MEASURING STATISTICS ATTITUDES: STRUCTURE OF THE SURVEY OF ATTITUDES TOWARD STATISTICS (SATS-36)

STIJN VANHOOF
Katholieke Universiteit Leuven
stijn_vanhoof@hotmail.com

SOFIE KUPPENS
Katholieke Universiteit Leuven
sofie.kuppens@ped.kuleuven.be

ANA ELISA CASTRO SOTOS
Katholieke Universiteit Leuven
anaelisa.castrosotos@bnpparibasfortis.com

LIEVEN VERSCHAFFEL
Katholieke Universiteit Leuven
lieven.verschaffel@ped.kuleuven.be

PATRICK ONGHENA
Katholieke Universiteit Leuven
patrick.onghena@ped.kuleuven.be

ABSTRACT

Although a number of instruments for assessing attitudes toward statistics have been developed, several questions with regard to the structure and item functioning remain unresolved. In this study, the structure of the Survey of Attitudes Toward Statistics (SATS-36), a widely used questionnaire to measure six aspects of students' attitudes toward statistics, is investigated. This study addresses the previously unexplored issue of individual item functioning. Based on confirmatory factor analysis using individual items, the results suggest that the SATS-36 can be improved by removing some poorly functioning items and that depending on the goals of a specific study either six subscales could be used or three of them (Affect, Cognitive Competence, and Difficulty) can be combined into one subscale without losing much information.

Keywords: *Statistics education research; Attitudes towards statistics; Assessment instrument; Confirmatory factor analysis*

1. INTRODUCTION

In recent years, attitudes toward statistics have received increasing attention in statistics education. In statistics education research, attitudes toward statistics are usually broadly defined as a multidimensional concept referring to distinct, but related dispositions pertaining to favourable or unfavourable responses with regard to statistics and statistics learning (Chiesi & Primi, 2009; Gal, Ginsburg, & Schau, 1997; Schau,

Statistics Education Research Journal, 10(1), 35-51, http://www.stat.auckland.ac.nz/serj

Stevens, Dauphinee, & Del Vecchio, 1995). The importance of attitudes in the context of introductory statistics courses is widely recognized (e.g., Gal et al., 1997; Leong, 2006). Students in the social and behavioral sciences tend to experience a course in statistics as intimidating and/or feel insufficiently competent in order to acquire the necessary conceptual understanding (Finney & Schraw, 2003). Such negative attitudes are often considered a major obstacle for effective learning (Waters, Martelli, Zakrajsek, & Popovich, 1988).

In research and in practice it is important to assess dimensions of students' attitudes regarding statistics. Attitude assessment can be used to provide information to students and instructors, or to evaluate the effectiveness of different curricula or didactical approaches. For instance, researchers and teachers believe that if teachers choose challenging activities that promote investigation and are clearly related to everyday life, this can have a positive impact on students' beliefs and attitudes toward statistics (Carnell, 2008; Keeler & Steinhorst, 1995; Leong, 2006; Mills, 2004; Shultz & Koshino, 1998; Suanpang, Petocz, & Kalceff, 2004). Furthermore, attitude information can specifically be used to identify students who are potentially at risk for failing the statistics course. Such identification may be the primary step in assisting them to become successful (Cashin & Elmore, 2005; Roberts & Saxe, 1982).

Evaluation of attitudes toward statistics and their associations with other variables is only possible if proper assessment instruments are available. Such work has already been initiated by a number of researchers (e.g., Roberts & Bilderback, 1980; Schau et al., 1995; Wise, 1985). To improve teaching practice, however, constructing an instrument is not enough. In order to correctly evaluate students' attitudes toward statistics it is essential to evaluate the reliability, validity, and possible pitfalls or flaws of an instrument. More specifically, evidence that the presupposed structure of the instrument demonstrates an acceptable fit to the data and that all items measure the underlying constructs of interest should be gathered (Hatcher, 1994). If such evidence is not available, results cannot be unambiguously interpreted.

Although some studies have already been conducted to investigate the structure of the Survey of Attitudes Toward Statistics (SATS) (Cashin & Elmore, 2005; Chiesi & Primi, 2009; Dauphinee, Schau, & Stevens, 1997; Hilton, Schau, & Olsen, 2004; Schau et al., 1995; Tempelaar, van der Loeff, & Gijselaers, 2007), further clarification and investigation is needed. In particular, there is some disagreement on the number of factors that best represent the attitudes toward statistics assessed via the SATS. Furthermore, previous studies are limited because combinations of items ("parcels"; see Section 2) rather than individual items have been used. Therefore, these studies only provide partial insight into the underlying structure of the SATS and the functioning of individual SATS-items.

The goal of the present study is to address some of the above-mentioned remaining issues of previous research on the structure of the Survey of Attitudes Toward Statistics. Specifically, the structure will be investigated using information from individual items rather than parcels. Furthermore, the relative merits of two previously identified factor models (a four- and a six-factor model) will be compared.

This paper is organized in five sections. In Section 2 we briefly outline the SATS and then introduce the research goals in detail and relate them to existing studies on the structure of the SATS. We also discuss advantages and disadvantages of using parcels and items for confirmatory factor analysis. In Section 3 we describe the methodology of this study. In Section 4, we present our results, followed, in Section 5, by a discussion.

