STRUCTURAL EQUATION MODELING, *11*(1), 92–109

Survey of Attitudes Toward Statistics: Factor Structure Invariance by Gender and by Administration Time

Sterling C. Hilton
Department of Educational Leadership and Foundations
Brigham Young University

Candace Schau
CS Consultants
Albuquerque, NM

Joseph A. Olsen
College of Family, Home, and Social Sciences
Brigham Young University

In addition to student learning, positive student attitudes have become an important course outcome for many introductory statistics instructors. To adequately assess changes in mean attitudes across introductory statistics courses, the attitude instruments used should be invariant by administration time. Attitudes toward statistics from 4,910 students enrolled in an introductory statistics course were measured using the Survey of Attitudes Toward Statistics (SATS) both at the beginning and at the end of the semester. Confirmatory factor analysis on the covariance structure was used to determine the gender and time invariance properties of the SATS. Results indicate that the SATS is gender, time, and Gender × Time invariant with respect to factor loadings and factor correlations. Gender was invariant with respect to 3 of the 4 factor variances; variances from these same 3 factors were larger at the end than at the beginning of the course. Having established that the SATS is factorially invariant with respect to gender, time, and Gender × Time, its component scores can be used appropriately to examine mean attitude differences for these 2 variables and their interaction.

Many undergraduate postsecondary programs in the United States require successful completion of an introductory statistics course for graduation. Large and increasing numbers of students enroll in these courses (Loftsgaarden & Watkins, 1998).

Traditionally, the major course outcome assessed in postsecondary introductory statistics courses has been student learning. Recently, however, positive student atti-

Requests for reprints should be sent to Sterling C. Hilton, Department of Educational Leadership and Foundations, Brigham Young University, 306 MCKB, Provo, UT 84602. E-mail: hiltons@byu.edu

tudes have been included as another important course outcome (e.g., Gal, Ginsburg, & Schau, 1997; Garfield, Hogg, Schau, & Whittinghill, 2000). Students and instructors alike believe that students' attitudes toward statistics affect course enrollment, persistence, achievement, and climate (e.g., Gal et al., 1997). Evidence to support this belief is slowly growing. Most studies report small to moderate relations between several components of attitudes toward statistics, measured using a variety of surveys, and course achievement, usually measured by test scores or course grades. (See Gal et al. and Harris and Schau [1999] for brief summaries of this research.)

If students' attitudes toward statistics are important, we must be able to assess them. To be most useful, a statistics attitudes measure needs to yield valid scores at least at the beginning and at the end of courses as well as across groups of students with differing characteristics.

STATISTICS ATTITUDE SURVEYS

Currently, scores from at least 10 Likert-type instruments have been used to assess attitudes toward statistics. These instruments include the Multifactorial Scale of Attitudes Toward Statistics (Auzmendi, 1991), Bad Attitude Toward Statistics (Berk & Nanda, 1998), Statistical Anxiety Rating Scale (Cruise, Cash, & Bolton, 1985), Statistics Attitude Scale (McCall, Belli, & Madjidi, 1990), Statistics Attitude Survey (Roberts & Bilderback, 1980), Survey of Attitudes Toward Statistics (SATS; Schau, Stevens, Dauphinee, & Del Vecchio, 1995), Student's Attitudes Toward Statistics (Sutarso, 1992), Attitudes Toward Statistics (Wise, 1985), unnamed (Zanakis & Valenzi, 1997), and Statistics Anxiety Inventory (Zeidner, 1991).

Although somewhat similar in content, the number of subscale scores (and so by implication the number of dimensions or components of attitudes toward statistics) included in these instruments ranges from one (Roberts & Bilderback, 1980) to six (Cruise et al., 1985; Sutarso, 1992). The internal structure of these measures has seldom been explored statistically. When this exploration has occurred, researchers almost universally have used exploratory factor analysis to identify the survey's internal structure or to confirm its hypothesized factor structure.

SATS

The SATS (for a description of the development process, see Schau et al., 1995) is the most carefully developed of these measures (Harris & Schau, 1999). The results from this process indicated that scores from the SATS assess four components of attitudes toward statistics (see Dauphinee, Schau, & Stevens, 1997; Schau et al., 1995). These components include the following:

1. *Affect* (six items): Students' positive and negative feelings about statistics.

2. *Cognitive Competence* (six items): Attitudes about the students' intellectual knowledge and skills when applied to statistics.
3. *Value* (nine items): Attitudes about the usefulness, relevance, and worth of statistics in personal and professional life.
4. *Difficulty* (seven items): Attitudes about the difficulty of statistics as a domain.

Each of these components is consistent with a variety of theories, including expectancy value (e.g., Atkinson, 1957; Eccles & Wigfield, 1995), attribution (e.g., Weiner, 1979), social cognition (e.g., Bandura, 1977), and goal theories (e.g., Maehr, 1984), although no one theory postulates the importance of all four.

Using confirmatory factor analysis (CFA), this four-factor structure provided a good fit to responses collected at the beginning of introductory statistics courses using a sample of undergraduate students attending a major Southwestern university (Schau et al., 1995). This structure was also tested for gender invariance. The four-factor model fit well for presemester responses from White male and female subgroups from the original undergraduate sample. With two minor exceptions, the fit was invariant across gender (Dauphinee et al., 1997). The standardized variance of Value was greater for women (0.69) than for men (0.47). The correlation between Value and Affect was greater for men (0.48) than for women (0.33).

Coefficient alpha values have varied somewhat by component. Across various samples, alpha values for Affect have ranged from .81 to .85; for Cognitive Competence, .77 to .83; for Value, .80 to .85; and for Difficulty, .64 to .77 (Schau et al., 1995).

