# CONFIRMING THE STRUCTURE OF THE SURVEY OF ATTITUDES TOWARD STATISTICS (SATS-36) BY SWEDISH STUDENTS

INGER PERSSON
*Uppsala University*
*inger.persson@statistics.uu.se*

KATRIN KRAUS
*Uppsala University*
*katrin.kraus@statistics.uu.se*

LISBETH HANSSON
*Uppsala University*
*lisbeth.hansson@statistics.uu.se*

FAN YANG WALLENTIN
*Uppsala University*
*fan.yang@statistics.uu.se*

## ABSTRACT

*Research on students' attitudes toward statistics has attracted many statistics instructors and statistics education researchers. In this study, we use confirmatory factor analysis to analyze data collected from an introductory statistics course using the Survey of Attitudes toward Statistics. The results suggest that the items and six factors are conceptually relevant, confirming the six-factor structure of the pretest version of SATS-36 on this sample of Swedish students, with a few suggested modifications of the original model structure. Two items are excluded from the Difficulty component, two items on the Affect component are allowed to correlate, and two items on the Cognitive competence component are also allowed to correlate.*

***Keywords:*** *Statistics education research; Structural equation models; Confirmatory factor analysis*

## 1. INTRODUCTION

Being a statistician is ranked as the best job one can have, according to a recent ranking by the job hunting site CareerCast.com (http://www.careercast.com/jobs-rated/best-jobs-2017). The demand for statisticians or data analysts has been constantly increasing for a few years. This puts emphasis on statistics education. Not only do we need to meet the increasing demand for statisticians but we also need to train statistics students to meet the high expectations from future employers.

Many suggest that there is a positive association between student attitudes towards their chosen subject and a favorable outcome of their studies. It is also believed that their attitudes have a large impact on whether they will use what they have learned in the future (e.g., Wentzel & Wigfield, 2009; and other references presented by VanHoof, Kuppens, Castro Sotos, Verschaffel, & Onghena, 2011). These attitudes are perhaps especially important in statistics, where an old and stubborn rumor of statistics classes being extremely difficult and boring is still alive. Therefore, it is important to (1) investigate students' perceptions of statistics and the relationship between these perceptions and their attitudes toward statistics, (2) assess the different dimensions of students' attitudes towards statistics in a sufficient way, and (3) provide information to students and instructors that can be used to evaluate and improve our courses.

Evaluation of attitudes can be done in a number of different ways and in recent years the Survey of Attitudes Toward Statistics (SATS) developed by Schau, Stevens, Dauphinee, and Del Vecchio (1995)

*Statistics Education Research Journal, 18(1), 83–93, http://www.stat.auckland.ac.nz/serj*

has become widely used. This instrument has been evaluated to some extent, but further research of its structure is needed. Firstly, two versions of the instrument have been proposed: a six-factor pretest version (SATS-36) and a four-factor version (SATS-28). Tempelaar, Schim van der Loeff, and Gijselaers (2007) show that the six-factor version (SATS-36) is preferred over the four-factor version (SATS-28), and the six-factor version is now being used predominantly. Secondly, while the intention of the SATS was for components to be scored using all of the items within each component, early studies of the SATS used parceling, where average values of subgroups of items within the same factor are used, in conjunction with confirmatory factory analysis (CFA) and structural equation modeling (SEM) as was the suggested practice at the time. VanHoof et al. (2011) show that using individual items and not parceling is preferred, based on their CFA confirming the structure of the SATS-36. Lastly, the suggested factors (dimensions) and the items used to capture these dimensions have also been discussed. VanHoof et al. show that the model fit can be improved by removing some items, and that some of the dimensions could be combined into one without losing a lot of information.

Extending the work of VanHoof et al. (2011), this study addresses the two above-mentioned structure issues of the six-factor version of the Survey of Attitudes toward Statistics (SATS-36). We studied information from individual items using confirmatory factor analysis based on a sample of Swedish students taking an introductory statistics course. We evaluate the suggested six factors of the instrument and their respective items.

