the SATS-28, in which a formal method (Nominal Group Technique; Moore, 1987) involving students and faculty was detailed (Schau, Stevens, Dauphinee, & Del Vecchio, 1995).

Substantive validity evidence was even more sparse than content validity evidence. Only three of the fifteen surveys identified a theoretical basis for interpretation of survey responses. Authors of the SATS-28 and SATS-36 claim that scores represent attitudes defined by expectancy-value, social cognition, and goal theories of learning (Tables 2c and 2d), and developers of the STACS utilize self-efficacy theory of learning to interpret attitudes scores.

Structural Validity Tables 2a-d and Table 3 indicate that the structural validity of survey scores has been examined for the SCAS, SAS, ATS, SASc, SATS-28, SATS-36, BATS, SATSQ, and SASTSc; however, multiple studies of structural validity evidence exist for scores from only the ATS (Table 2b), SATS-28 (Table 2c), SATS-36 (Table 2d), and SATSQ (Table 3). The structural validity evidence for scores from these four surveys is summarized next.

For the scores of the ATS, principal-axis factor analysis with SATS-28 item responses (Cashin & Elmore, 2005), principal components factor analysis with SAS item responses (Waters, Martelli, Zakrajsek, & Popovich, 1988a), and principal factor solution analyses (Wise, 1985) have indicated two moderately correlated factors (Table 2b). These results are consistent with its author's claim that the ATS assesses two underlying dimensions: *Field* and *Course*, where *Field* refers to students' attitudes toward the value of statistics in their field and *Course* refers to attitudes toward taking a statistics course.

The structural validity of SATS-28 scores has been assessed by parceled CFA on at least seven occasions. Each study confirmed that the SATS-28 items measure four underlying dimensions: *Affect* refers to students' feelings regarding statistics, *Cognitive Competence* refers to their attitudes about intellectual knowledge and skills applied to statistics, *Value* refers to their perception of the value of statistics generally and in their field, and *Difficulty* refers to their perception of the difficulty of statistics as a subject (Schau et al., 1995). These subscale scores have generally been shown to be moderately to highly intercorrelated (Table 2c). Parallel exploratory factor analysis with SATS-28 item responses, and principal-axis factor analysis with SATS-28 and ATS item responses, performed by Cashin and Elmore (2005), indicated only two underlying dimensions (*Affect-Cognitive Competence-Difficulty* and *Value*). However, when one, three, and four factor models were compared using parceled CFA, only the four factor model (*Affect*, *Cognitive Competence*, *Value*, and *Difficulty*) produced an acceptable (p -value > 0.01) goodness-of-fit χ^2 result (Dauphinee, Schau, & Stevens, 1997).

Regarding the SATS-36, Table 2d shows that parceled and unparceled CFA yield six factors. This result is consistent with its authors' claim that it assesses six underlying dimensions (Coetzee & Merwe, 2010; Nasser, 2004; Tempelaar, Schim van der Loeff, & Gijselaers, 2007; VanHoof, Kuppens, Sotos, Verschaffel, & Onghena, 2011). Vanhoof et al. have also suggested that a four-factor model, combining the *Affect*, *Cognitive Competence*, and *Difficulty* subscales, does not result in a substantial loss of information.

The structural validity of the fourth scale, the SATSQ, has been investigated twice. Exploratory factor analysis identified four moderately correlated factors, consistent with its author's claim that it assesses two bi-dimensional factors: *Affective*, which includes *Interest* and *Anxiety*, and *Evaluative*, which include students' perceptions of the *Current Value of Statistics* and *Future Value of Statistics* (Table 3).

Convergent and Discriminant Validity Convergent validity has only been established among only three instruments: the SAS, ATS, and SATS-28 (Cashin & Elmore, 2005; Chiesi & Primi, 2009; Roberts & Reese, 1987; Schau et al., 1995; Waters, Martelli, Zakrajsek, & Popovich, 1988b) (Tables 2a-c). Specifically, the total score of the SAS has high positive correlations with the *Course* and *Field* subscales (Waters et al., 1988b) and total score of the ATS (Roberts & Reese, 1987), and moderate to strong positive correlations with the *Affect*, *Cognitive Competence*, *Value* and *Difficulty* subscales of the SATS-28 (Table 2a).

Table 2a. Summary of peer-reviewed validity evidence and relationships to other variables for the SAS (see Table 2d for definitions of table abbreviations)

Survey	Content and Substantive	Structural	Convergent and Discriminant	Predictive
SAS	<p><i>Content</i></p> <p>³⁶Items assessing a single dimension based on Dutton's¹⁵ mathematics content domain</p> <p>³⁸Assesses students' perceptions of how competent they are and how useful statistics is</p> <p><i>Substantive</i></p> <p>-</p>	<p>⁴⁹PCFA with ATS items: two factors (Field, Course), but total score could also be used</p> <p><i>Between-Factor Correlation</i></p> <p>$r = ^{49} -0.46^{\text{Pre}}, -0.43^{\text{Post}}$</p>	<p><i>Convergent with Statistics Attitudes Survey:</i></p> <p>³⁷ATS Total $r = 0.88$</p> <p>⁵⁰ATS</p> <p><i>Field</i> $r = 0.73^{\text{Post}}$</p> <p><i>Course</i> $r = 0.83^{\text{Post}}$</p> <p>⁸SATS-28</p> <p><i>A</i> $r = 0.69^{\text{Pre}}, 0.80^{\text{Pre}}, 0.70^{\text{Post}}, 0.87^{\text{Post}}$</p> <p><i>CC</i> $r = 0.64^{\text{Pre}}, 0.77^{\text{Pre}}, 0.64^{\text{Pre}}, 0.82^{\text{Post}}$</p> <p><i>V</i> $r = 0.59^{\text{Pre}}, 0.67^{\text{Pre}}, 0.41^{\text{Post}}, 0.74^{\text{Post}}$</p> <p><i>D</i> $r = 0.55^{\text{Pre}}, 0.64^{\text{Pre}}, 0.48^{\text{Post}}, 0.66^{\text{Post}}$</p> <p><i>Discriminant with Other Attitude Surveys</i></p> <p>³⁸Pre-course calculator attitudes</p> <p>$r = 0.17^{\text{Pre}}, \text{n.s.}^{\text{Post}}$</p> <p>Post-course calculator attitudes</p> <p>$r = \text{n.s.}^{\text{Pre, Post}}$</p>	<p>³⁶Total course points</p> <p>$r = 0.33^{\text{S1}}, 0.48^{\text{S2}}, 0.54^{\text{S3}}$</p> <p>³⁷Grade $r = 0.14$</p> <p>³⁸Grade $r = 0.26^{\text{Pre}}, 0.41^{\text{Post}}$</p> <p>⁵⁰Grade $r = 0.38^{\text{Post}}$</p>

Table 2b. Summary of peer-reviewed validity evidence and relationships to other variables for ATS (see Table 2d for definitions of table abbreviations)