2. LITERATURE REVIEW

2.1. THE SURVEY OF ATTITUDES TOWARD STATISTICS (SATS)

The Survey of Attitudes Toward Statistics (Schau et al., 1995) instrument was developed to assess students' attitudes toward statistics. The SATS is a Likert-type instrument with seven response possibilities for each statement ranging from *strongly disagree* to *strongly agree*. The first version of the SATS consisted of four subscales: (a) *Affect* (six items): positive and negative feelings concerning statistics; (b) *Cognitive Competence* (six items): attitudes about intellectual knowledge and skills applied to statistics; (c) *Difficulty* (seven items): attitudes about the difficulty of statistics as a subject; and (d) *Value* (nine items): attitudes about the usefulness, relevance, and worth of statistics in personal and professional life. Afterwards (Schau, 2003; Schau, personal communication, September 29, 2005), two subscales were added to the instrument: *Interest* (four items), students' level of individual interest in statistics and *Effort* (four items), the amount of effort students spend on learning statistics. Depending on the number of subscales and corresponding items, the developers labeled the instrument as SATS-28 and SATS-36. Additionally, two versions are available; one to administer before (SATS-28-pre/SATS-36-pre) and one to administer after a statistics course (SATS-28-post/ SATS-36-post). These two versions are identical, except for tense.

A complete version of the SATS-36 and detailed scoring information can be consulted online via http://www.evaluationandstatistics.com/index.html.

2.2. STRUCTURE OF THE SATS

In one study on the most recent version, namely SATS-36, the presupposed six-factor solution was supported by confirmatory factor analysis (Tempelaar et al., 2007). However, the factors *Affect* and *Cognitive Competence* appeared to be very strongly correlated, and the *Difficulty* scale was moderately to strongly correlated with these two subscales. These findings coincide with empirical studies on the SATS-28 (Cashin & Elmore, 2005; Chiesi & Primi, 2009; Dauphinee et al., 1997; Hilton et al., 2004; Schau et al., 1995). In the above studies, correlations ranged between 0.80 and 0.94 for *Affect* and *Cognitive Competence*, between 0.57 and 0.73 for *Affect* and *Difficulty*, and between 0.46 and 0.64 for *Cognitive Competence* and *Difficulty*. In all studies except the one by Cashin and Elmore (2005) these related constructs were represented as three distinct latent factors.

Dauphinee et al. (1997) explicitly compared the original four-factor model of the SATS-28 to a three-factor model that combined *Affect* and *Cognitive Competence*. They concluded that the factors should remain distinct because: (1) the four-factor model fit better, (2) the two factors operated differently in terms of course completion, and (3) it is important in statistics education to have a distinct construct (i.e., *Cognitive Competence*) that corresponds with Mathematics Self-Concept in the area of mathematics education.

Conversely, based on one exploratory factor analysis on the SATS-28 conducted by Cashin and Elmore (2005), a more parsimonious solution was suggested with *Affect*, *Cognitive Competence*, and *Difficulty* combined into one factor. They argued that the SATS-28 may only pertain to two underlying factors, namely (1) the value of statistics as a tool in students' respective fields of study (*Value*) and (2) different aspects of how a student will perform in his or her statistics course (measured by the *Affect*, *Cognitive Competence*, and *Difficulty* items). Interestingly, these two dimensions correspond to the two subscales (*Field* and *Course*) of the Attitudes Toward Statistics (ATS) scale (Wise, 1985), another instrument that measures students' attitudes toward statistics. As was the case in the other studies on the structure of the SATS, in the study of Cashin and Elmore (2005) especially, the relationship between *Affect* and *Cognitive Competence* was pronounced and considered meaningful as both subscales relate to feelings concerning the specific course, skills, or personal capabilities to complete the coursework. In contrast to their prior expectations *Difficulty* did not form a separate factor in their exploratory factor analysis. Hence they suggested further research to investigate the factor structure of the SATS and the relationship between *Affect*, *Cognitive Competence*, and *Difficulty* in more detail.

Theoretically, the more parsimonious *Course-Field* difference for the SATS-28 relates to the distinction often made in attitude research between students' attitudes about a specific domain (i.e., the value of statistics) and their attitudes about themselves as learners of a domain (i.e., affect, self-efficacy, and perceived difficulty regarding statistics) (e.g., see Gal & Ginsburg, 1994; McLeod, 1992). For instance, students' attitudes about a specific domain are generally considered resistant to change, whereas their attitudes about themselves as learners of a domain are more likely to change depending on changing circumstances during the progress of students' curriculum (Gal & Ginsburg, 1994).

2.3. PARCEL VERSUS ITEM-LEVEL CONFIRMATORY FACTOR ANALYSIS

As already stated in the introduction, all available confirmatory factor analytic studies of the SATS have performed analyses on item parcels rather than individual items (Chiesi & Primi, 2009; Dauphinee et al., 1997; Hilton et al., 2004; Schau et al., 1995; Tempelaar et al., 2007). An item parcel refers to a simple sum or mean of several items from the same factor. It is assumed that the items of a parcel assess the same construct and that they are psychometrically unidimensional (Bandalos, 2002; Hau & Marsh, 2004; Nasser & Wisenbaker, 2003). In such analyses, parcels are treated as continuous indicators (Kline, 2005).

The main reasons for using the technique of item parceling in the context of SATS data are: "to improve reliability" (Dauphinee et al., 1997, p. 133; Schau et al., 1995, p. 872), "to avoid inherent non-normality associated with single item distributions" (Hilton et al., 2004, p. 97), and "to reduce the number of model parameters to achieve a more attractive variable to sample size ratio, and to get more stable parameter estimates" (Tempelaar et al., 2007, p. 85).

Although the technique of item parceling has its advantages, it remains controversial. Parceling might be seen as "tricky" because modeled data should resemble the observed responses as much as possible (Little, Cunningham, Shahar, & Widaman, 2002). In this sense, parceling introduces a potential source of subjective bias (Little et al., 2002), especially because several methods of parceling are available and the choice of the method can affect the results (Kline, 2005). Furthermore, the assumption of unidimensionality within parcels is often not investigated or even not met (Bandalos & Finney, 2001). If a set of items assigned to the same parcel is not unidimensional, analysis of the aggregate score across the items may be meaningless (Kline, 2005). Moreover, in some instances parceling can mask a multidimensional factor structure which may lead to a seriously misspecified CFA-model fitting the data reasonably well (Bandalos, 2002).