## 2. THE SURVEY OF ATTITUDES TOWARD STATISTICS

### 2.1. SATS-36

The SATS was developed by Schau et al. (1995) to help understand student attitudes toward statistics and how learning and teaching are affected by these attitudes. The current version of the survey (SATS-36) consists of 36 response items, designed to capture information on six different attitude components. The descriptions of the attitude components below are cited from Schau’s wegpage http://www.evaluationandstatistics.com:

(i) *Affect*: students’ feelings concerning statistics (6 items)
(ii) *Cognitive competence*: students’ attitudes about their intellectual knowledge and skills when applied to statistics (6 items)
(iii) *Value*: students’ attitudes about the usefulness, relevance, and worth of statistics in personal and professional life (9 items)
(iv) *Difficulty*: students’ attitudes about the difficulty of statistics as a subject (7 items)
(v) *Interest*: students’ level of individual interest in statistics (4 items)
(vi) *Effort*: amount of work the student expends to learn statistics (4 items)

The 36 items are Likert-type with seven response alternatives for each item: 1 = *Strongly disagree*, 4 = *Neither Disagree nor Agree*, and 7 = *Strongly agree*. Some items are positively worded and some are negatively worded, and negatively worded items are reversed before scoring. The average item score is calculated for each of the six components, where higher scores indicate more positive attitudes (including perceiving the course as less difficult and expending more effort). The full version of SATS-36 including information on the scoring of the components can be requested from http://www.evaluationandstatistics.com.

Two versions of the SATS are available: one to administer before a statistics course, and one to administer after the course (same questions, but with corresponding changes in verb tense). In this study, we used only the pretest version of the SATS.

### 2.2. INDIVIDUAL ITEM RESPONSES PREFERRED OVER PARCELING OF ITEMS

VanHoof et al. (2011) provide an overview of the main reasons for using item parceling in the context of SATS data, that is to use mean values for subgroups of items instead of using the individual item responses. The motivation behind parceling is generally to simplify the model, in order to get more stable estimates and to improve reliability, which was the accepted practice at the time. There are many reasons presented against parceling in more recent literature, which points to disadvantages such as the introduction of unnecessary bias, the sensitivity of choice of parceling method, the possibility of masking an underlying multidimensional structure, etc. (see VanHoof et al., 2011, for further references). This study extends VanHoof et al. using the preferred confirmatory factor analysis on individual item scores of the SATS-36 rather than parcel scores, in order to give insight into the pretest factor structure based on a sample of Swedish students.

## 3. METHOD

### 3.1. DATA

The pretest version of SATS-36 was administered to 707 students (51% female, 49% male) in an introductory undergraduate statistics course at the Department of Statistics at Uppsala University in Sweden. The most common majors of these students are business (47%), political science (22%), and social science (8%). For three courses in 2014 and 2015, the survey was administered immediately before the first lecture and completed in class. Because Swedish students have very good English skills, there was no need to translate the survey.

In order to include as much of the collected information as possible, missing values have been imputed by pseudomedian imputation. Very few students had more than a few missing values, which makes imputation appropriate. Three individuals had a large share of missing values (two responded to only 33% of the items and one responded to fewer than 60% of the items) and have been excluded from the analyses, which leads to 704 students being included. Of these, almost 89% have complete data (see Appendix Table A1). Nine percent of the students have 1 missing response, 1.56% (11 students) have 2 missing responses, and three students have between 3 and 6 missing responses. In total, 0.4% of the overall data (100 item responses) are missing and subsequently imputed. Item 3 and item 15 (*Affect*) have the largest share of missing values, 1.28% each (see Appendix Table A2). Items 8 (*Difficulty*), 9 (*Value*), 29 (*Interest*), and 30 (*Difficulty*) have 0.99% missing values, and the rest of the items have 0.71% missing values or less. There is no relationship between the percentage of missing values and whether an item is positively or negatively worded, and the items with most missing values are not different from the other items considering the characteristics of the items.

There are no patterns in the data concerning missing values, in other words they appear to be missing at random. The analyses presented in Section 3.2 below have been repeated with listwise deletion of missing data, giving similar results (not shown).