Survey	Content and Substantive	Structural	Convergent and Discriminant	Predictive
ATS	<p><i>Content</i></p> <p>⁵²Items assessing two subscales were assessed for content validity by five statistics instructors</p> <p><i>Substantive</i></p> <p>-</p>	<p>⁸Principal-axis FA with SATS-28 items: two factors (Field 37%, Course 10%, Total 48%)</p> <p>⁴⁹PCFA with SAS items: two factors (Field, Course), but total score could also be used</p> <p>⁵²Principal factor solution^{S1}: two factors (Field, Course, Total 49%)</p> <p><i>Between-Factor Correlation</i></p> <p>$r = ^{35} 0.32^{\text{Pre}}, 0.56^{\text{Post}}, ^{8} -0.44^{\text{Post}}, ^{49} -0.46^{\text{Pre}}, -0.43^{\text{Post}}, ^{50} 0.56^{\text{Pre}}, 0.55^{\text{Post}}, ^{52} 0.33^{\text{S1}}$</p>	<p><i>Convergent with Statistics Attitudes Surveys</i></p> <p>SAS (see Table 2a)</p> <p>⁴⁰SATS-28</p> <p><i>Field-A</i> $r = 0.34$</p> <p><i>Field-CC</i> $r = 0.38$</p> <p><i>Field-V</i> $r = 0.76$</p> <p><i>Field-D</i> $r = \text{n.s.}$</p> <p><i>Course-A</i> $r = 0.79$</p> <p><i>Course-CC</i> $r = 0.76$</p> <p><i>Course-V</i> $r = 0.40$</p> <p><i>Course-D</i> $r = 0.42$</p> <p>⁸SATS-28</p> <p><i>Field-A</i> $r = 0.44^{\text{Pre}}, 0.48^{\text{Pre}}, 0.46^{\text{Post}},$</p>	<p>⁵²Grade</p> <p><i>Field</i> $r = \text{n.s.}^{\text{S2Post}}$</p> <p><i>Course</i> $r = 0.27^{\text{S2Post}}$</p> <p>⁵⁰Grade</p> <p><i>Field</i> $r = \text{n.s.}^{\text{Pre}}, 0.17^{\text{Post}}$</p> <p><i>Course</i> $r = 0.20^{\text{Pre}}, 0.42^{\text{Post}}$</p> <p>³⁷Grade</p> <p><i>Total</i> $r = 0.16$</p> <p>³⁵Grade</p> <p><i>Field</i> $r = \text{n.s.}^{\text{Pre, Post}}$</p> <p><i>Course</i> $r = 0.29^{\text{Pre, Post}}$</p> <p>⁴⁸Statistics exam results for 1st year undergraduates</p> <p><i>Course</i> $r = 0.33^{\text{Pre}}, 0.47^{\text{Post}}$</p>

Table 2b. Summary of peer-reviewed validity evidence and relationships to other variables for ATS (see Table 2d for definitions of table abbreviations)

Survey	Content and Substantive	Structural	Convergent and Discriminant	Predictive
			0.60 ^{Post} <i>Field-CC</i> $r = 0.38^{\text{Pre}}, 0.47^{\text{Pre}}, 0.42^{\text{Post}}, 0.56^{\text{Post}}$ <i>Field-V</i> $r = 0.66^{\text{Pre}}, 0.78^{\text{Pre}}, 0.69^{\text{Post}}, 0.83^{\text{Post}}$ <i>Field-D</i> $r = 0.29^{\text{Pre}}, 0.31^{\text{Pre}}, 0.22^{\text{Post}}, 0.30^{\text{Post}}$ <i>Course-A</i> $r = 0.68^{\text{Pre}}, 0.81^{\text{Pre}}, 0.74^{\text{Post}}, 0.89^{\text{Post}}$ <i>Course-CC</i> $r = 0.62^{\text{Pre}}, 0.71^{\text{Pre}}, 0.68^{\text{Post}}, 0.84^{\text{Post}}$ <i>Course-V</i> $r = 0.47^{\text{Pre}}, 0.51^{\text{Post}}, 0.60^{\text{Post}}$ <i>Course-D</i> $r = 0.50^{\text{Pre}}, 0.61^{\text{Pre}}, 0.51^{\text{Post}}, 0.61^{\text{Post}}$ <i>Total-A</i> $r = 0.62^{\text{Pre}}, 0.71^{\text{Pre}}, 0.64^{\text{Post}}, 0.79^{\text{Post}}$ <i>Total-CC</i> $r = 0.56^{\text{Pre}}, 0.69^{\text{Pre}}, 0.58^{\text{Post}}, 0.75^{\text{Post}}$ <i>Total-V</i> $r = 0.66^{\text{Pre}}, 0.73^{\text{Pre}}, 0.68^{\text{Post}}, 0.81^{\text{Post}}$ <i>Total-D</i> $r = 0.45^{\text{Pre}}, 0.49^{\text{Pre}}, 0.38^{\text{Post}}, 0.48^{\text{Post}}$	General exam results for 1 st , 2 nd , 3 rd , 4 th , and 5 th year undergraduates $r = \text{n.s.}$ Statistics exam results for 2 nd and 3 rd , and 5 th year undergraduates $r = \text{n.s.}$ ²⁴ Statistics grade <i>Field</i> $r = -0.15$ <i>Course</i> $r = -0.26$ <i>Total</i> $r = -0.24$ (low ATS scores meant more positive attitudes)
			⁹ SATS-28 <i>Field-A</i> $r = 0.19^{\text{Pre}}, 0.33^{\text{Post}}$ <i>Field-CC</i> $r = 0.21^{\text{Pre}}, 0.26^{\text{Post}}$ <i>Field-V</i> $r = 0.55^{\text{Pre}}, 0.62^{\text{Post}}$ <i>Field-D</i> $r = \text{n.s.}^{\text{Pre}}, 0.24^{\text{Post}}$ <i>Course-A</i> $r = 0.57^{\text{Pre}}, 0.69^{\text{Post}}$ <i>Course-CC</i> $r = 0.60^{\text{Pre}}, 0.67^{\text{Post}}$ <i>Course-V</i> $r = 0.27^{\text{Pre}}, 0.47^{\text{Post}}$ <i>Course-D</i> $r = 0.27^{\text{Pre}}, 0.50^{\text{Post}}$	
			<i>Discriminant with Other Attitude Surveys</i> -	

Table 2c. Summary of peer-reviewed validity evidence and relationships to other variables for SATS-28 (see Table 2d for definitions of table abbreviations)