GENDER DIFFERENCES AND CHANGE IN ATTITUDES TOWARD STATISTICS

The two areas of research most relevant to the purposes of this project include gender differences in attitudes toward statistics and change in attitudes across statistics courses. Much more research exists on gender differences.

About one half of the studies that have examined gender differences reported that men expressed more positive attitudes toward statistics (e.g., Auzmendi, 1991; Roberts & Bilderback, 1980; Roberts & Saxe, 1982; Waters, Martelli, Zakrajsek, & Popovich, 1988); these differences usually were small. The other one half found no gender differences (e.g., Cherian & Glencross, 1997; Faghihi & Rakow, 1995; Schau, Dauphinee, & Del Vecchio, 1992; Sutarso, 1992; Tomazic & Katz, 1988; Wisenbaker & Scott, 1997). A few studies have reported more positive attitudes for women on one scale but not on the others (Rhoads & Hubele, 2000; Zeidner, 1991) or at one time of administration but not at the other (Rhoads & Hubele, 2000).

Most of the limited number of studies that have examined attitude change reported that at least some components of attitudes become more positive across statistics courses (e.g., Harlow, Burkholder, & Morrow, 2002; Katz & Tomazic, 1998; Perney & Ravid, 1990; Roberts & Saxe, 1982; Sorge, Schau, Hubele, & Kennedy, 2000; Waters et al., 1988). Some found no change (Green, 1993; Rhoads & Hubele, 2000; Shultz & Koshino, 1998). Schau (2000) reported an increase in one attitude component and decreases in two others.

The inconsistent results from these studies are likely due in part to the variety of attitude measures, student samples, and administration times used. In addition, some of the more recent research tested attitude change as part of the evaluation of an innovative course based on elements of the reform movement in statistics education (e.g., Harlow et al., 2002). All of these studies tested mean differences in attitudes. These differences should be tested only after the factorial equivalence of the measures' internal structure has been established for the groups and administration times involved (e.g., Rensvold & Cheung, 1998). The SATS is the only measure with any evidence of gender equivalence in its internal structure; none of the surveys have been tested for equivalence of factor structure from pre- to postsemester.

PURPOSE

Using data collected almost one decade later from a population of students attending a different university, this research is designed to continue the exploration of the invariance of the SATS four-factor structure begun by Dauphinee et al. (1997). The purpose of this research is twofold: (a) to test the gender invariance of the four-factor structure of the SATS to determine if we can replicate Dauphinee et al.'s results; and (b) to extend the invariance testing to include administration time (at the beginning and end of a semester in an introductory statistics course), as well as gender and administration time simultaneously.

METHODS

Sample

The original sample consisted of 5,360 students attending Brigham Young University, who enrolled in an undergraduate introductory statistics course offered by the statistics department over a four-semester period: Fall 1998, Winter 1999, Fall 1999, and Winter 2000. This course can be used to satisfy a university general education requirement, and it is a required course for approximately one third of the students enrolled. Over the course of the data collection phase of our study, class size ranged from 160 to 200 students with either six or eight classes taught per semester. Classes met three times each week for a 50 min lecture; students also enrolled in 1-hr weekly recitation session consisting of 20 students. Types of assessments used in the course were homework problems, which sometimes required the use of computers and exams that consisted of both multiple-choice and short-answer essay questions.

The sample was evenly split on gender: 51% male and 49% female. The average age of the students was 21.4 ($SD = 2.6$) years. Ethnicity was not collected on the students in the sample; however, given the large number of students enrolled in the course each semester, we expect approximately the same ethnic mix as that found at the university level: approximately 89% White, 4% Asian and Pacific Islander, 3% Hispanic, 0.6% Native American, 0.4% African American, and 3% unknown.

The analysis sample, whose demographic characteristics closely matched those of the original sample, consisted of the 4,910 students (92%) who had complete responses on the SATS both pre- and postsemester. Forty-five students (1%) completed the postsemester SATS but did not complete the presemester SATS; 206 students (4%) completed the presemester SATS but did not complete the post. The majority of these latter students dropped the class during the first 2 weeks of the semester. There were 98 students (2%) missing one response from either the pre- or the postsemester administration; the remaining 101 students (2%) had more than one missing response from the pre-, post-, or both semester administrations.

For the model-testing phase of the analysis, we selected a random sample of 567 students (281 women and 286 men) from the Fall 1998 semester who had complete data. For the model-confirmatory phase, we used the remaining 4,343 students with complete data; see Appendixes A and B for the covariance matrices used in these analyses.

Instrument

The SATS, which can be found at www.unm.edu/~cschau/infopage.htm, contains 28 Likert-type items. The presemester and postsemester items are identical except for tense. Students rate each item on a scale from 1 (*strongly disagree*) to 7 (*strongly agree*). Although we recognize that in general Cronbach's α coefficients tend to underestimate reliability (Raykov, 1997), we report the coefficient alpha values by gender and administration time from the analysis sample in Table 1 because they represent a lower bound for the true reliability of the factors. Because the observed alpha coefficients range between .68 and .90, we are satisfied that these alpha coefficients are sufficiently high to indicate scale reliability.

Procedure

As part of a larger course evaluation project, students were encouraged to respond to the SATS instrument twice during the semester. Presemester SATS responses were collected during the first 2 weeks of the 16-week semester. For the first two semesters of data collection, the presemester responses were collected at the university testing center, and for the last two semesters they were collected online via an internal university Web site. Postsemester responses were collected in paper-and-pencil form at the end of the final exam.