Besides these general disadvantages of the technique of item parceling, specific problems pertain to the parceling schemes used in previous research on the SATS. First, in the parceling scheme of Dauphinee et al. (1997), Hilton et al. (2004) and Schau et al.

(1995) the *Affect*, *Cognitive Competence*, and *Difficulty* factors comprised only two parcels, whereas a minimum of three has been suggested (Hau & Marsh, 2004; Nasser & Wisenbaker, 2003). Second, in the parceling scheme of Tempelaar et al. (2007), there were exactly three parcels per factor, but some parcels necessarily contained only one item because there were only four *Interest* and *Effort* items.

As the technique of item parceling may jeopardize a good understanding of the true factor structure of the SATS-36 items (Bandalos, 2002; Bandalos & Finney, 2001), the present study performed a confirmatory factor analysis on individual items of the SATS-36 for the first time using statistical approaches for categorical item-indicators (see Section 3.3).

Two main research questions on the structure of the SATS-36 will be investigated using confirmatory factor analysis on the individual items:

1. Can the six-factor structure be confirmed for the SATS-36?
2. Is the six-factor structure preferable to a four-factor structure that represents *Affect*, *Cognitive Competence*, and *Difficulty* as a single factor?

The answers to these two questions are important to guide interpretation of the instrument not only in research but also in teaching. For instance, if *Affect*, *Cognitive Competence*, and *Difficulty* measure the same construct or show similar correlations with other SATS-factors, a more parsimonious interpretation of the instrument may be preferred.

3. METHOD

3.1. PARTICIPANTS

Participants of this observational study are 514 first year Educational Sciences (321 female, 22 male) and Speech Pathology and Audiology (163 female, 8 male) students from two cohorts of an introductory undergraduate statistics course at the Department of Educational Sciences of the Katholieke Universiteit Leuven. As the numbers indicate, students in Educational Sciences and Speech Pathology and Audiology are mainly female. Most participants took limited to moderate mathematics-oriented programs in secondary education. Specifically, 154 participants followed programs including one, two, or three weekly hours of mathematics; 334 participants followed programs including four, five, or six weekly hours of mathematics; and 26 students followed programs including seven or eight weekly hours of mathematics. Four hundred sixty-six participants indicated that an introduction to statistics was part of their secondary school education. The introductory statistics course was taught during the first semester of the students' first academic year. In general, the course dealt with some introductory methodological and statistical concepts (such as tables, figures, and descriptive statistics), but not with formal probability theory or statistical inference. The mathematical background required for following the course was limited.

The data were collected at the very beginning of this first year statistics course. For one cohort this occurred in September 2005, for the other cohort in September 2006. The instrument was completed voluntarily and handed in during class time. It was stressed that the data would be analyzed anonymously.

3.2. MEASURES

Attitudes toward statistics are assessed with a Dutch translation of the pre-test version of the SATS-36 (Schau et al., 1995). Negatively formulated items were reversed to assure that a high score equals a positive attitude. Like Tempelaar et al. (2007), the focus of this study is on students' attitudes when entering university.

The translation from English into Dutch took place in August/September 2005, using the following procedure. First, the instrument was translated by the first author and by an expert translator with a Master's degree in German languages. Afterwards, both translations were compared and differences and possible ambiguities were discussed. This resulted in a Dutch version of the (pre-test version of the) SATS-36. Second, this translation was validated using a back-translation technique (Brislin, 1970). Statistics experts translated the items of the Dutch version back into English. Afterwards, the quality of the translation was judged by comparing this version with the original English version. Differences and possible ambiguities were discussed. Third, the Dutch version of the SATS-36 was administrated to a small number of people (n = 6) with a diverse statistical background. Participants were instructed to write down suggestions when filling out the instrument. The comments did not point out any problems with the wording, merely some minor comments on the punctuation, so no changes to the translation of any item were made. The translated Dutch version of the SATS-36 is available on request from the first author.

3.3. STATISTICAL ANALYSES

Confirmatory factor analysis Confirmatory factor analyses (CFA) are performed using the software Lisrel 8.70 (Jöreskog & Sörbom, 2004) to test the proposed factor models for the SATS-36.

The ordinal nature of the SATS-36 items has to be respected and important assumptions (such as multivariate normality) of the techniques used have to be investigated. Two Robust Maximum Likelihood (RML) estimation techniques were performed: RML for ordinal data with polychoric correlations (Jöreskog, 1993) and RML with covariances (e.g., Kline, 2005). RML was preferred to weighted least squares estimation (WLS), because the latter requires a very large sample size (e.g., Finney & DiStefano, 2006; Flora & Curran, 2004; Kline, 2005) and more easily results in convergence problems or improper solutions as model complexity increases (Boomsma & Hoogland, 2001).

RML is preferred to standard ML for both polychoric correlations and covariances to correct for nonnormality in the data. When deviation from multivariate normality is present and data are categorical, parameter estimates, chi-square statistics, and standard errors tend to be biased (Finney & DiStefano, 2006; Kline, 2005). If this is the case, RML, such as using a Satorra-Bentler adjustment of the chi-square statistic, is recommended to adjust estimations based on the degree of nonnormality (Finney & DiStefano, 2006; Kline, 2005; Satorra & Bentler, 1994).

First, polychoric correlations are analyzed. A polychoric correlation estimates what the Pearson correlation coefficient between two ordinal variables would be if both were continuous and normally distributed in the population (Kline, 2005). The estimated underlying continuous population variables are then used in the confirmatory factor models, instead of the observed variables (Jöreskog, 2005).

Second, the covariance matrix is analyzed, treating the observed variables as continuous. It has been shown that such an analysis combined with Satorra-Bentler adjustment produces sufficiently accurate parameter estimates for Likert scales with more than five response categories under conditions of nonnormality (Finney & DiStefano, 2006). Because the SATS-36 is based on a seven-point Likert scale, this technique seems appropriate.