Survey	Content and Substantive	Structural	Convergent and Discriminant	Predictive
SATS-28	<p><i>Content</i></p> <p>⁴⁰Using modified NGT³¹ two undergraduate, two graduate, and two instructors developed items for four subscales.</p> <p><i>Substantive</i></p> <p>²³Congruent with expectancy value, social cognition, and goal theories of learning.</p>	<p>⁴⁰Parceled CFA: four factors (<i>A, CC, V, D</i>)</p> <p>⁴⁴Parceled CFA: four factors (<i>A, CC, V, D</i>)</p> <p>⁹Parceled CFA: four factors (<i>A, CC, V, D</i>)</p> <p>Structural invariance by time, except <i>A</i> and <i>D</i> covariance (minor)</p> <p>¹²CFA: four factors (<i>A, CC, V, D</i>)</p> <p>Structural invariance by gender, except <i>V</i> and <i>A</i> covariance and <i>V</i> variance (minor)</p> <p>²³CFA: four factors (<i>A, CC, V, D</i>)</p> <p>Structural invariance by gender, time, and gender*time, except all post variances larger than pre variances but male <i>CC</i> equal across time</p> <p>²⁸CFA: four factors (<i>A, CC, V, D</i>)</p> <p>⁸Parallel FA: two factors (<i>A-CC-D</i> 38%, <i>V</i> 10.3%, Total 47%)</p> <p><i>Between-Factor Correlation</i></p> <p><i>A-CC</i> $r =$⁴⁰0.92, ⁴⁴0.91, ¹²0.94, ²³0.94, ³²0.62, ⁹0.86, ¹³0.71^{Pre}, 0.75^{Post}</p> <p><i>A-V</i> $r =$⁴⁰0.37, ⁴⁴0.42, ¹²0.32, ²³0.56, ³²0.45, ⁹0.26, ¹³n.s.^{Pre}, 0.26^{Post}</p> <p><i>A-D</i> $r =$⁴⁰0.69, ⁴⁴0.67, ¹²0.73, ²³0.72, ³²0.41, ⁹0.57, ¹³0.66^{Pre}, 0.50^{Post}</p> <p><i>CC-V</i> $r =$⁴⁰0.43, ⁴⁴0.50, ¹²0.39, ²³0.55, ³²0.36, ⁹0.30, ¹³0.38^{Pre}, 0.45^{Post}</p> <p><i>CC-D</i> $r =$⁴⁰0.62, ⁴⁴0.71, ¹²0.64, ²³0.70, ³²0.32, ⁹0.46, ¹³0.55^{Pre}, 0.44^{Post}</p> <p><i>V-D</i> $r =$⁴⁰0.09, ⁴⁴0.14, ²³0.29, ³²0.27, ⁹0.20, ¹³0.23^{Pre}, n.s.^{Post}</p> <p><i>A-CC-D</i> with <i>V</i> $r =$⁸0.44^{Post}</p> <p><i>Course-Value</i> $r =$⁸-0.44^{Post}</p>	<p><i>Convergent with Statistics Attitude Surveys</i></p> <p>SAS (see Table 2a)</p> <p>ATS (see Table 2b)</p> <p><i>Discriminant with Other Attitude Surveys:</i></p> <p>³²Attitudes toward math</p> <p><i>A</i> $r =$0.49</p> <p><i>CC</i> $r =$0.47</p> <p><i>D</i> $r =$0.18</p>	<p>³²Statistics achievement</p> <p><i>A</i> $r =$0.18</p> <p><i>CC</i> $r =$0.28</p> <p><i>V</i> $r =$n.s.</p> <p><i>D</i> $r =$n.s.</p> <p>14% of variance explained with attitudes toward math, math anxiety and motivation to succeed (SEM)</p> <p>⁸Statistics achievement</p> <p><i>A</i> $r =$0.36^{Pre}, 0.45^{Post}</p> <p><i>CC</i> $r =$0.32^{Pre}, 0.43^{Post}</p> <p><i>V</i> $r =$0.28^{Pre}, 0.32^{Post}</p> <p><i>D</i> $r =$0.26^{Pre}, 0.30^{Post}</p> <p>21% of variance explained (MLR)</p> <p>⁹Statistics achievement</p> <p><i>A</i> $r =$0.19^{Pre}, 0.29^{Post}</p> <p><i>CC</i> $r =$0.21^{Pre}, 0.29^{Post}</p> <p><i>V</i> $r =$0.10^{Pre}, 0.20^{Post}</p> <p><i>D</i> $r =$0.04^{Pre}, 0.17^{Post}</p> <p>2%^{Pre} and 10%^{Post} of variance explained (SLR)</p> <p>¹³Statistics achievement</p> <p><i>A</i> $r =$n.s.^{Pre}, 0.23^{Post}</p> <p><i>CC</i> $r =$n.s.^{Pre}, 0.31^{Post}</p> <p><i>V</i> $r =$n.s.^{Pre}, 0.23^{Post}</p> <p><i>D</i> $r =$n.s.^{Pre, Post}</p> <p>⁴¹Course grade</p> <p><i>A</i> $r =$0.21</p> <p><i>V</i> $r =$n.s.</p>

Table 2d. Summary of peer-reviewed validity evidence and relationships to other variables for the SATS-36

Survey	Content and Substantive	Structural	Convergent and Discriminant	Predictive
SATS-36	<p><i>Content</i> ⁴⁶Items developed for two additional subscales</p> <p><i>Substantive</i> ⁴⁶<i>Interest</i> and <i>Effort</i> subscales increase structural fidelity with expectancy-value theory of learning</p>	<p>⁴⁶Parceled CFA: six factors (<i>A</i>, <i>CC</i>, <i>V</i>, <i>D</i>, <i>I</i>, <i>E</i>) ¹⁰Parceled CFA: six factors (<i>A</i>, <i>CC</i>, <i>V</i>, <i>D</i>, <i>I</i>, <i>E</i>) ⁴⁷Unparceled CFA: six (<i>A</i>, <i>CC</i>, <i>V</i>, <i>D</i>, <i>I</i>, <i>E</i>) or four (<i>A-CC-D</i>, <i>V</i>, <i>I</i>, <i>E</i>) factors</p> <p><i>Between-Factor Correlation</i> <i>A-CC</i> $r = ^{46}0.80, ^{47}0.88$ <i>A-V</i> $r = ^{46}0.40, ^{47}0.39$ <i>A-D</i> $r = ^{46}0.61, ^{47}0.85$ <i>A-I</i> $r = ^{46}0.42, ^{47}0.48$ <i>A-E</i> $r = ^{46}$n.s., ⁴⁷-0.14 <i>CC-V</i> $r = ^{46}0.43, ^{47}0.43$ <i>CC-D</i> $r = ^{46}0.62, ^{47}0.86$ <i>CC-I</i> $r = ^{46}0.35, ^{47}0.48$ <i>CC-E</i> $r = ^{46}-0.17, ^{47}-0.12$ <i>V-D</i> $r = ^{46}$ n.s., ⁴⁷0.36 <i>V-I</i> $r = ^{46}0.63, ^{47}0.72$ <i>V-E</i> $r = ^{46}0.34, ^{47}0.16$ <i>D-I</i> $r = ^{46}$ n.s., ⁴⁷0.47 <i>D-E</i> $r = ^{46}-0.28, ^{47}-0.23$ <i>I-E</i> $r = ^{46}0.44, ^{47}0.20$</p>	<p><i>Convergent with Statistics Attitude Surveys</i> - <i>Discriminant with Other Attitude Surveys</i> -</p>	<p>⁴⁷Course performance 10% of variance explained with statistical reasoning assessment (SEM) ⁴³Correct statistical reasoning <i>A</i> $r = 0.12$ <i>CC</i> $r = 0.12$ <i>V</i> $r = 0.10$ <i>D</i> $r = 0.11$ <i>I</i> $r =$ n.s. <i>E</i> $r =$ n.s. Statistical reasoning misconceptions <i>A</i> $r = -0.07$ <i>CC</i> $r = -0.06$ <i>V</i> $r = -0.06$ <i>D</i> $r = -0.10$ <i>I</i> $r =$ n.s. <i>E</i> $r = 0.17$ ⁴⁶Exam scores <i>A</i> $r =$ n.s. <i>CC</i> $r = 0.39$ <i>V</i> $r =$ n.s. <i>D</i> $r = -0.17$ <i>I</i> $r =$ n.s. <i>E</i> $r =$ n.s. ⁶GPA <i>A</i> $r_s =$ n.s., ^{Pre}, ^{Post}0.32 <i>CC</i> $r_s =$ n.s., ^{Pre}, ^{Post}0.39 <i>V</i> $r_s =$ n.s., ^{Pre, Post} <i>D</i> $r_s =$ n.s., ^{Pre}, ^{Post}0.44 <i>I</i> $r_s =$ n.s., ^{Pre, Post} <i>E</i> $r_s =$ n.s., ^{Pre, Post}</p>