TABLE 1
Cronbach Coefficient Alpha Values for the Four Subscales of the SATS

	Presemester			*Postsemester*		
Factor	*Male*	*Female*	*Total*	*Male*	*Female*	*Total*
Affect	.80	.84	.83	.81	.86	.84
Cognitive Competence	.82	.84	.84	.81	.85	.84
Value	.88	.87	.87	.90	.90	.90
Difficulty	.68	.74	.72	.72	.79	.76

Note. SATS = Survey of Attitudes Toward Statistics.

Analysis Strategy

Parcels. Responses to negatively worded items were reversed. Therefore, higher scores on each item always represented a more positive affect, greater perceived competence and value, and less difficulty. To help avoid the inherent nonnormality associated with single item distributions, items within each hypothesized subscale were grouped into parcels (Gribbons & Hocevar, 1998). We used the same parcels used by Schau et al. (1995) and by Dauphinee et al. (1997); the latter article lists the items comprising each parcel and describes the parcel development process. Two parcels contained the Affect items, two contained the Cognitive Competence items, and two contained the Difficulty items. Three parcels contained the Value items. Each parcel included either three or four items. Each student was given a mean score on each parcel; these scores served as the observed measures in the subsequent analyses. The resulting 36 parcel distributions (9 for men and 9 for women from presemester responses and another 18 from postsemester responses) were smooth with no apparent outliers. All but 2 had kurtosis values between ±1; those 2 were less than +1.25.

Model. Figure 1 presents the path diagram of the model for a single gender. Parcels were allowed to load only on their hypothesized factor, and factors were allowed to covary. We made the standard assumption that parcel errors were independent with the exception that the errors from the same pre- and postsemester parcels were allowed to covary. The mean structure of the model was unconstrained in all analyses.

Invariance tests. The equality of model parameters was tested across both gender and time of administration (beginning of semester vs. end of semester) using nested model comparison chi-square tests. Due to the relatively large sample size, we used an alpha level of 0.01 to determine statistical significance. Our general strategy imposed constraints first across gender but not time, then across time but not gender, and finally, simultaneously across both gender and time. Using this ap-

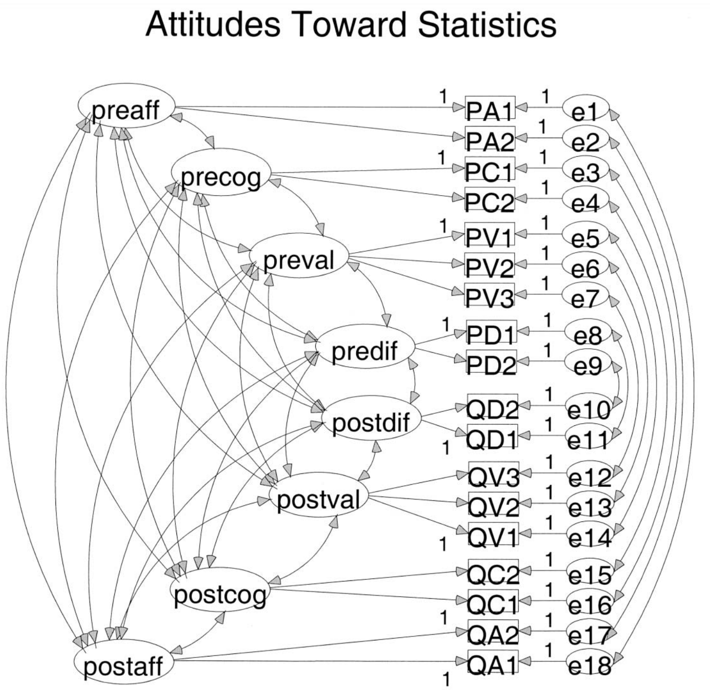

FIGURE 1 Path diagram of the model for a single gender. preaff = presemester affect; precog = presemester cognitive competence; preval = presemester value; predif = presemester difficulty; postdif = postsemester difficulty; postval = postsemester value; postcog = postsemester cognitive competence; postaff = postsemester affect; PA1 = parcel 1 for presemester affect and QA1 = parcel 1 for postsemester affact. The other P* and Q* abbreviations have similar meanings. e1 … e18 are the error terms for the 18 parcels.

proach, we first tested the invariance of factor loadings, then factor variances, and finally, factor correlations. All tests were conducted on the appropriate covariance matrices using maximum likelihood estimation in LISREL 8.14.

When we simultaneously constrained variances to be equal across gender and time, we encountered convergence problems using LISREL. The reason for these problems is not clear; however, by reparameterizing the factor variances as standard deviations, convergence was achieved. Because this reparameterization does not impact the maximized likelihood, all reported results and discussions are in terms of factor variances. In addition, we replicated these results using AMOS 4.0 with constrained factor variances.

We tested the invariance of correlations rather than covariances because correlations are the parameters of interest to most researchers, and tests of covariances are confounded tests of variances and correlations. All analyses were conducted using the sample covariance matrix because it has been shown that analyzing the correlation matrix with most structural equation modeling (SEM) software (such as LISREL) can produce incorrect standard errors (Cudeck, 1989). Because correlations are a function of covariances and variances, we constrained correlations to be equal by constraining the appropriate functions of the covariance and variance parameters (Steiger & Browne, 1984). Note, however, that we did not constrain the correlations between the pre- and postsemester factors to be invariant across gender because we felt this was a structural, rather than a measurement, constraint.