A major advantage of analyses based on covariances as compared to polychoric correlations is that covariance results allow interpretation of factor loadings and other estimates based on actual results or raw data. Therefore, discussion of the parameter estimates will be based on results from analysis of covariances when similar results are produced by both approaches (covariances and polychoric correlations).

Details about data and assumptions Before describing the decision rules and fit statistics used, details about the data and underlying assumptions of the techniques are investigated. When assumptions are violated, biased results may occur in terms of model fit, parameter estimates, and related significance tests (Finney & DiStefano, 2006; Schumacker & Lomax, 2004).

Summary statistics (means, standard deviations, kurtosis, skewness) for the items, polychoric correlations and covariance matrices are available upon request from the first author. Tests of model assumptions (Jöreskog, 1999; SSICentral, n.d.) on the items showed deviations from multivariate normality (Skewness z = 21.749, p < 0.001; Kurtosis z = 399.592, p < 0.001; Skewness and Kurtosis χ^2 = 303.972, p < 0.001). All items, except item 5 (*Cognitive Competence* 1), item 10 (*Value* 3), item 6 (*Difficulty* 1) and item 30 (*Difficulty* 5), show significant deviations from univariate normality. However, when inspecting the size of skewness and kurtosis, only item 27 (*Effort* 4) showed substantive deviation from normality (Skewness = -1.892; Kurtosis = 4.823). Because there were indications of multivariate nonnormality, a Satorra-Bentler adjustment was performed (Satorra & Bentler, 1994).

The assumption of underlying bivariate normality is required to analyze polychoric correlations. Based on the RMSEA-values for population discrepancy (Jöreskog, 2004), no violations of this assumption were observed.

Global model fit Satorra-Bentler-scaled chi-square statistics (SBS χ^2; Satorra & Bentler, 1994) were obtained during the CFA-analyses to assess the magnitude of discrepancy between the sample and fitted matrices (Hu & Bentler, 1999). As mentioned earlier, the Satorra-Bentler-scaled chi-square statistic corrects for nonnormality in the data. However, it is widely known that the chi-square-based statistics are very sensitive to sample size (e.g., Kline, 2005). This may result in the rejection of reasonable models because in the presence of large sample sizes, small degrees of lack of fit already result in small p-values (Byrne, 1989; Hu, Bentler, & Kano, 1992). For this reason additional goodness-of-fit indices were used to evaluate model fit: Root Mean Square Error of Approximation (*RMSEA*), Comparative Fit Index (*CFI*), and Non-normed fit index (*NNFI*). It has been suggested that a value of the *RMSEA* of less than 0.05 is an indication of a good fit whereas values between 0.05 and 0.08 still show a reasonable fit of the model. The indices *NNFI* and *CFI* normally range between zero and one, with higher values indicating a better fit. As a benchmark for good fit, the value 0.90 is often used (Hoyle & Panter, 1995; Hu & Bentler, 1999; Schumacker & Lomax, 2004).

Because one of the goals of this study was to compare the presupposed six-factor structure model of the SATS-36 to a four-factor model where *Affect*, *Cognitive Competence,* and *Difficulty* are combined in one factor, additional fit statistics were inspected to assess the relative fit of these two nested models.

The six- and four-factor models were compared by means of the scaled chi-square difference test (scaled-$\chi^2\Delta$; Satorra & Bentler, 2001). However, because this significance test is sensitive to relatively small deviations when sample size is large, the Akaike Information Criterion (*AIC*; Wang & Liu 2006) and the Bayesian Information Criterion (*BIC*; Wang & Liu 2006) were additionally used to compare models. The *AIC* and *BIC* take into account both model fit and model complexity. When comparing two models, the model with the lowest *AIC* and *BIC* is the preferred one (Jöreskog, 1993; Kline, 2005).

Local model fit Because this is the first study that examined the factor structure of the SATS using confirmatory factor analysis on the individual items, we considered evaluating local model fit to be important. It is possible for a model to be misspecified in some parts but very well specified in other parts (Jöreskog, 1993). For the presented models, size and significance of the factor loadings, standardized residuals, and modification indices are reviewed (e.g., Kline, 2005). Also correlations between latent factors are checked and discussed especially to investigate whether *Affect*, *Cognitive Competence*, and *Difficulty* show similar or different correlations with the other SATS-36 factors.

Although standardized residuals show which relationships are not properly explained, they do not indicate how the model should be modified to fit the data better (e.g., Kline, 2005). Therefore, modification indices rather than standardized residuals were used to guide model modification. Clear guidelines or cut-off values regarding modification indices are not available. The best option is to initially consider the modification indices with the highest values. In addition, substantive and theoretical arguments were used to guide the modification of the models in order to avoid the risk of capitalizing on chance and building models that do not generalize to Educational Sciences and Speech Pathology groups at other countries/universities (Hatcher, 1994; Hoyle & Panter, 1995; Jöreskog, 1993; MacCallum, Roznowski, & Necowitz, 1992; Schumacker & Lomax, 2004).

Items with factor loadings below 0.40 (e.g., Hatcher, 1994) were considered for deletion from the SATS-36, because such items may not sufficiently relate to the expected underlying construct. As will be discussed in detail later, cross-validation of such modifications is needed.

4. RESULTS

As the results from polychoric correlations corroborated the results from covariances, only the latter results are presented. Details of the results from polychoric correlations are available upon request from the first author. Tables 1 to 4 present information on the hypothesized six-factor and four-factor models as well as on the modified models that resulted from consideration of the model characteristics. More information about the modified models will be presented after describing the original models.

4.1. SIX-FACTOR SOLUTION

As presented in Table 1, adequate fit indices were obtained for the hypothesized six-factor SATS-36 model (Model 1). Associations among the latent factors for Model 1 are shown in Table 2. Note that the *Difficulty* factor should be interpreted as perceived easiness, because high scores represent a positive attitude which explains the negative correlation with the *Effort* factor (see further). As reported in Table 2, *Affect*, *Cognitive Competence*, and *Difficulty* were highly correlated.