Notes: PCFA = Principal Components Factor Analysis, FA = Factor Analysis, CFA = Confirmatory Factor Analysis, *A* = Affect, *CC* = Cognitive Competence, *V* = Value, *D* = Difficulty, *I* = Interest, *E* = Effort, SLR = Simple Linear Regression, MLR = Multiple Linear Regression, SEM = structural equation modeling, r = Pearson's correlation coefficient, r_s = Spearman's rank order correlation coefficient, n.s. = Non-significant, - = No evidence found. Superscripted numbers preceding entries indicate citations. Superscripts S1, S2, and S3 indicate subsamples if applicable. Superscripts Pre and Post indicate data from surveys administered at the beginning and end of courses, respectively, if applicable. Peer-reviewed validity evidence for the remaining surveys may be found in Table 3.

Table 3. Peer-reviewed validity and internal consistency evidence for other surveys designed to assess students' attitudes toward statistics

Survey	Content and Substantive	Structural	Convergent and Discriminant	Predictive	Internal Consistency
SCAS	<p><i>Content</i> ⁴Items based on in-class discussion and anonymously written paragraphs</p> <p><i>Substantive</i> -</p>	<p>⁴Variance in survey scores explained by 10% Computational preference 13% Persuasive preference 12% Amount of high school math 11% Amount of college math 23% Number of psychology courses (MLR^{S2})</p> <p>-</p>	-	<p>⁴Quantified letter grade $r = 0.24^{S1}, 0.21^{S2}$</p>	-
ATQC	<p><i>Content</i> ¹⁸Modification of Dutton's scale¹⁵ by removing an item</p> <p><i>Substantive</i> -</p>	-	-	<p>¹⁸Course performance $r = 0.35$ Predicted success, failure, or drop-out (χ^2 test, $p < 0.05$)</p>	-
SASc	<p><i>Content</i> ²²McCall²⁵ reportedly developed three clusters of qualitatively different items</p> <p><i>Substantive</i> -</p>	<p>²²PCFA: one factor (77%)</p>	-	-	²² $\alpha = 0.95$
BATS	<p><i>Content</i> ⁵Items assessing two subscales based on the ATS</p> <p><i>Substantive</i> -</p>	<p>⁵PCFA: two factors (<i>Usefulness</i> 32%, <i>Course</i> 11%)</p> <p><i>Between-Factor Correlation</i> $r = .53$</p>	-	<p>⁵Statistics achievement <i>Usefulness</i> $r = n.s.^{S1-3Pre, S2Post}$, $0.38^{S3}, 0.45^{S1}$ <i>Course</i> $r = n.s.^{S1-3Pre, Post}$ <i>Total</i> $r = n.s.^{S1-3Pre, Post}$</p>	<p>⁵<i>Usefulness</i> $\alpha = 0.81$ <i>Course</i> $\alpha = 0.81$</p>
STACS	<p><i>Content</i> ¹⁷Developed to measure two dimensions based on pre-</p>	-	-	<p>¹⁷Course grade <i>Attitudes</i> $r = 0.20^{Pre}, 0.25^{Post}$ <i>Concepts</i> $r = n.s.^{Pre, Post}$</p>	<p>¹⁷<i>Attitudes</i> $\alpha = 0.92$ <i>Concepts</i> $\alpha = 0.59$</p>

Survey	Content and Substantive	Structural	Convergent and Discriminant	Predictive	Internal Consistency
	existing surveys.				
	<i>Substantive</i> ¹⁶ Claimed to be congruent with Bandura's self-efficacy theory of learning				
SATSQ	<i>Content</i> ³ Items based on Wise ⁵² and Auzmendi ² developed to assess two bi-dimensional factors <i>Substantive</i> -	³ FA: four factors (Total 53%) ³⁰ FA: four factors <i>Between-Factor Correlations</i> <i>I-A</i> $r = 0.39^{\text{Males}}, 0.48^{\text{Females}}$ <i>I-C</i> $r = 0.60^{\text{Males}}, 0.50^{\text{Females}}$ <i>I-F</i> $r = 0.62^{\text{Males}}, 0.55^{\text{Females}}$ <i>A-C</i> $r = 0.51^{\text{Males}}, 0.44^{\text{Females}}$ <i>A-F</i> $r = 0.48^{\text{Males}}, 0.32^{\text{Females}}$ <i>C-F</i> $r = 0.64^{\text{Males}}, 0.62^{\text{Females}}$	-	-	³ <i>Affective Total</i> $\alpha = 0.87$ <i>Evaluative Total</i> $\alpha = 0.85$ ³⁰ <i>Interest</i> $\alpha = 0.84^{\text{Males}}, 0.85^{\text{Females}}$ <i>Anxiety</i> $\alpha = 0.84^{\text{Males}}, 0.82^{\text{Females}}$ <i>Current Utility</i> $\alpha = 0.74^{\text{Males}}, 0.72^{\text{Females}}$ <i>Future Utility</i> $\alpha = 0.76^{\text{Males}}, 0.82^{\text{Females}}$
SASTSc	<i>Content</i> ¹ Items assessing five subscales based on student input <i>Substantive</i> -	¹ PCFA: five factors (SCC 26%, TCC 16%, LWT 11%, Value 8%, Affect 7%)	-	-	¹ Total $\alpha = 0.91$

Note. PCFA = Principal Components Factor Analysis; r = Pearson's correlation coefficient; I = Interest; A = Anxiety; C = Current Value; C = Future Value; SCC = Statistics Cognitive Competence; TCC = Technology Cognitive Competence; LWT = Learning With Technology; MLR = multiple linear regression; - = No evidence found. Superscripted numbers preceding entries indicate citations. Superscripts S1 and S2 indicate subsamples if applicable. Superscripts Pre and Post indicate data from surveys administered at the beginning and end of courses, respectively, if applicable. No peer-reviewed evidence for any aspect of validity or reliability could be found for the MSAS, STATS-A, STATS-B, and ARGSS surveys.