Notation. Throughout this article, we use the following notation: Λ_M and Λ_F are the 18×8 matrices of factor loadings for men and women, respectively. The first nine rows of each matrix (Λ_{M1}, Λ_{F1}) correspond to the four presemester factors—Affect (2 parcels), Cognitive Competence (2 parcels), Value (3 parcels), and Difficulty (2 parcels). The last nine rows correspond to these same postsemester factors (Λ_{M2}, Λ_{F2}). Φ_M and Φ_F are the 8×8 variance–covariance matrices of the unobserved factors for men and for women. It is helpful to partition Φ_M into the following four submatrices: Φ_{M1} is the 4×4 variance–covariance matrix associated with the presemester factors for men; Φ_{M2} is the 4×4 variance–covariance matrix associated with the postsemester factors for men; and $\Phi'_{M12} = \Phi_{M21}$ are the 4×4 covariance matrices between the pre- and postsemester factors for men. Φ_F is partitioned in similar fashion. Let R_M and R_F be the 8×8 correlation matrices of the unobserved factors for men and women with partitions similar to Φ_M and Φ_F. Finally, $\Theta_{\delta M}$ and $\Theta_{\delta F}$ are the 18×18 variance–covariance matrices of errors (uniquenesses) associated with the nine pre- and nine postsemester parcels for men and women, respectively.

RESULTS

Based on the invariance test results presented in Table 2, the SATS factor loadings were invariant across gender (Model 1a), time (Model 1b), and both simultaneously (Model 1c). The reference model used to test for factor loading invariance was the four-factor model that imposed no additional constraints other than the number of latent factors (Model 0).

After establishing that the factor loadings were invariant, the constrained model (Model 1c) served as the reference model for testing the invariance structure of the factor variances. The factor variances were essentially invariant across gender (Model 2a) but not across time (Model 2b). To determine which factor variances were statistically different pre- to postsemester, we constrained each one individually and tested for lack of fit (Table 3).

Each postsemester factor variance was statistically significantly larger than its corresponding presemester variance with the exception of the variances for Cognitive Competence in men (Model 2c). However, a logical problem results when one simultaneously imposes a constraint that *all* factor variances be invariant across gender (results from the test of Model 2a) and a constraint that only the male Cognitive Competence variances be invariant across time (results from the test of Model 2c). Further testing revealed that, for the Cognitive Competence factor,

TABLE 2
Structural Invariance Test Results

Model	*Factor Loadings*	*Factor Variances*	*Within-Factor Correlations*	*df*	χ^2	*Reference Model*	$\Delta \chi^2$	Δ *df*	*p*
0.	Unconstrained			196	370.45				
1a.	$\Lambda_{M1} = \Lambda_{F1}, \Lambda_{M2} = \Lambda_{F2}$			206	381.23	0	10.78	10	.375
1b.	$\Lambda_{M1} = \Lambda_{M2}, \Lambda_{F1} = \Lambda_{F2}$			206	389.16	0	18.71	10	.044
1c.	$\Lambda_{M1} = \Lambda_{F1} = \Lambda_{M2} = \Lambda_{F2}$			211	394.52	0	24.07	15	.064
2a.	$\Lambda_{M1} = \Lambda_{F1} = \Lambda_{M2} = \Lambda_{F2}$	$\Phi_{M1} = \Phi_{F1}, \Phi_{M2} = \Phi_{F2}$		219	411.79	1c	17.27	8	.027
2b.	$\Lambda_{M1} = \Lambda_{F1} = \Lambda_{M2} = \Lambda_{F2}$	$\Phi_{M1} = \Phi_{M2}, \Phi_{F1} = \Phi_{F2}$		219	474.06	1c	79.54	8	.000
2c.	$\Lambda_{M1} = \Lambda_{F1} = \Lambda_{M2} = \Lambda_{F2}$	$\Phi_{M1} = \Phi_{M2}, \Phi_{F1} = \Phi_{F2}$[a]		212	395.40	1c	0.88	1	.348
2d.	$\Lambda_{M1} = \Lambda_{F1} = \Lambda_{M2} = \Lambda_{F2}$	$\Phi_{M1} = \Phi_{F1}, \Phi_{M2} = \Phi_{F2}$[b]		219	405.70	1c	11.18	8	.192
3a.	$\Lambda_{M1} = \Lambda_{F1} = \Lambda_{M2} = \Lambda_{F2}$	$\Phi_{M1} = \Phi_{F1}, \Phi_{M2} = \Phi_{F2}$[b]	$R_{M1} = R_{F1}, R_{M2} = R_{F2}$	231	427.29	2d	21.59	12	.042
3b.	$\Lambda_{M1} = \Lambda_{F1} = \Lambda_{M2} = \Lambda_{F2}$	$\Phi_{M1} = \Phi_{F1}, \Phi_{M2} = \Phi_{F2}$[b]	$R_{M1} = R_{M2}, R_{F1} = R_{F2}$	231	424.66	2d	18.96	12	.090
3c.	$\Lambda_{M1} = \Lambda_{F1} = \Lambda_{M2} = \Lambda_{F2}$	$\Phi_{M1} = \Phi_{F1}, \Phi_{M2} = \Phi_{F2}$[b]	$R_{M1} = R_{F1} = R_{M2} = R_{F2}$	237	435.82	2d	30.12	18	.036

[a]Only male Cognitive Competence variances were constrained to be equal across time.

[b]Cognitive Competence variances were constrained to be equal across time in men but not across gender at postsemester measurement.