Table 1. Fit indices for the models tested based on the covariance matrix

Model	*SBS* χ^2	*df*	*RMSEA*	*NNFI*	*CFI*	*BIC*	*AIC*
6-factor original (1)	1607.3	579	0.059	0.94	0.95	2150.4	1781.3
6-factor modified (2)	1136.2	479	0.052	0.96	0.96	1648.1	1300.2
4-factor original (3)	1655.6	588	0.060	0.94	0.94	2142.4	1811.6
4-factor modified (4)	1209.5	488	0.054	0.96	0.96	1665.1	1355.5

Note. SBS χ^2 = Satorra-Bentler-scaled chi-square; *df* = Degrees of Freedom; *RMSEA* = Root Mean Square Error of Approximation; *NNFI* = Non-Normed Fit Index; *CFI* = Comparative Fit Index; *BIC* = Bayesian Information Criterion; *AIC* = Akaike Information Criterion

Table 2. Estimated latent factor correlations for the six-factor models

Model 1: 6-factor original	*Affect*	*Cognitive Competence*	*Difficulty*	*Value*	*Interest*
Cognitive Competence	0.888***				
Difficulty	0.844***	0.855***			
Value	0.442***	0.431***	0.370***		
Interest	0.575***	0.484***	0.476***	0.715***	
Effort	-0.088	-0.120*	-0.221***	0.165**	0.200***

Model 2: 6-factor modified	*Affect*	*Cognitive Competence*	*Difficulty*	*Value*	*Interest*
Cognitive Competence	0.883***				
Difficulty	0.848***	0.860***			
Value	0.393***	0.432***	0.355***		
Interest	0.484***	0.483***	0.470***	0.715***	
Effort	-0.135*	-0.119*	-0.232***	0.164**	0.201***

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

To explore differences between *Affect*, *Cognitive Competence*, and *Difficulty*, we explored their differential associations with other SATS-factors for the original six-factor model. *Affect*, *Cognitive Competence*, and *Difficulty* showed similar correlations with *Value* (resp. $r = 0.44$, $r = 0.43$, $r = 0.37$) and *Interest* (resp. $r = 0.58$, $r = 0.48$, $r = 0.48$). The correlations with *Effort* showed more diversity ($r = -0.09$, $r = -0.12$, $r = -0.22$), with *Difficulty* (perceived easiness) most negatively related to *Effort*. In other words, the more that students perceived statistics to be difficult, the more *Effort* they expected to spend on learning statistics. Students' affect and competence seemed less associated with the expected amount of *Effort* compared to the associations with *Difficulty*, *Value*, and *Interest*. Note that a relatively high correlation was estimated between *Interest* and *Value* ($r = 0.715$, $p < 0.001$). Parameter loading estimates for all items in the original six-factor model are shown in the third column of Table 3. All factor loadings except for three *Difficulty* items were above 0.40 and significant. In a modified six-factor model the three items with factor loadings below 0.40 (item 22, item 34, and item 36) were deleted. Two of these items (item 22 and item 36) ask about most people's attitudes regarding the *Difficulty* of statistics, rather than—like the other *Difficulty* items—students' own attitudes regarding the *Difficulty* of the field of statistics. This may explain the poor functioning of these two items.

Table 3. Parameter estimates of the models tested

		6-factor model	6-factor modified		4-factor model	4-factor modified
Affect	Item 3	0.650	0.569	Integration	0.619	0.584
	Item 4	0.585	0.616		0.583	0.595
	Item 15	0.685	0.714	of	0.670	0.678
	Item 18	0.659	0.690		0.633	0.641
	Item 19	0.624	0.532		0.577	0.535
	Item 28	0.749	0.777		0.738	0.742
				Affect,		
Cognitive	Item 5	0.705	0.711		0.676	0.686
Competence	Item 11	0.544	0.542		0.523	0.522
	Item 26	0.546	0.547		0.541	0.546
	Item 31	0.477	0.478		0.448	0.450
	Item 32	0.608	0.609	*Cognitive*	0.579	0.582
	Item 35	0.827	0.823	*Competence*	0.794	0.796
Difficulty	Item 6	0.615	0.624		0.588	0.57
	Item 8	0.651	0.649		0.582	0.579
	Item 22	0.338	Deleted	and	0.277	Deleted
	Item 24	0.523	0.514		0.456	0.458
	Item 30	0.438	0.426		0.402	0.402
	Item 34	0.246	Deleted		0.248	Deleted
	Item 36	0.341	Deleted	*Difficulty*	0.320	Deleted
Value	Item 7	0.544	0.544	*Value*	0.544	0.544
	Item 9	0.570	0.571		0.568	0.568
	Item 10	0.512	0.513		0.510	0.510
	Item 13	0.501	0.501		0.500	0.500
	Item 16	0.620	0.619		0.621	0.621
	Item 17	0.567	0.567		0.570	0.570
	Item 21	0.424	0.424		0.424	0.424
	Item 25	0.519	0.519		0.517	0.517
	Item 33	0.617	0.617		0.618	0.618
Interest	Item 12	0.592	0.592	*Interest*	0.591	0.591
	Item 20	0.822	0.819		0.820	0.819
	Item 23	0.850	0.853		0.852	0.853
	Item 29	0.844	0.843		0.845	0.844
Effort	Item 1	0.720	0.718	*Effort*	0.720	0.719
	Item 2	0.749	0.750		0.747	0.747
	Item 14	0.783	0.783		0.784	0.784
	Item 27	0.700	0.699		0.700	0.700
Error covariance between item 3 and item 19			0.410			0.400

Besides dropping three *Difficulty* items with low factor loadings, inspection of the standardized residuals and modification indices shows that another substantial justifiable improvement can be made. Including an error covariance between item 3 and item 19 (the first and fifth *Affect* items) resulted in a substantial decrease in chi-square. The existence of this error covariance was not surprising as both items were the only two positively formulated items of the *Affect* factor which also share closely related meanings (i.e., item 3 “I will like statistics”; item 19 “I will enjoy taking statistics courses”). In other words, there seems to be a unique association (method or content similarity) between these two items that is not accounted for by the latent factor.