### 3.2. STATISTICAL ANALYSES

The six factor model (SATS-36) is evaluated by confirmatory factor analysis, using LISREL 9.30. The 36 response items are measured on an ordinal scale, which has to be taken into account when choosing the estimation technique. Yang-Wallentin, Jöreskog, and Luo (2010) show that Robust Maximum Likelihood (RML), Unweighted Least Squares (ULS), or Diagonally Weighted Least Squares (DWLS), based on polychoric correlations, are recommended for this type of study. In order to compare the results to VanHoof et al. (2011), RML is chosen (ULS and RDWLS give similar results). The polychoric correlation approach assumes bivariate normality, which is investigated by the Root Mean Square Error of Approximation (RMSEA) values for population discrepancy (Jöreskog, Olsson, & Wallentin, 2016), and no violations of this assumption are observed.

To evaluate the model fit, the following goodness-of-fit indices are used: The RMSEA, the Comparative Fit Index (CFI), and the Parsimony Normed Fit Index (PNFI). Hair, Black, Babin, and Anderson (2014) suggest that a model has a good fit for a study of this size if RMSEA is 0.07 or less, with CFI of 0.90 or higher. We used the PNFI to compare models, with the highest value suggesting a better fit relative to the complexity of the model.

Because one of the aims of this study is to evaluate the factor structure, different models will be evaluated, where modifications are made based on the results from the first analysis. For this purpose, the chi-squared statistic is used. Following the recommendations from Yang-Wallentin et al. (2010), we used the Satorra and Bentler (1988) scaled chi-squared value for this purpose. It is well-known that the chi-squared-based statistic is very sensitive to sample size. For example, Kline (2005) and Hair et al. (2014) state that significant p-values are expected for a study of this size, therefore the chi-squared statistic is not used to establish the fit of the model. Additionally, the Akaike Information Criterion

(AIC) and Bayesian Information Criterion (BIC) are used, where the model with the lowest AIC and BIC is to be preferred (Jöreskog, 1993; Kline, 2005).

Suggestions of how to modify the model are based on modification indices, where the indices with the highest values primarily are chosen. We made changes one at a time, and re-evaluated the modification indices after each change. We considered modifications as long as there is/are any relatively high modification index value(s) suggesting substantial improvements of the model. Also, following the recommendation of Hooper, Coughlan, and Mullen (2008), we removed items with low multiple $R^2$ from the analysis as that indicates very high levels of error. To be consistent with the outline of VanHoof et al. (2011), items with factor loadings below 0.40 are considered for deletion from the model.

Cronbach's (1951) alpha is used to assess the internal reliability, indicating the degree of interrelationship among item responses within each component. The alpha coefficient is supplemented by the Construct Reliability (CR), indicating the reliability and internal consistency of each component (Hair, Black, Babin, & Anderson, 2010). Hair et al. (2010) state that a reliability estimate above 0.7 suggests good reliability.

## 4. RESULTS AND DISCUSSION

We present the goodness-of-fit indices from the confirmatory factor analysis of the original SATS-36 instrument in Table 1, together with results for a model with suggested modifications. The original model does not meet the recommended cut-off values, which means that these results do not initially confirm the structure of the SATS-36 for this group of Swedish students. The RMSEA of 0.080 is above the benchmark for a good model fit. Confirming this, the CFI also does not pass the recommended value of 0.90 (CFI = 0.885).

As a first step, we modified the original model by deleting items with loadings < 0.4, namely items 22, 30, and 34. Two of the deleted items (22 and 34) were also found to have weak loadings by VanHoof et al. (2011). All three items are associated with the *Difficulty* component. Item 30 "Statistics involves massive computations" and item 34 "Statistics is highly technical" are not necessarily indicators of *Difficulty* among young students today. Most students are comfortable using computers; therefore these items may be somewhat misleading when trying to capture the difficulty of the subject. The third item having a weak loading is item 22 "Statistics is a subject quickly learned by most people" which as VanHoof et al. pointed out, asks about *most* people's attitudes and not the participant's own attitudes. The deletion of these items is therefore justifiable not only based on the weak loadings, but also due to their potential obfuscation and the lack of a clear connection to the difficulty of the subject.