TABLE 3
Single Degree of Freedom Invariance Tests for Factor Variances Across Time

Model	*Factor Loadings*	*Factor Variances*	*df*	χ^2	*Reference Model*	$\Delta\chi^2$	Δ *df*	*p*
1c.	$\Lambda_{M1} = \Lambda_{F1} = \Lambda_{M2} = \Lambda_{F2}$		211	394.52				
2b1.	$\Lambda_{M1} = \Lambda_{F1} = \Lambda_{M2} = \Lambda_{F2}$	Male/Affect equal over time	212	410.72	1c	16.20	1	<.0001
2b2.	$\Lambda_{M1} = \Lambda_{F1} = \Lambda_{M2} = \Lambda_{F2}$	Male/Cognitive Competence equal over time	212	395.40	1c	0.88	1	.3482
2b3.	$\Lambda_{M1} = \Lambda_{F1} = \Lambda_{M2} = \Lambda_{F2}$	Male/Value equal over time	212	416.79	1c	22.27	1	<.0001
2b4.	$\Lambda_{M1} = \Lambda_{F1} = \Lambda_{M2} = \Lambda_{F2}$	Male/Difficulty equal over time	212	400.29	1c	5.77	1	.0163
2b5.	$\Lambda_{M1} = \Lambda_{F1} = \Lambda_{M2} = \Lambda_{F2}$	Female/Affect equal over time	212	403.27	1c	8.75	1	.0031
2b6.	$\Lambda_{M1} = \Lambda_{F1} = \Lambda_{M2} = \Lambda_{F2}$	Female/Cognitive Competence equal over time	212	409.58	1c	15.06	1	<.0001
2b7.	$\Lambda_{M1} = \Lambda_{F1} = \Lambda_{M2} = \Lambda_{F2}$	Female/Value equal over time	212	409.55	1c	15.03	1	<.0001
2b8.	$\Lambda_{M1} = \Lambda_{F1} = \Lambda_{M2} = \Lambda_{F2}$	Female/Difficulty equal over time	212	422.54	1c	28.02	1	<.0001

male pre-, male post-, and female presemester variances were equal, whereas the female postsemester variance was significantly larger than the others (Model 2d). Therefore, with the exception of this pattern for the Cognitive Competence factor, the factor variances were invariant across gender but not across time.

The reference model used for testing factor correlation invariance was one that constrained factor loadings to be invariant and factor variances to be invariant across gender but not time, except for Cognitive Competence as explained earlier (Model 2d). The SATS factor correlations were invariant across gender (Model 3a), time (Model 3b), and both gender and time simultaneously (Model 3c).

Model 3c from Table 2 formed the basis for the CFA using all of the remaining data (4,343 individuals with complete pre- and postsemester responses). The resulting chi-square was 1279.55 with 237 *df.* Other goodness-of-fit measures were root mean square error of approximation (0.045), Goodness-of-Fit Index (0.970), Nonnormed Fit Index (0.980), and comparative fit index (0.980). Each of these goodness-of-fit measures indicates that Model 3c fits the data well.

Parameter estimates from the CFA are presented in Tables 4 through 6. The within-group completely standardized factor loadings are presented in Table 4. Although the factor loadings were constrained to be invariant across gender and time, the completely standardized estimates were not identical across gender and time because of the differences in the factor variances. Table 5 presents the maximum likelihood estimates of the factor variances. With the exception of Cognitive Competence in men, the postsemester factor variances were larger than the presemester variances. The smallest increases in factor variances from pre- to postsemester were about one third (Cognitive Competence in women and Affect), and the largest increase was over 90% for the Difficulty factor. Estimates of the factor correlations are presented in Table 6. Affect and Cognitive Competence were strongly related, whereas the relation between Value and Difficulty was small.

DISCUSSION

As others have noted (e.g., Marsh, Hau, Chung, & Siu, 1998), there is not just one hierarchy for testing the invariance across two variables. Factorial invariance is often considered to be the most important aspect of invariance. This aspect of structural equivalence should be established across groups and time before testing for group or time mean differences, as so often occurs in the research on attitudes toward statistics (e.g., Marsh et al., 1998; Rensvold & Cheung, 1998). Therefore, we first tested for factorial invariance. The SATS was factorially invariant across gender, administration time, and both simultaneously. These invariance results suggest that mean differences by gender, by administration time, and by their interaction in the SATS' components can be examined meaningfully. Our gender factorial invariance findings replicate those reported by Dauphinee et al. (1997) for their students; they did not examine administration time invariance.

TABLE 4
Within-Group Completely Standardized Factor Loading Estimates From Confirmatory Factor Analysis

	Males		*Females*	
Parcels	*Pre*	*Post*	*Pre*	*Post*
A_1	0.83	0.81	0.84	0.82
A_2	0.91	0.92	0.90	0.90
C_1	0.84	0.77	0.83	0.81
C_2	0.91	0.86	0.90	0.89
V_1	0.77	0.80	0.76	0.79
V_2	0.88	0.89	0.87	0.89
V_3	0.89	0.92	0.86	0.90
D_1	0.69	0.73	0.71	0.73
D_2	0.83	0.88	0.85	0.89

Note. A_i = the *i*th parcel for the Affect factor; C_i = the *i*th parcel for the Cognitive Competence factor; V_i = the *i*th parcel for the Value factor; D_i = the *i*th parcel for the Difficulty factor.

TABLE 5
Factor Variance Estimates by Gender and Administration Time From Confirmatory Factor Analysis

	Males		*Females*	
Factor	*Pre*	*Post*	*Pre*	*Post*
Affect	0.884	1.188	0.884	1.188
Cognitive Competence	0.706	0.706	0.706	0.923
Value	0.563	0.846	0.563	0.846
Difficulty	0.303	0.578	0.303	0.578

Second, factor variances were tested for invariance. With one exception, factor variances were gender invariant. This finding generally supports that reported by Dauphinee et al. (1997), although they reported one small source of gender noninvariance in Value variances. With one exception, we found that factor variances were not invariant over administration time. Three of the four postsemester factor variances were larger than their corresponding presemester variances; factor variances increased from one third to 90% across the semester. That is, students exhibited greater variability in their attitudes toward statistics after experiencing the course than they did early in the course. This finding was expected because variances usually increase across a period of growth as should occur in introductory statistics courses. Researchers exploring presemester to postsemester differences in attitudes using the SATS should be aware that the component variances are likely to be heterogeneous across these two administration times.