For the ATS, the *Course* subscale scores were moderately and positively correlated with the *Value* and *Difficulty* subscale scores, and highly correlated with the *Affect* and *Cognitive Competence* subscale scores of the SATS-28. In contrast, the relationships between the *Field* subscale scores of the ATS and the SATS-28 subscale scores ranged from weak (for *Difficulty*) to strong (for *Value*). Its correlation with the *Value* subscale of the SATS-28 was high (Cashin & Elmore, 2005; Schau et al., 1995) (Table 2b).

Little evidence of discriminant validity was found in the research (Table 2a-d and Table 3). Scores from the SAS have been compared with scores from a measure of students' attitudes toward calculators, revealing weak and non-significant relationships (Roberts & Saxe, 1982) (Table 2a). Also, scores from the SATS-28 were compared with scores from a measure of students' attitudes toward mathematics, with significant small to moderate relationships observed (Nasser, 2004) (Table 2c).

Predictive Validity Eight instruments have been used to investigate the degree to which measures of students' attitudes toward statistics are able to predict academic performance in statistics courses (SCAS, ATQC, SAS, ATS, SATS-28, SATS-36, BATS, and STACS). As Tables 2a-d and Table 3 indicate, academic performance was measured in several ways, such as academic success (passing versus failing), letter grade, course grade, and examination scores.

Across instruments, correlations between measures of students' attitudes toward statistics and academic performance ranged from weak and non-significant to moderate, positive and significant (Tables 2a-d and Table 3), with the highest correlation between scores from the SAS and total course points ($r = 0.54$, Table 2a). Studies that collected pre-course and post-course measures of students' attitudes generally revealed higher correlations between post-course measures and academic performance (Tables 2a-d). Structural equation modeling and regression results indicate that students' attitudes toward statistics as measured by the SATS-28 and SATS-36 account for between 2% and 21% of the variance in students' achievement (Chiesi & Primi, 2009; Nasser, 2004; VanHoof et al., 2011).

3.3. RELIABILITY EVIDENCE

Internal consistency in the form of Cronbach's α was found for nine of the fifteen instruments (SAS, ATS, SASc, SATS-28, SATS-36, BATS, STACS, STATSQ, and SASTSc). There were only five instruments for which internal consistency had been examined multiple times: the SAS, ATS, SATS-28, SATS-36 (Table 4), and the STATSQ (Table 3). The method of reporting internal consistency differed across studies with some presenting pre-course estimates separately from post-course estimates, some pooling them together, and others using only one administration at any time during the course. Additionally, some studies reported internal consistency estimates separately for each subscale, whereas others reported totals. Overall, internal consistency was generally good or excellent ($\alpha \geq 0.75$), except for scores from the *Difficulty* subscale of the SATS-28 ($\alpha = 0.51$ to 0.91), which are considered at least acceptable. Interestingly, Table 4 reveals that the internal consistency of the total score of the ATS is very high (0.89 to 0.94). These values have been used as an argument for using total scores in place of subscale scores for the ATS (Roberts & Reese, 1987; Waters et al., 1988a).

Table 4. Peer-reviewed Internal Consistency Evidence (Cronbach's α) Presented as Ranges for the SAS, ATS, SATS-28 and SATS-36

Scale	Pre-Course Administration (α)	Post-Course Administration (α)	Pooled Administrations (α)	Single Administration (α)
SAS	0.93 (0.92 to 0.93) ^{8,37,38}	0.93 (0.93 to 0.94) ^{8,38,50}	-	0.94 (0.93 to 0.95) ³⁶
ATS				
<i>Field</i>	0.88 (0.83 to 0.93) ^{8,24,35,48,50}	0.91 (0.86 to 0.96) ^{8,35,48,50}	-	0.92 ^{34,52}
<i>Course</i>	0.86 (0.77 to 0.90) ^{8,24,35,48,50}	0.90 (0.85 to 0.92) ^{8,35,48,50}	-	0.90 ⁵² , 0.91 ³⁴
Total	0.92 (0.91 to 0.93) ^{8,24,37,48}	0.89 ⁴⁸ , 0.94 ⁸	-	-
SATS-28				
<i>Affect</i>	0.81 (0.73 to 0.85) ^{8,9,23,28,40,44,45}	0.84 (0.81 to 0.89) ^{8,9,23,41,51}	0.81 ¹³	0.74 ³²
<i>Cognitive Competence</i>	0.78 (0.71 to 0.84) ^{8,9,23,28,40,44,45}	0.80 (0.74 to 0.86) ^{8,9,23,51}	0.85 ¹³	0.79 ³²
<i>Value</i>	0.81 (0.75 to 0.88) ^{8,9,23,28,40,44,45}	0.79 (0.63 to 0.90) ^{8,9,23,41,51}	0.81 ¹³	0.80 ³²
<i>Difficulty</i>	0.68 (0.61 to 0.74) ^{8,9,23,28,40,44,45}	0.67 (0.51 to 0.76) ^{8,9,23,51}	0.75 ¹³	0.65 ³²
SATS-36				
<i>Affect</i>	0.82 ⁴⁶ , 0.88 ⁶	0.87 ⁶	0.88 ⁷	0.80 ¹⁰
<i>Cognitive Competence</i>	0.78 ⁴⁶ , 0.93 ⁶	0.86 ⁶	0.79 ⁷	0.80 ¹⁰
<i>Value</i>	0.78 ⁴⁶ , 0.93 ⁶	0.58 ⁶	0.90 ⁷	0.83 ¹⁰
<i>Difficulty</i>	0.68 ⁴⁶ , 0.91 ⁶	0.81 ⁶	0.81 ⁷	0.66 ¹⁰
<i>Interest</i>	0.80 ⁴⁶ , 0.89 ⁶	0.84 ⁶	0.85 ⁷	0.83 ¹⁰
<i>Effort</i>	0.76 ⁴⁶ , 0.83 ⁶	0.71 ⁶	0.79 ⁷	0.85 ¹⁰

Note. Where three or more estimates of internal consistency were available, the unweighted mean and range is presented. Otherwise, the values are presented. Internal consistency evidence for the remaining surveys can be found in Table 3.