Consequently, in a modified version of the six-factor model (Model 2) the three *Difficulty* items with low factor loadings were deleted and the error covariance between item 3 and item 19 added to the model. As presented earlier and discussed later, caution is needed when changes like these are made to a presupposed factor structure.

The modified six-factor model (Model 2) provided a considerable better fit to the data compared to the hypothesized six-factor model (Model 1, Table 1), as indicated by a significant scaled-chi-square difference test (scaled-$\Delta\chi^2(100) = 403.23$, $p < 0.001$). There was also a substantial decrease in *BIC* and *AIC* values in favour of the modified model ($\Delta BIC = 502.27$, $\Delta AIC = 481.06$).

In the modified six-factor model all parameter estimates were above 0.40 (see last column of Table 3) and no additional substantial and theoretically plausible changes seemed reasonable based on the modification indices. Inspection of the correlation matrix between the latent factors for Model 2 (Table 2) shows—as was the case for Model 1—high correlations between *Affect*, *Cognitive Competence*, and *Difficulty* (all correlations greater than 0.84). Together with similar results observed in earlier studies (see Section 2) this again suggests that it might be possible to integrate the items of these factors into one factor.

4.2. FOUR-FACTOR SOLUTION

Table 1 also shows the fit indices of the four-factor model where *Affect*, *Cognitive Competence*, and *Difficulty* were combined into one factor (Model 3). As was the case for the six-factor model, the absolute model fit was adequate.

Again, the same three *Difficulty* items (item 22, item 34, and item 36) had factor loadings below 0.40 (Table 3) and modification indices indicated a substantial improvement in fit by allowing an error covariance between item 3 and item 19. In the modified four-factor model (Model 4), the three *Difficulty* items were deleted from the model and the error covariance between item 3 and item 19 was included.

The modified Model 4 performed significantly better than the hypothesized four-factor Model 3 (scaled-Δ (100) = 391.65, $p < 0.001$) and there was a considerable impact on the *BIC* and *AIC* values in favour of Model 4 ($\Delta BIC = 477.29$, $\Delta AIC = 456.08$). The correlation structure of the latent factors for Model 3 and Model 4 are shown in Table 4.

Table 4. Estimated latent factor correlations for the four-factor models

Model 3: 4-factor original	Integration of *Affect*, *Cognitive Competence*, and *Difficulty*	*Value*	*Interest*
Value	0.445***		
Interest	0.540***	0.715***	
Effort	-0.320**	0.164**	0.201***

Model 4: 4-factor modified	Integration of *Affect*, *Cognitive Competence*, and *Difficulty*	*Value*	*Interest*
Value	0.426***		
Interest	0.509***	0.715***	
Effort	-0.147**	0.164**	0.201***

*$p < 0.05$; **$p < 0.01$; ***$p < 0.001$

4.3. FOUR- VERSUS SIX-FACTOR SOLUTION

A formal comparison of the Satorra-Bentler chi-square values of the two original models, Model 1 and Model 3, indicated that the six-factor model performed better (scaled-$\Delta\chi^2$ (9) = 49.15, p < 0.001). The difference in *BIC* was small (lower than 10; Raftery, 1995) and in favour of the four-factor model (Model 3). The difference in *AIC* was also small, but in favour of the six-factor model (Model 1) (ΔBIC = 7.93, ΔAIC = -30.25). When the modified models are compared, the results are all in favour of the six-factor model (scaled-$\Delta\chi^2$ (9) = 70.85, p < 0.001; ΔBIC = -17.05; ΔAIC = -55.23).

The four-factor model was based on high correlations between three factors. To further investigate differences between *Affect*, *Cognitive Competence*, and *Difficulty*, mean observed scores were compared for the modified model. The mean scores were 3.63 (*Affect*), 4.22 (*Cognitive Competence*), and 3.36 (*Difficulty*). Although all differences were smaller than one point on the Likert-scale, paired *t*-tests revealed that all differences were statistically significant (p < 0.001).

Although very high correlations are observed between *Affect*, *Cognitive Competence*, and *Difficulty* in this study, we also tested a five-factor model with only *Affect* and *Cognitive Competence* combined. The reason for doing this was that in earlier studies *Affect* and *Cognitive Competence* were clearly more related than *Affect* and *Difficulty* or *Cognitive Competence* and *Difficulty* (see Section 2.2). Because a very high correlation (r = 0.88) remained between the combined *Affect-Cognitive Competence* and *Difficulty scales*, this solution was not explored further.

5. DISCUSSION

In the following, we first discuss results for the confirmatory factor analyses of the original six-factor model of the SATS-36 and the alternative four-factor model, then we consider conditions when one model is preferred to the other. Lastly we present some limitations of the present study and suggestions for further research.

5.1. SIX- AND FOUR-FACTOR MODEL FOR THE SATS-36

In an absolute sense, the predefined six-factor structure of the SATS-36 could not be falsified in this study, which corroborates Tempelaar et al.'s (2007) study using parcels. In addition, similar to Cashin and Elmore (2005), a four-factor structure which integrated *Affect*, *Cognitive Competence*, and *Difficulty* also showed adequate absolute properties. As such, an important conclusion of this study is that both the six- and four-factor models appropriately describe the observed interrelationships between SATS-36-items. However, a closer investigation of item functioning suggested that some model modifications are in order.