The next suggested modification is to let the errors of item 3 and item 19 correlate. Item 3 "I will like statistics" and item 19,"I will enjoy taking statistics courses" are both used as indicators for the *Affect* component. It is easy to justify that these two could be correlated, they are the only two positively worded items for this component and the meanings of these items are closely related. This again confirms the results of the study by VanHoof et al. (2011).

The modification indices further suggest letting the errors of items 31 and 32 correlate. Item 31 "I can learn statistics" and item 32 "I will understand statistics equations" are both indicators for *Cognitive competence* and it is again easy to justify that these items could be correlated. The modified model results in a lower chi-square value, lower AIC and BIC values, and higher PNFI compared to the original model. The recommended cut-off value of CFI is now met (CFI = 0.919), and the RMSEA is approaching the suggested threshold of 0.07 (RMSEA = 0.074); see Table 1.

In addition to the modifications above, the indices suggest letting item 6,"Statistics formulas are easy to understand" be an indicator for both *Difficulty* and *Cognitive competence*. This further improves the model fit, but the loading for *Difficulty* drops far below 0.4 and implies that item 6 is no longer a good indicator for this component (results not shown). This would otherwise be a reasonable modification. "Statistics formulas are easy to understand" could very well capture students' attitudes about their intellectual knowledge and skills in addition to the difficulty of the subject. In order to evaluate whether item 6 should be used as an indicator for Cognitive competence *instead of* being an indicator for *Difficulty*, one more modification was evaluated. This time item 6 was used *only* as an indicator for *Cognitive competence* (not for *Difficulty*). This however resulted in a worse model fit,

*Table 1. Fit indices for SATS-36 evaluated on pretest data (n = 704)*

| Model | SBS $\chi^2$ | Df | RMSEA | CFI | PNFI | AIC | BIC |
|---|---|---|---|---|---|---|---|
| Original | 1891.6 | 579 | 0.080 | 0.885 | 0.775 | 16682.7 | 17079.2 |
| Modified* | 1371.4 | 478 | 0.074 | 0.919 | 0.797 | 14189.9 | 14568.1 |
| *Affect*/*Cognitive* | combined (four-component model) | | | | | | |
| *Competence*/*Difficulty* | 1474.0 | 487 | 0.076 | 0.910 | 0.804 | 14323.9 | 14661.1 |
| *Affect*/*Cognitive* | combined (five-component model) | | | | | | |
| *Competence* | 1391.9 | 483 | 0.074 | 0.917 | 0.804 | 14205.6 | 14561.1 |
| *Affect*/*Cognitive* | combined (five-component model), item 36 deleted | | | | | | |
| *Competence* | 1330.3 | 452 | 0.075 | 0.919 | 0.804 | 13603.8 | 13950.1 |

*Modified model: Deleting items with loadings < 0.4 (items 22, 30, 34), letting errors of items 3 and 19 correlate, and letting errors of items 31 and 32 correlate.

confirming the suggestion of using item 6 as an indicator for *Difficulty*. No other substantial changes were suggested from the modification indices.

Parameter estimates (loadings) for all items in the original and modified model are significant, with *p*-values for all items < 0.0005 (Table 2). Notice that the loading for item 36 is relatively weak in both models, just above 0.4. In the study by VanHoof et al. (2011) this item was excluded due to its weak loading, and if the recommendations by Hooper et al. (2008) were adhered to, namely to remove items with multiple $R^2$ less than 0.20, this item would be excluded here as well.