4. DISCUSSION

4.1. PRIMARY FINDINGS

A systematic review of the research identified fifteen instruments that purportedly assess students' attitudes toward statistics; however, the content, substantive, structural, and external validity evidence, as well as internal consistency, were substantial for only four instruments: the SAS, ATS, SATS-28, and SATS-36. Furthermore, all four types of validity evidence, including all three subtypes of external validity evidence (convergent, discriminant, and predictive) were only found for scores from one instrument: the SATS-28. A discussion of these findings for content, substantive, structural, and predictive validity, and internal consistency follows.

Content and Substantive Validity Generally, articles introducing surveys of students' attitudes toward statistics did not provide adequate evidence of content validity by thoroughly detailing the process by which survey items were developed. Thus, scores from these surveys may not adequately or accurately sample the domain of students' attitudes toward statistics. The SATS-28 items appear to be the most rigorously developed, and, hence, most likely to represent the population of items assessing students' attitudes toward statistics.

Substantive validity evidence for the instruments was weak. The revision of the SATS-28 to the SATS-36 to increase congruency with the expectancy-value model of learning suggests the use of an *a priori* theory to create the *Interest* and *Effort* subscales. If the items in these two subscales were developed through the same process used for the SATS-28, then scores from the SATS-36 are likely to have the most convincing content and substantive validity evidence. However, as the initial development and validation of the SATS-36 has not been published, it is difficult to draw this conclusion with certainty.

Structural Validity Evidence of structural validity generally confirmed authors' claims about their surveys' factors, but not unanimously so. In the case of the SAS, for which there was only one study that examined the structural validity of its scores, a unidimensional factor could not be confirmed (Waters et al., 1988a).

Criticisms of the SAS, namely that it was more likely to measure previous mathematics exposure or mathematics ability than attitudes toward statistics, motivated Wise to develop the ATS (Wise, 1985). The two factor structure of the ATS has been confirmed multiple times; however, the high internal consistency for the ATS total score has led some researchers to advocate for the use of its total score instead of subscale scores (Roberts & Reese, 1987).

Structural validity evidence for the SATS-28 and SATS-36 scores has been somewhat conflicting, likely due to the fact that the *Affect*, *Cognitive Competence*, and *Difficulty* subscales are moderately to highly intercorrelated. Confirmatory factor analyses of SATS-28 scores, in which items were parceled, have supported a four factor model (Dauphinee et al., 1997; Schau et al., 1995); however, parallel EFA has indicated that the *Affect*, *Cognitive Competence*, and *Difficulty* subscales could be collapsed into a single factor, resulting in a two factor model similar to the ATS (Cashin & Elmore, 2005). Similarly, whereas parceled CFA of SATS-36 scores has confirmed a six factor model, unparceled CFA has indicated that a four factor solution in which the *Affect*, *Cognitive Competence*, and *Difficulty* subscales were collapsed into a single factor performed similarly to the six factor model (VanHoof et al., 2011).

The structural validity evidence suggests that students' attitudes toward statistics can be considered a unidimensional construct or a multidimensional one, according to the measure used (Cashin & Elmore, 2005; VanHoof et al., 2011). Several reasons have been given to maintain the dimensions of the SATS-28 and SATS-36 (Dauphinee et al., 1997; Hilton, Schau, & Olsen, 2004; Schau et al., 1995). The four factor model for the SATS-28 and six factor model for the SATS-36 have been confirmed in several studies. Maintaining these factors allows for comparisons with previous studies, and differential patterns of association have been observed for these three subscales. Vanhoof et al. suggest that a two factor model of the SATS-28 and a four factor model of the SATS-36 may be used when a global measure of students' attitudes toward statistics is all that is needed, but

that subscales should be considered separately when associations between specific subscales or between subscales and other measures are of interest.

As was noted in the introduction of this paper, a multiplicity of interpretations of the construct of students' attitudes toward statistics exist, and was evidenced in this study by the array of scales and subscales in the fifteen surveys identified. From this group of surveys, however, three common elements did emerge, namely affect, perceived ability to learn and/or understand statistics, and perceived value of statistics. For statistics educators interested in improving students' attitudes toward statistics, developing materials that target these three areas may be most beneficial.

External Validity In spite of structural differences, when the SAS, ATS, and SATS-28 were co-administered, scores were generally moderately to strongly correlated across some subscales. In particular, the ATS *Field* subscale was strongly correlated with the SATS-28 *Value* subscale and the ATS *Course* subscale was strongly correlated with the SATS-28 *Affect* and *Cognitive Competence* subscale scores and moderately correlated with the *Value* and *Difficulty* subscale scores (Table 2a-d). These results indicate that all three instruments are likely assessing at least some of the same underlying dimensions.

Discriminant validity evidence was especially scant, making it impossible to conclude whether or not patterns of weak relationships exist between students' attitudes toward statistics surveys and other attitudinal surveys. Thus, it is possible that the measures of statistics attitudes are measuring other constructs. For example, scores from the SATS-28 *Affect*, *Cognitive Competence*, and *Value* subscales were shown to be moderately and significantly correlated with measures of attitudes toward mathematics, indicating a possible lack of discriminant validity.

Students' attitudes toward statistics as measured by a number of the instruments identified in this study have successfully been used to predict academic performance in statistics courses. The most sophisticated demonstrations of this relationship were structural equation models using the SATS-28 (Nasser, 2004) and SATS-36 (Tempelaar, Schim van der Loeff, et al., 2007) instruments, in which students' attitudes toward statistics were shown to account for as much as 14% of the variance in student achievement. This evidence supports the claim that students' attitudes influence academic performance. Using these surveys to identify students who are at risk of underperforming or failing statistics courses is an example of how educators could apply these results to improving academic performance in statistics courses.

Internal Consistency The results of this systematic review indicate moderate to high internal consistency among many of the instruments identified. Only scores from the *Difficulty* subscale from both versions of the SATS sometimes exhibited lower values; however, even these values were in the acceptable range. Vanhoof et al. (2011) and Coetzee and Merwe (2010) deleted items from the *Difficulty* subscale on the SATS-36, which may have contributed to improved item performance on that version of the survey, and not on the SATS-28.

In this study, only evidence of internal consistency (Cronbach's α) was examined; however, it is important to acknowledge that a number of factors influence the reliability of survey scores including population demographics, administration time, and environment. Many of the studies included in this review sampled North American undergraduate students enrolled in introductory statistics courses in social sciences programs. Thus, it is not possible to conclude that using these instruments in samples from other populations (e.g., students in other programs, professionals), in the future, or under different conditions (e.g., online) would elicit equally valid and reliable scores.

Summary of Primary Findings If the construct validity evidence for scores from the SATS-28 can be applied to scores from the SATS-36, then interpretations of students' attitudes toward statistics based on SATS-36 scores appear to have the strongest evidence of construct validity and internal consistency, at this time. An advantage of using this survey is that educators and researchers can choose whether to use a four or six dimensional model of students' attitudes toward statistics, depending on the level of detail they are interested in. However, it should be emphasized that this conclusion depends on the assumption that validity and reliability evidence for SATS-28 scores can be applied to SATS-36 scores. Also, it should be noted that some of the surveys in Table 3 are quite recent, and have not yet been systematically studied.