TABLE 6
Factor Correlation Estimates From Confirmatory Factor Analysis

	Affect	*Cognitive Competence*	*Value*	*Difficulty*
Affect	1.00			
Cognitive Competence	0.94	1.00		
Value	0.56	0.55	1.00	
Difficulty	0.72	0.70	0.29	1.00

The exception in both the gender and time invariance results involved the postsemester Cognitive Competence variance in men. This variance, and only this variance, did not increase from presemester to postsemester. The time invariance of the Cognitive Competence variance in men is puzzling because there is no theoretical reason to expect it. No one else has explored the invariance of factor variances across time, much less across time and gender, for any statistics attitude survey; therefore, no other research results in this area exist. Perhaps this finding is unique to our student population or to this sample; perhaps it is not. Further research is needed to clarify this finding.

Third, factor correlations were tested for invariance. SATS factor correlations were invariant across gender, administration time, and both simultaneously. This result was similar to the gender invariance results reported by Dauphinee et al. (1997); they, however, did report that the correlation between Value and Affect was somewhat larger for men than for women with their presemester data.

In our results, the correlation between Affect and Cognitive Competence was 0.94. Schau et al. (1995) and Dauphinee et al. (1997), using a subsample from the Schau et al. (1995) sample, reported correlation values of 0.92 and 0.94, respectively, between these two latent attitudes. Sorge and Schau (2002), using a different sample, reported a value of 0.91. All of these values are large and suggest a possible redundancy in these two components of statistics attitudes. Also, correlations this high can cause collinearity problems in the SEM analyses; however, we saw no evidence of such problems in our analyses.

We believe that these two components represent distinct, but highly related, attitude constructs for three reasons. First, Dauphinee et al. (1997) showed a better fit for the four-factor model of the SATS that includes Affect and Cognitive Competence as separate components than the three-factor model that combines them. Second, these two components of statistics attitudes appear to operate differently in their relations with other variables. For example, Del Vecchio (1994) found that Cognitive Competence was strongly and positively related to successful completion of an undergraduate introductory statistics course for both women and men; Affect was not related to course completion for women and was only weakly related for men. In their hierarchical regression models, Schutz, Drogosz, White, and Distefano (1999) reported that Affect was unrelated to graduate student grade in an introductory statistics course unless a measure of confidence regarding the student's ability to master statistics (an aspect of Cognitive Competence) was included in the model. Third, affective measures like the Affect scale of the SATS and self-efficacy–confidence measures like the Cognitive Competence scale of the SATS often are treated as distinct constructs in educational theory (e.g., see Eccles & Wigfield's, 1995, application of expectancy-value models of behavior to mathematics education) and in research on attitudes toward statistics (e.g., Harlow et al., 2002).

We did not test the uniquenesses for invariance for two reasons. First, uniquenesses often are not invariant but are likely to be specific to the group or time being tested (e.g., Marsh et al., 1998). Second, researchers using attitude surveys are rarely interested in group or time differences in uniquenesses.

There certainly are other reasonable hierarchical orders for invariance testing of two variables, and the chosen hierarchy may influence the results due to the use of different reference models as one moves through the testing hierarchy. We feel that the hierarchy used in our analysis is sensible.

Overall, the final model fit the SATS data well. Factor loadings and factor correlations were completely invariant by gender and administration time. The major source of invariance in the SATS was factor variances. Postcourse factor variances are clearly larger than precourse variances for women for all four factors and for men for three factors.

The next step in our research is to address the issue of stability and change in student attitudes toward statistics. This issue requires working with the mean structure as well as the covariance structure of the SATS. Because the SATS data used in this study is linked with student performance measures and other student characteristics, we can investigate relations between attitudes and performances, as well as academic and demographic predictors of initial attitudes and changes in attitudes over the semester. Our large and growing database will permit us to examine potential instructor and class effects using multiple group and multilevel SEM approaches. Finally, because our work, as well as prior work, on the SATS has used parceling, we will empirically pursue some of the issues concerning parceling raised by Bandalos (2002) and Little, Cunningham, Shahar, and Widaman (2002).

REFERENCES

Atkinson, J. W. (1957). Motivational determinants of risk-taking behavior. *Psychological Review 64,* 359–372.

Auzmendi, E. (1991, April). *Factors related to statistics: A study with a Spanish sample.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL.

Bandalos, D. L. (2002). The effects of item parceling on goodness-of-fit and parameter estimate bias in structural equation modeling. *Structural Equation Modeling, 9,* 78–102.

Bandura, A. (1977). *Social learning theory.* Engelwood Cliffs, NJ: Prentice Hall.

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,* 383–409.

Cherian,V. I., & Glencross, M. J. (1997). Sex, socioeconomic status, and attitude toward applied statistics among postgraduate education students. *Journal of the Royal Anthropological Institute 3,* 1385–1386.

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

Cudeck, R. (1989). Analysis of corrrelation matrices using covariance structure models. *Psychological Bulletin, 105,* 317–327.

Dauphinee, T. L., Schau, C., & Stevens, J. J. (1997). Survey of Attitudes Toward Statistics: Factor structure and factorial invariance for females and males. *Structural Equation Modeling, 4,* 129–141.

Del Vecchio, A. M. (1994). *A psychological model of statistics course completion.* Unpublished doctoral dissertation, University of New Mexico, Albuquerque.

Eccles, J. S., & Wigfield, A. (1995). In the mind of the actor: The structure of adolescents' achievement task value and expectancy-related beliefs. *Personality and Social Psychology Bulletin, 21,* 215–225.

Faghihi, F., & Rakow, E. A. (1995, November). *The relationship of instructional methods with student responses to the survey of attitudes toward statistics.* Paper presented at the Annual Meeting of the Mid-South Educational Research Association, Biloxi, MS.