*Table 2. Parameter estimates (loadings) of the evaluated models*

| | Item | Original | Modified | | Item | Original | Modified |
|---|---|---|---|---|---|---|---|
| *Affect* | 3 | 0.589 | 0.534 | *Value* | 7 | 0.558 | 0.558 |
| | 4 | 0.694 | 0.710 | | 9 | 0.545 | 0.544 |
| | 15 | 0.516 | 0.521 | | 10 | 0.569 | 0.567 |
| | 18 | 0.658 | 0.670 | | 13 | 0.603 | 0.602 |
| | 19 | 0.584 | 0.522 | | 16 | 0.681 | 0.682 |
| | 28 | 0.808 | 0.817 | | 17 | 0.544 | 0.544 |
| | | | | | 21 | 0.460 | 0.460 |
| *Cognitive* | 5 | 0.723 | 0.734 | | 25 | 0.706 | 0.706 |
| *competence* | 11 | 0.633 | 0.627 | | 33 | 0.762 | 0.763 |
| | 26 | 0.700 | 0.708 | | | | |
| | 31 | 0.571 | 0.527 | *Interest* | 12 | 0.703 | 0.705 |
| | 32 | 0.592 | 0.552 | | 20 | 0.800 | 0.798 |
| | 35 | 0.772 | 0.775 | | 23 | 0.808 | 0.811 |
| | | | | | 29 | 0.874 | 0.872 |
| *Difficulty* | 6 | 0.548 | 0.552 | | | | |
| | 8 | 0.616 | 0.593 | *Effort* | 1 | 0.644 | 0.643 |
| | 22 | 0.352 | | | 2 | 0.873 | 0.873 |
| | 24 | 0.658 | 0.617 | | 14 | 0.778 | 0.778 |
| | 30 | 0.287 | | | 27 | 0.515 | 0.516 |
| | 34 | 0.295 | | | | | |
| | 36 | 0.431 | 0.417 | | | | |
| Error covariance between items 3 and 19 (*p*-value) | | | | | | | 0.443 (0.0354) |
| Error covariance between items 31 and 32 (*p*-value) | | | | | | | 0.377 (0.0393) |

After the suggested improvements of the *Difficulty*, *Affect*, and *Cognitive Competence* components, Figure 1 shows the modified model for which all items load not only significantly but also relatively strongly (Table 4) on all the six attitude components, confirming the six-factor structure.