4.2. RECOMMENDATIONS FOR FUTURE RESEARCH

Recommendations for future research center on the need for additional, peer-reviewed validation research and improved consistency in reporting reliability evidence. Although four instruments were identified that had been used in multiple validation studies, none of the items and scores from the underlying dimensions had accumulated a large amount of content, substantive, structural, or external validity, and none specifically possessed evidence of all four. Furthermore, eleven additional instruments exist, which require extensive additional validation research before interpretations can be deemed adequate and appropriate. Content, substantive, convergent, and divergent validity evidence appear to be aspects of the validity argument that are commonly overlooked.

The reporting of reliability evidence was inconsistent across studies, and should be improved. Ideally, researchers should report Cronbach's α for each subscale and for each administration time. In the future, generalizability studies should be undertaken to examine the extent to which population, temporal, and environmental factors influence the generalizability of survey scores among samples with differing characteristics, from different institutions, and enrolled in different course. This is important considering that much research has been conducted with North American undergraduate students in social sciences programs.

Finally, during this systematic review, it became apparent that much research on this topic exists in the form of unpublished manuscripts, doctoral dissertations, and conference abstracts. For example, only half of the instruments identified were published in peer-reviewed journals, and less than a third have been used in more than one published, peer-reviewed validation or reliability study. Unpublished works were frequently cited in the literature; however, they should be interpreted with caution until subjected to the complete process of peer-reviewed publication. Furthermore, we encourage researchers to pursue peer-reviewed publication of any validation research conducted.

4.3. LIMITATIONS

As in any systematic review, there are potential sources of bias, which may have influenced our findings, such as publication bias (i.e., only peer-reviewed articles were included in the analysis of validity and reliability data), language bias (i.e., articles were restricted to English language), or insufficient search strategy (i.e., failure to develop all appropriate search terms). However, several large databases were searched, and manual searching of article references was conducted to minimize the impact of some of these limitations.

4.4. CONCLUSION

The valid and reliable assessment of students' attitudes toward statistics is an important tool for predicting academic performance, allowing educators to identify and assist those whose negative attitudes may lead to sub-optimal performance. Improving students' attitudes through educational interventions may also increase confidence in their ability to understand statistical information and problem-solve using statistics beyond the classroom. This systematic review benefits statistics educators and researchers by providing a summary of the validity and reliability evidence for the construct of students' attitudes toward statistics underlying currently available survey instruments.

ACKNOWLEDGEMENTS

The authors would like to acknowledge the assistance of L. Toews in conducting the electronic database searches, and would also like to sincerely thank the reviewers and editorial team, especially C. Schau, whose comments and suggestions made a significant contribution to this manuscript.

REFERENCES

* Denotes articles included in the systematic review of validity and reliability evidence.

- ^{*1}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.
- ²Auzmendi, E. (1991, April). *Factors related to attitudes toward statistics: A study with a Spanish sample*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.
- ³Bayot, A., Mondejar, J., Mondejar, J. A., Monsalve, F., & Vargas, M. (2005). The difficulties of learning concepts in the social sciences. In M. Misztal & M. Trawinski (Eds.), *Studies in teacher education: Psychopedagogy* (pp. 242–258). Krakow: Wydawnictwo Naukowe Akademii Pedagogicznej.
- ^{*4}Bendig, A. W., & Hughes, J. B. (1954). Student attitude and achievement in a course in introductory statistics. *Journal of Educational Psychology*, 45(5), 268–276.
- ^{*5}Berk, R. A., & Nanda, J. P. (1998). Effects of jocular instructional methods on attitudes, anxiety, and achievement in statistics courses. *HUMOR: International Journal of Humor Research*, 11(4), 383–410.
- ^{*6}Carlson, K. A., & Winqvist, J. R. (2011). Evaluating an active learning approach to teaching introductory statistics: A classroom workbook approach. *Journal of Statistics Education*, 19(1). [Online: <http://www.amstat.org/publications/jse/v19n1/carlson.pdf>]
- ^{*7}Carnell, L. J. (2008). The effect of a student-designed data collection: Project on attitudes toward statistics. *Journal of Statistics Education*, 16(1). [Online: <http://www.amstat.org/publications/jse/v16n1/carnell.html>]
- ^{*8}Cashin, S. E., & Elmore, P. B. (2005). The *Survey of Attitudes Toward Statistics* scale: A construct validity study. *Educational and Psychological Measurement*, 65(3), 509–524.
- ^{*9}Chiesi, F., & Primi, C. (2009). Assessing statistics attitudes among college students: Psychometric properties of the Italian version of the *Survey of Attitudes toward Statistics* (SATS). *Learning and Individual Differences*, 19(2), 309–313. [Online: <http://dx.doi.org/10.1016/j.lindif.2008.10.008>]
- ^{*10}Coetzee, S., & Merwe, P. v. d. (2010). Industrial psychology students' attitudes towards statistics. *SA Journal of Industrial Psychology*, 36, 1–8.
- ¹¹Cronbach, L. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16(3), 297–332.
- ^{*12}Dauphinee, T. L., Schau, C., & Stevens, J. J. (1997). *Survey of Attitudes Toward Statistics*: Factor structure and factorial invariance for women and men. *Structural Equation Modeling*, 4(2), 129–141.
- ^{*13}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). [Online: <http://www.amstat.org/publications/jse/v17n2/dempster.html>]
- ¹⁴Douzenis, C. (1994, April). *Graduate statistics education: A dilemma for students and faculty*. Paper presented at the annual meeting of the American Educational Research Association, New Orleans.
- ¹⁵Dutton, W. H., & Blum, M. P. (1968). The measurement of attitudes toward arithmetic with a Likert-type test. *The Elementary School Journal*, 68(5), 259–264.
- ¹⁶Evans, B. (2005). *Student attitudes, conceptions, and achievement in introductory undergraduate college statistics* (Doctoral dissertation). Temple University. Available from Dissertation Abstracts International. (AAI3178774)
- ^{*17}Evans, B. (2007). Student attitudes, conceptions, and achievement in introductory undergraduate college statistics. *Mathematics Educator*, 17(2), 24–30.
- ^{*18}Feinberg, L. E., & Halperin, S. (1978). Affective and cognitive correlates of course performance in introductory statistics. *Journal of Experimental Education*, 46(4), 11–18.
- ¹⁹Gal, I., & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework. *Journal of Statistics Education*, 2(2), 1–15. [Online: <http://www.amstat.org/publications/jse/v2n2/gal.html>]