Gal, 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). The Netherlands: IOS.

Garfield, J., Hogg, B., Schau, C., & Whittinghill, D. (2000, August). *Best practices in introductory statistics.* Paper presented at the American Statistical Association's Undergraduate Statistics Education Initiative, Indianapolis, IN.

Green, K. E. (1993, April). *Affective, evaluative, and behavioral components of attitudes toward statistics.* Paper presented at the annual meeting of the American Educational Research Association, Atlanta, GA.

Gribbons, B. C., & Hocevar, D. (1998). Levels of aggregation in higher level confirmatory factor analysis: Application for academic self-concept. *Structural Equation Modeling, 5,* 377–390.

Harlow, L. L., Burkholder, G. J., & Morrow, J. A. (2002). Evaluating attitudes, skill, and performance in a learning-enchanced quantitative methods course: A structural modeling approach. *Structural Equation Modeling, 9,* 413–430.

Harris, M. B., & Schau, C. (1999). Successful strategies for teaching statistics. In S. N. Davis, M. Crawford, & J. Sebrechts (Eds.), *Coming into her own: Educational success in girls and women* (pp. 193–210). San Francisco: Jossey-Bass.

Katz, B. M., & Tomazic, T. J. (1998). Reducing statistical anxiety through a nonquantitative approach to advanced statistical techniques. *American Statistical Association Proceedings of the Section on Statistical Education,* 222–227.

Little, T. D., Cunningham, W. A., Shahar, G., & Widaman, K. F. (2002). To parcel or not to parcel: Exploring the question, weighing the merits. *Structural Equation Modeling, 9,* 151–173.

Loftsgaarden, D. O., & Watkins, A. E. (1998). Statistics teaching in colleges and universities: Course, instructors, and degrees in fall 1995. *American Statistician, 52,* 308–314.

Maehr, M. (1984). Meaning and motivation: Toward a theory of personal investment. In R. Ames & C. Ames (Eds.), *Research on motivation in education, Vol 1: Student motivation* (pp. 115–144). Orlando, FL: Academic.

Marsh, H. W., Hau, K., Chung, C., & Siu, T. L. P. (1998). Confirmatory factor analyses of Chinese students' evaluations of university teaching. *Structural Equation Modeling, 5,* 143–164.

McCall, C., Belli, G., & Madjidi, F. (1990, August). *The complexities of teaching graduate students in educational administration introductory statistical concepts.* Paper presented at the International Conferences on Teaching Statistics (ICOTS–3), Dunedin, New Zealand.

Perney, J., & Ravid, R. (1990, April). *The relationship between attitudes toward statistics, math self-concept, test anxiety and graduate students' achievement in an introductory statistics course.* Paper presented at the Annual Meeting of the American Educational Research Association, Boston, MA.

Raykov, T. (1997). Scale reliability, Cronbach's coeffeicent alpha, and violations of essential tau-equivalence with fixed congeneric components. *Multivariate Behavioral Research, 32,* 329–354.

Rensvold, R. B., & Cheung, G. W. (1998). Testing measurement models for factorial invariance: A systematic approach. *Educational and Psychological Measurement, 58,* 1017–1034.

Rhoads, T. R., & Hubele, N. F. (2000). Student attitudes toward statistics before and after a computer-integrated introductory statistics course. *IEEE Transactions in Education, 43,* 182–187.

Roberts, D. M., & Bilderback, E. W. (1980). Reliability and validity of a statistics attitude survey. *Educational and Psychological Measurement, 40,* 235–238.

Roberts, D. M., & Saxe, J. E. (1982). Validity of a statistics attitude survey: A follow-up study. *Educational and Psychological Measurement, 42,* 907–912.

Schau, C. G. (2000, November). *Teaching and learning in statistics courses: Importance of students' attitudes and beliefs.* Paper presented to the faculty at Arizona State University.

Schau, C. G., Dauphinee, T., & Del Vecchio, A. (1992, April). *The development of the Survey of Attitudes Toward Statistics.* Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Schau, C. G., 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,* 868–875.

Schutz, P. A., Drogosz, L. M., White, V. E., & Distefano, C. (1999). Prior knowledge, attitude, and strategy use in an introduction to statistics course. *Learning and Individual Differences, 10,* 291–308.

Shultz, K. S., & Koshino, H. A. (1998). Evidence of reliability and validity for Wise's Attitude Toward Statistics scale. *Psychological Reports, 82,* 27–31.

Sorge, C., & Schau, C. (2002, April). *Impact of engineering students' attitudes on achievement in statistics.* Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA.

Sorge, C., Schau, C., Hubele, N., & Kennedy, D. (2000, August). *Measuring engineering students' attitudes toward statistics and probability using Internet data collection techniques.* Paper presented at the Joint Statistical Meetings, Indianapolis, IN.

Steiger, J. H., & Browne, M. W. (1984). The comparison of interdependent correlations between optimal linear composites. *Psychometrika, 49,* 11–24.

Sutarso, T. (1992, November). *Students' Attitudes Toward Statistics (STATS).* Paper presented at the Mid-South Educational Research Association, Knoxville, TN.

Tomazic, T. J., & Katz, B. M. (1988). Statistical anxiety in introductory applied statistics. *American Statistical Association Proceedings of the Section on Statistical Education,* 239–244.

Waters, L. K., Martelli, A., Zakrajsek, T., & Popovich, P. M. (1988). Attitudes toward statistics: An evaluation of multiple measures. *Educational and Psychological Measurement, 48,* 513–516.

Weiner, B. (1979). A theory of motivation for some classroom experiences. *Journal of Educational Psychology, 71,* 3–25.

Wise, S. L. (1985). The development and validation of a scale measuring attitudes toward statistics. *Educational and Psychological Measurement, 45,* 401–405.