First, as reflected by low factor loadings, several *Difficulty* items (item 22, item 34, and item 36) should be deleted from the instrument because they show less common variance with other *Difficulty* items. For item 22 and item 36 this may relate to the fact that both items refer to how statistics is perceived by most people, whereas other *Difficulty* items pertain to students' attitudes toward the field of statistics as such. The reason for the poor functioning of item 34 remains unclear and should be further investigated.

Second, item 3 and item 19 seemed to have more in common than represented by the *Affect* factor. This unique association likely refers to method and content similarity because they are the only two positively formulated items of the *Affect* factor and share closely related meanings. In this study—to take this unique association and the individual contribution of both items to *Affect* into account—an error covariance between the two items was added to the CFA model. A more practical solution would have been to delete one of these items without much loss of information because of the similarity in meaning (see for example John & Soto, 2007).

Based on these results we argue that—when assessing students' attitudes toward statistics—statistics educators and researchers should take these improvements to the SATS-36 into account. However, caution is needed regarding this conclusion. Results of this observational study might be idiosyncratic and might not necessarily generalize to students from other curricula or Educational Sciences and Speech Pathology groups at other countries/universities (Hoyle & Panter, 1995; MacCallum et al., 1992). Replication and further research is needed. Nevertheless, it seems important that users of the SATS-36 pay explicit attention to the suggested improvements to the instrument.

5.2. COMPARISON OF SIX- AND FOUR-FACTOR MODEL

Results from the explicit comparison of both models were not straightforward. From a technical viewpoint, the original six-factor model demonstrated a better fit than the four-factor model, but the difference was very small when the extra model complexity of the six-factor model was penalized (cf. *BIC* fit index). From a more substantive viewpoint, several reasons can be formulated that favor the original six-factor version of the SATS-36 over the more parsimonious four-factor model. First, it will be easier to compare results to earlier studies. Second, in this study mainly correlations were considered, but it is still possible that mean scores for *Affect*, *Cognitive Competence*, and *Difficulty* differ considerably. In our study, all differences were significant, with a minor difference between *Affect* and *Difficulty* (3.63 versus 3.36) and a substantial difference between *Cognitive Competence* and *Difficulty* (4.22 versus 3.36). Furthermore, it can be important to situate individual students on all subscales; it is possible that an individual student has a relatively low score (for instance compared to the class average) on one of these subscales, and at the same time a relatively high score on another subscale. Third, in line with previous research, differential associations emerged between these three factors and other SATS-factors.

In the absence of a univocal preference for one structure, the choice may depend on the goals of a specific study or specific educational setting. Researchers or educators who require a more global description of students' attitudes toward statistics, may employ the more parsimonious subscale structure when using the SATS-36 (note that this does not mean that students have to answer fewer items). In such case, in line with the distinction formulated by Gal and Ginsburg (1994) and McLeod (1992) (see also Section 2.2), the combination of the *Affect*, *Cognitive Competence*, and *Difficulty* subscales represent students' attitudes about themselves as learners of statistics (or as formulated by Cashin and Elmore (2005): different aspects of how a student will perform in his or her statistics course), as opposed to attitudes about the statistics field itself (i.e., *Value* and *Interest*).

In other instances, a more detailed picture of students' attitudes on statistics may be required. For example, when an examination of associations among the subscales or with others measures is at the forefront, it may be more informative to include the six constructs because a differential pattern may otherwise go unnoticed.

5.3. LIMITATIONS AND SUGGESTIONS FOR FURTHER RESEARCH

The main limitations of the present study are the homogeneity of the participants (Educational Sciences and Speech Pathology and Audiology students; mainly female participants) and the fact that data from only one statistics course and one administration were considered. Future longitudinal studies including students from other academic fields and other statistics courses are needed to further validate the SATS-36 and generalize the present findings. Specifically, the authors are looking forward to new studies on the differences in mean scores and on the differential relationships between *Affect*, *Cognitive Competence*, and *Difficulty* on the one hand and other SATS factors or external variables on the other hand. Also further examination of the relatively high association between *Interest* and *Value* that was observed in this study would be interesting. Such studies would contribute to an understanding of the similarities and differences between the three SATS factors that are highly correlated. Although our results on the pre-test version of the SATS-36 likely transfer to the post-test version, research specifically addressing the post-test version is needed. Despite these limitations, the results from the present study clearly addressed important questions regarding the measurement of attitudes toward statistics and showed the additional value of analyzing individual items rather than item parcels.

ACKNOWLEDGEMENTS

This research was partially supported by Grant GOA 2006/01 "Developing adaptive expertise in mathematics education" from the Research Fund K.U.Leuven, Belgium.

REFERENCES

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*(1), 78–102.

Bandalos, D. L., & Finney, S. J. (2001). Item parceling issues in structural equation modeling. In G. A. Marcoulides and R. E. Schumacker (Eds.), *New developments and techniques in structural equation modeling* (pp. 269–296). Hillsdale, NJ: Lawrence Erlbaum Associates.

Boomsma, A., & Hoogland, J. J. (2001). The robustness of LISREL modeling revisited. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation models: Present and future. A Festschrift in honor of Karl Jöreskog* (pp. 139–168). Chicago: Scientific Software International.

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural Psychology, 1*(3), 185–216.

Byrne, B. M. (1989). *A primer of LISREL: Basic applications and programming for confirmatory factor analytic models*. New York: Springer.

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]

Cashin, S. E., & Elmore, P. B. (2005). The Survey of Attitudes Toward Statistics scale: A construct validity study. *Educational and Psychological Measurement, 65*(3), 1–16.

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.

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*(2), 129–141.

Finney, S. J., & DiStefano, C. (2006). Dealing with nonnormal and categorical data in structural equation models. In G. R. Hancock & R. O. Mueller (Eds.), *A second course in structural equation modeling.* Greenwich, CT: Information Age.

Finney, S. J., & Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology, 28*(2), 161–186.

Flora, D. B., & Curran P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*(4), 466–491.