- ²⁰Gal, I., Ginsburg, L., & Schau, C. (1997). Monitoring attitudes and beliefs in statistics education. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 37–51). Amsterdam: IOS Press.
- ²¹Garfield, J. (1995). How students learn statistics. *International Statistical Review*, 63(1), 25–34.
- ²²Glencross, M. J., & Cherian, V. I. (1992). Attitudes toward applied statistics of postgraduate education students in Transkei. *Psychological Reports*, 70(1), 67–75.
[Online: <http://dx.doi.org/10.2466/pr0.1992.70.1.67>]
- ²³Hilton, S. C., Schau, C., & Olsen, J. A. (2004). Survey of Attitudes Toward Statistics: Factor structure invariance by gender and by administration time. *Structural Equation Modeling*, 11(1), 92–109.
- ²⁴Kottke, J. L. (2000). Mathematical proficiency, statistics knowledge, attitudes towards statistics, and measurement course performance. *College Student Journal*, 34(3), 334–347.
- ²⁵McCall, C. H., Belli, G., & Madjidi, F. (1990, August). *The complexities of teaching graduate students in educational administration introductory statistical concepts*. Paper presented at the Third International Conference on Teaching Statistics, Dunedin, New Zealand.
- ²⁶Messick, S. (1989). Validity. In R. L. Linn (Ed.), *Validity* (3rd ed., pp. 13–103). New York: Macmillan.
- ²⁷Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749.
- ²⁸Mills, J. D. (2004). Students' attitudes toward statistics: Implications for the future. *College Student Journal*, 38(3), 349.
- ²⁹Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & Group, T. P. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7).
[Online: <http://www.plosmedicine.org/article/info%3Adoi%2F10.1371%2Fjournal.pmed.1000097>]
- ³⁰Mondejar-Jimenez, J., & Vargas-Vargas, M. (2010). Determinant factors of attitude towards quantitative subjects: Differences between sexes. *Teaching and Teacher Education*, 26(3), 688–693.
[Online: <http://dx.doi.org/10.1016/j.tate.2009.10.004>]
- ³¹Moore, C. M. (1987). *Group techniques for idea building* (Vol. 9). Thousand Oaks, CA: Sage Publications.
- ³²Nasser, F. (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).
[Online: <http://www.amstat.org/publications/jse/v12n1/nasser.html>]
- ³³Onwuegbuzie, 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.
- ³⁴Perepiczka, M., Chandler, N., & Becerra, M. (2011). Relationship between graduate students' statistics self-efficacy, statistics anxiety, attitude toward statistics, and social support. *The Professional Counselor: Research and Practice*, 1(2), 99–108.
- ³⁵Rhoads, T. R., & Hubele, N. F. (2000). Student attitudes toward statistics before and after a computer-integrated introductory statistics course. *IEEE Transactions on Education*, 43(2), 182–187.
- ³⁶Roberts, D. M., & Bilderback, E. W. (1980). Reliability and validity of a statistics attitude survey. *Educational and Psychological Measurement*, 40(1), 235–238.
- ³⁷Roberts, D. M., & Reese, C. M. (1987). A comparison of two scales measuring attitudes towards statistics. *Educational and Psychological Measurement*, 47(3), 759–764.
- ³⁸Roberts, D. M., & Saxe, J. E. (1982). Validity of a statistics attitude survey: A follow-up study. *Educational and Psychological Measurement*, 42(3), 907–912.
- ³⁹Rogness, N. T. (1993). *The development and validation of a multi-factorial instrument to measure students' attitudes toward a course in statistics*. (Doctoral dissertation), University of Northern Colorado. Available from Proquest. (AAT 9413644)

- ^{*40}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.
[Online: <http://dx.doi.org/10.1177/0013164495055005022>]
- ^{*41}Schutz, P. A., Drogosz, L. M., White, V. E., & Distefano, C. (1998). Prior knowledge, attitude, and strategy use in an introduction to statistics course. *Learning and Individual Differences*, 10(4), 291–308.
[Online: <http://dx.doi.org/10.1016/S1041-6080%2899%2980124-1>]
- ^{*42}Sutarso, T. (1992, November). *Students' attitudes toward statistics (STATS)*. Paper presented at the annual meeting of the Mid-South Educational Research Association, Knoxville, TN.
- ^{*43}Tempelaar, D. T., Gijsselaers, W. H., & Schim van der Loeff, S. (2006). Puzzles in statistical reasoning. *Journal of Statistics Education*, 14(1).
[Online: www.amstat.org/publications/jse/v14n1/tempelaar.html]
- ^{*44}Tempelaar, D. T., Gijsselaers, W. H., Schim van der Loeff, S., & Nijhuis, J. F. (2007). A structural equation model analyzing the relationship of student achievement motivations and personality factors in a range of academic subject matter areas. *Contemporary Educational Psychology*, 32(1), 105–131.
[Online: <http://dx.doi.org/10.1016/j.cedpsych.2006.10.004>]
- ^{*45}Tempelaar, D. T., Schim van der Loeff, S., Gijsselaers, W. H., & Nijhuis, J. F. (2011). On subject variations in achievement motivations: A study in business subjects. *Research in Higher Education*, 52(4), 395–419.
[Online: <http://dx.doi.org/10.1007/s11162-010-9199-7>]
- ^{*46}Tempelaar, D. T., Schim van der Loeff, S., & Gijsselaers, 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: http://www.stat.auckland.ac.nz/~iase/serj/SERJ6%282%29_Tempelaar.pdf]
- ^{*47}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), 35–51.
[Online: http://www.stat.auckland.ac.nz/~iase/serj/SERJ10%281%29_Vanhoof.pdf]
- ^{*48}Vanhoof, S., Sotos, A. E. C., Onghena, P., Verschaffel, L., Van Dooren, W., & Van den Noortgate, W. (2006). Attitudes toward statistics and their relationship with short-and long-term exam results. *Journal of Statistics Education*, 14(3).
[Online: <http://www.amstat.org/publications/jse/v14n3/vanhoof.html>]
- ^{*49}Waters, L., Martelli, T., Zakrajsek, T., & Popovich, P. M. (1988a). Factor analyses of two measures of attitudes toward statistics. *Educational and Psychological Measurement*, 48(4), 1037–1041.
[Online: <http://dx.doi.org/10.1177/0013164488484020>]
- ^{*50}Waters, L., Martelli, T. A., Zakrajsek, T., & Popovich, P. M. (1988b). Attitudes toward statistics: An evaluation of multiple measures. *Educational and Psychological Measurement*, 48(2), 513–516.
[Online: <http://dx.doi.org/10.1177/0013164488482026>]
- ^{*51}Wiberg, M. (2009). Teaching statistics in integration with psychology. *Journal of Statistics Education*, 17(1).
[Online: <http://www.amstat.org/publications/jse/v17n1/wiberg.html>]
- ^{*52}Wise, S. L. (1985). The development and validation of a scale measuring attitudes toward statistics. *Educational and Psychological Measurement*, 45(2), 401–405.

MEAGHAN NOLAN

Department of Veterinary Clinical and Diagnostic Sciences
Faculty of Veterinary Medicine
Health Sciences G380
University of Calgary
3330 Hospital Drive NW
Calgary, AB, Canada