Wisenbaker, J. M., & Scott, J. S. (1997, April). *Modeling aspects of students' attitudes and achievement in introductory statistics courses.* Paper presented at the Annual Meeting of the American Educational Research Association, Chicago, IL.

Zanakis, S. H., & Valenzi, E. R. (1997). Student anxiety and attitudes in business statistics. *Journal of Education for Business, 73*(1), 10–16.

Zeidner, M. (1991). Statistics and mathematics anxiety in social science students: Some interesting parallels. *British Journal of Educational Psychology, 61,* 319–332.

APPENDIX A
Male Covariance Matrix Analyzed

	PA_1	PA_2	PC_1	PC_2	PV_1	PV_2	PV_3	PD_1	PD_2	QA_1	QA_2	QC_1	QC_2	QV_1	QV_2	QV_3	QD_1	QD_2
PA_1	0.98																	
PA_2	0.66	0.99																
PC_1	0.59	0.69	0.98															
PC_2	0.61	0.71	0.71	0.97														
PV_1	0.41	0.41	0.37	0.45	0.89													
PV_2	0.37	0.35	0.44	0.45	0.62	0.94												
PV_3	0.45	0.43	0.42	0.50	0.66	0.75	1.05											
PD_1	0.33	0.32	0.24	0.33	0.12	0.09	0.12	0.63										
PD_2	0.31	0.32	0.26	0.31	0.10	0.06	0.11	0.36	0.49									
QA_1	0.52	0.48	0.48	0.45	0.31	0.30	0.33	0.17	0.20	1.62								
QA_2	0.44	0.47	0.42	0.39	0.31	0.24	0.26	0.14	0.17	1.15	1.46							
QC_1	0.33	0.40	0.46	0.44	0.28	0.17	0.24	0.16	0.09	0.73	0.75	1.08						
QC_2	0.35	0.31	0.37	0.35	0.33	0.18	0.23	0.07	0.10	0.90	0.98	0.75	1.23					
QV_1	0.16	0.17	0.24	0.19	0.38	0.31	0.28	–0.05	0.01	0.68	0.75	0.40	0.64	1.31				
QV_2	0.26	0.28	0.32	0.27	0.39	0.45	0.39	0.03	0.04	0.86	0.89	0.47	0.72	1.13	1.66			
QV_3	0.20	0.25	0.29	0.26	0.46	0.43	0.46	–0.04	–0.01	0.78	0.87	0.47	0.68	1.12	1.37	1.55		
QD_1	0.31	0.30	0.20	0.31	0.16	0.10	0.17	0.28	0.21	0.68	0.50	0.32	0.35	0.16	0.25	0.19	0.99	
QD_2	0.31	0.23	0.25	0.23	0.10	0.07	0.06	0.14	0.20	0.67	0.59	0.42	0.52	0.26	0.28	0.25	0.51	0.95

Note. P = presemester parcels; Q = postsemester parcels; A_i = the *i*th parcel for the Affect factor; C_i = the *i*th parcel for the Cognitive Competence factor; V_i = the *i*th parcel for the Value factor; D_i = the *i*th parcel for the Difficulty factor.

APPENDIX B
Female Covariance Matrix Analyzed

	PA_1	PA_2	PC_1	PC_2	PV_1	PV_2	PV_3	PD_1	PD_2	QA_1	QA_2	QC_1	QC_2	QV_1	QV_2	QV_3	QD_1	QD_2
PA_1	1.12																	
PA_2	0.83	1.15																
PC_1	0.65	0.67	0.90															
PC_2	0.70	0.71	0.66	1.03														
PV_1	0.43	0.46	0.39	0.45	0.93													
PV_2	0.44	0.41	0.39	0.37	0.57	0.99												
PV_3	0.45	0.47	0.42	0.43	0.67	0.84	1.18											
PD_1	0.37	0.41	0.26	0.32	0.20	0.12	0.16	0.47										
PD_2	0.37	0.35	0.27	0.31	0.15	0.12	0.14	0.28	0.46									
QA_1	0.49	0.44	0.37	0.40	0.26	0.36	0.31	0.21	0.11	1.64								
QA_2	0.55	0.55	0.50	0.42	0.32	0.44	0.42	0.18	0.15	1.19	1.54							
QC_1	0.49	0.46	0.49	0.41	0.25	0.30	0.25	0.18	0.19	0.85	0.98	1.22						
QC_2	0.55	0.52	0.52	0.49	0.35	0.39	0.37	0.24	0.19	1.12	1.22	1.06	1.58					
QV_1	0.26	0.30	0.21	0.24	0.50	0.42	0.45	0.04	0.05	0.52	0.53	0.45	0.70	1.28				
QV_2	0.35	0.28	0.29	0.25	0.48	0.56	0.55	0.07	0.03	0.70	0.72	0.60	0.80	1.01	1.34			
QV_3	0.28	0.24	0.22	0.25	0.51	0.56	0.64	0.05	0.00	0.87	0.81	0.59	0.86	1.08	1.26	1.70		
QD_1	0.20	0.24	0.20	0.21	0.15	0.12	0.16	0.21	0.15	0.69	0.57	0.41	0.59	0.23	0.30	0.36	1.01	
QD_2	0.26	0.25	0.26	0.27	0.21	0.22	0.18	0.16	0.15	0.71	0.75	0.59	0.82	0.34	0.39	0.46	0.62	1.02

Note. P = presemester parcels; Q = postsemester parcels; A_i = the *i*th parcel for the Affect factor; C_i = the *i*th parcel for the Cognitive Competence factor; V_i = the *i*th parcel for the Value factor; D_i = the *i*th parcel for the Difficulty factor.