Gal, I., & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: Towards an assessment framework. *Journal of Statistics Education, 2*(2). [Online: http://www.amstat.org/publications/jse/v2n2/gal.html]

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). Amsterdam: IOS Press.

Hatcher, L. (1994). *A step-by-step approach to using the SAS system for factor analysis and structural equation modeling.* Carey, NC: SAS Institute, Inc.

Hau, K. T., & Marsh, H. W. (2004). The use of item parcels in structural equation modelling: Non-normal data and small sample sizes. *British Journal of Mathematical Statistical Psychology, 57*(2), 327–351.

Hilton, 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.

Hoyle, R. H., & Panter, A. T. (1995). Writing about structural equation models. In R. H. Hoyle (Ed.), *Structural equation modeling: Concepts, issues, and applications* (pp. 158–176). Thousand Oaks, CA: Sage Publications.

Hu, L., Bentler, P. M., & Kano, Y. (1992). Can test statistics in covariance structure analysis be trusted? *Psychological Bulletin, 112*(2), 351–362.

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

John, O. P., & Soto, C. J. (2007). The importance of being valid: Reliability and the process of construct validation. In R. W. Robins, R. C. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 461–494). New York: The Guilford Press.

Jöreskog, K. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 294–316). Newbury Park, NJ: Sage Publications.

Jöreskog, K. G. (1999). *Formulas for skewness and kurtosis.* Unpublished manuscript. [Online: http://www.ssicentral.com/lisrel/techdocs/kurtosis.pdf]

Jöreskog, K. G. (2004). *Structural equation modeling with ordinal variables using LISREL*. Unpublished manuscript.

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

Jöreskog, K. G., & Sörbom, D. (2004). LISREL 8.7 for Windows [Computer Software]. Lincolnwood, IL: Scientific Software International, Inc.

Keeler, C. M., & Steinhorst, R. K. (1995). Using small groups to promote active learning in the introductory statistics course: A report from the field. *Journal of Statistics Education, 3*(2).
[Online: http://www.amstat.org/publications/jse/v3n2/keeler.html]

Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York: The Guilford Press.

Leong, J. L. (2006). *High school students' attitudes and beliefs regarding statistics in a service-learning based statistics course* (Unpublished doctoral dissertation). Georgia State University. Georgia.

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*(2), 151–173.

MacCallum, R., Roznowski, M., & Necowitz, L. B. (1992). Model modifications in covariance structure analysis: The problem of capitalization on chance. *Psychological Bulletin, 111*(3), 490–504.

McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D. A. Grouws (Ed.), *Handbook of research on mathematics teaching and learning* (pp. 575–596). New York: Macmillan.

Mills, J. (2004). Students' attitudes toward statistics: Implications for the future. *College Student Journal, 38*(3), 349–362.

Nasser F., & Wisenbaker J. (2003). A Monte Carlo study investigating the impact of item parceling on measures of fit in confirmatory factor analysis. *Educational and Psychological Measurement, 63*(5), 729–757.

Raftery, A. E. (1995). Bayesian model selection in social research. In A. E. Raftery (Ed.), *Sociological methodology* (pp. 111–164). Oxford: Blackwell.

Roberts, D. M., & Bilderback, E. W. (1980). Reliability and validity of a statistics attitude survey. *Educational and Psychological Measurement, 40*(1), 235–238.

Roberts, D. M., & Saxe, J. E. (1982). Validity of a statistics attitude survey: A follow-up study. *Educational and Psychological Measurement, 42*(3), 907–912.

Satorra, A., & Bentler, P. M. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye & C. C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Thousand Oaks, CA: Sage.

Satorra, A., & Bentler, P. M. (2001). A scaled difference chi-square test statistic for moment structure analysis. *Psychometrika, 66*(4), 507–514.

Schau, C. (2003, August). *Students' attitudes: The "other" important outcome in statistics education*. Paper presented at the Joint Statistical Meetings, San Francisco, CA.
[Online: http://evaluationandstatistics.com/JSM2003.pdf]

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.

Schumacker, R. E., & Lomax, R. G. (2004). *A beginners' guide to Structural Equation Modelling.* Mahwah, NJ: Lawrence Erlbaum.

Shultz, K. S., & Koshino, H. (1998). Evidence of reliability and validity for Wise's Attitude Toward Statistics scale. *Psychological Reports, 82*(1), 27–31.

SSICentral (n.d.). *The diagnosis and treatment of non-normality.* Unpublished manuscript.
[Online: http://www.ssicentral.com/lisrel/techdocs/Session4.pdf]

Suanpang, P., Petocz, P., & Kalceff, W. (2004). Student attitudes to learning business statistics: Comparison of online and traditional methods. *Educational Technology & Society, 7*(3), 9–20.

Tempelaar, D. T., van der Loeff, S. S., & Gijselaers, W. H. (2007). A structural equation model analyzing the relationship of students' attitudes toward statistics, prior reasoning abilities and course performance. *Statistics Education Research Journal, 6*(2), 78-102.
[Online: http://www.stat.auckland.ac.nz/~iase/serj/SERJ6%282%29_Tempelaar.pdf]

Wang, Y., & Liu, Q. (2006). Comparison of Akaike information criterion (AIC) and Bayesian information criterion (BIC) in selection of stock–recruitment relationships. *Fisheries Research, 77*(2), 220–225.

Waters, L. K., Martelli, T. A., Zakrajsek, T., & Popovich, P. M. (1988). Attitudes toward statistics: An evaluation of multiple measures. *Educational and Psychological Measurement, 48*(2), 513–516.

Wise, S. L. (1985). The development and validation of a scale measuring attitudes toward statistics. *Educational and Psychological Measurement, 45*(2), 401–405.

STIJN VANHOOF
Centre for Methodology of Educational Research
Katholieke Universiteit Leuven
Vesaliusstraat 2 bus 3762
3000 Leuven (Belgium)
tel.: +32 485-93.01.99
fax: +32 16-32.59.34