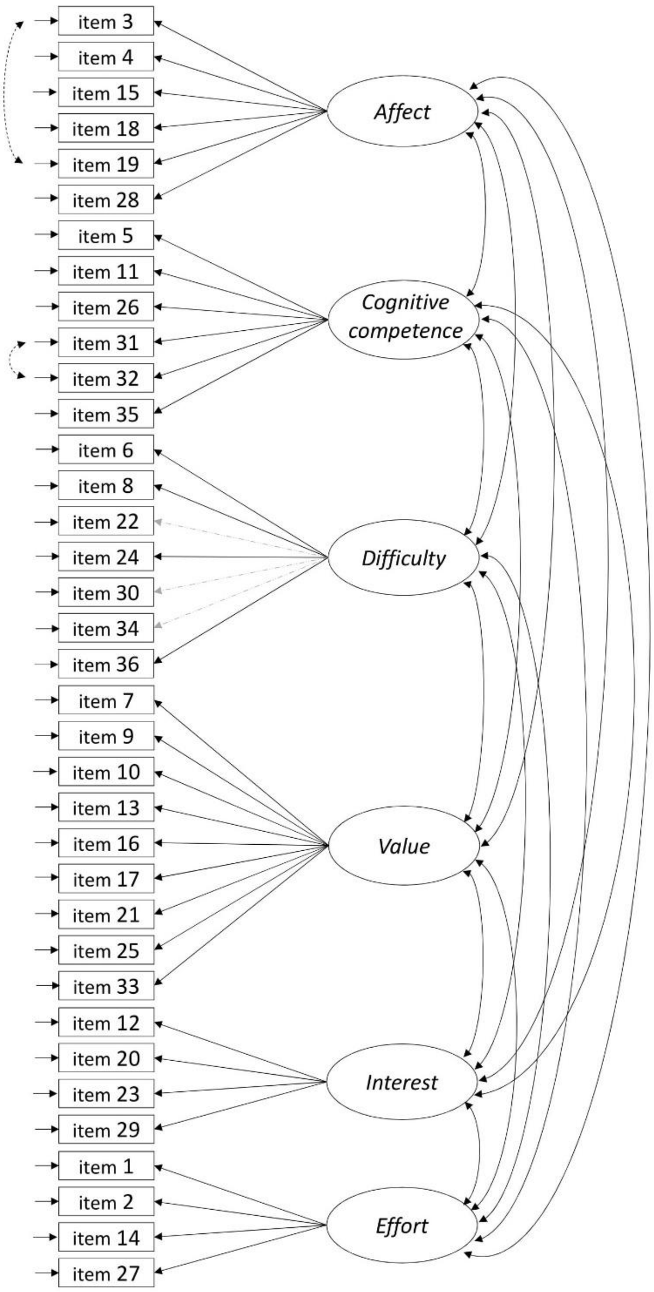


*Figure 1. The six factor structure of the 36 SATS items*

In Figure 1, all relations between components are correlations. The grey dotted lines from the Difficulty component to items 22, 30, and 34 indicate that corresponding relations are omitted in the modified model. The dotted arrows between the errors of items 3 and 19, and between items 31 and 32, respectively, indicate the added covariance between the corresponding error terms in the modified model.

Looking at the correlations between the components (Figure 1 and Table 3), we can see strong and significant correlations (0.8 and above) between *Affect*, *Cognitive Competence,* and *Difficulty*. This again confirms the results of VanHoof et al. (2011), suggesting that it could be possible to combine these three factors into one. Especially the nearly perfect correlation between *Affect* and *Cognitive Competence* (0.989, Table 3) suggests that at least a combination of these two components should be investigated. Furthermore, we see a strong correlation also between *Value* and *Interest* (0.725), which VanHoof et al. also found.

*Table 3. Estimated latent factor correlations (p-values) for the modified model*

| | *Affect* | *Cognitive competence* | *Difficulty* | *Value* | *Interest* | *Effort* |
|---|---|---|---|---|---|---|
| *Affect* | 1.000 | | | | | |
| *Cognitive competence* | 0.989 (0.015) | 1.000 | | | | |
| *Difficulty* | 0.853 (0.035) | 0.802 (0.037) | 1.000 | | | |
| *Value* | 0.400 (0.041) | 0.495 (0.038) | 0.099 (0.055) | 1.000 | | |
| *Interest* | 0.341 (0.047) | 0.404 (0.043) | 0.031 (0.058) | 0.725 (0.028) | 1.000 | |
| *Effort* | –0.163 (0.048) | –0.043 (0.050) | –0.393 (0.050) | 0.277 (0.047) | 0.404 (0.047) | 1.000 |

Table 1 shows the results of two additional models, one five-component model where the *Affect* and *Cognitive Competence* components are combined, and one four-component model where *Affect*, *Cognitive Competence*, and *Difficulty* are combined into one component. With the four-component model, all fit indices except *PNFI* indicate that this model has a worse model fit than the modified six-component model. The five-component model, combining only the two components *Affect* and *Cognitive Competence*, improves the model fit compared to the four-component model. When comparing to the modified six-component model, however, only the values of PNFI and BIC improve, whereas Chi-square, CFI, and AIC suggest a worse model fit. The loading for item 36 now drops below 0.4 (results not shown). Therefore, as a last step we estimated a five-component model without item 36 (Table 1). This model slightly improves the Chi-square, PNFI, AIC, and BIC compared to the modified six-component model. The loadings of this model (Table 4) are similar to the loadings of the modified six-component model in Table 2, and the latent factor correlations (Table 5) are also similar to those in Table 3.

Finally, we present the measures of internal reliability and consistency for the modified six-component model in Table 6. Cronbach's alpha and Construct Reliability both exceed the suggested threshold of 0.7 for all components except *Difficulty*, which is in line with previous research (see e.g. Schau, 2003; Tempelaar et al., 2007; Tempelaar & Nijhuis, 2007; Stanisavljevic, Trajkovic, Marinkovic, Bukumiric, Cirkovic, & Milic, 2014). This means that the items consistently represent the same component, for all components but *Difficulty*.

## 5. CONCLUSIONS

### 5.1. SUMMARY OF FINDINGS

The major conclusion of this study is that the six-factor structure captures the observed relationships between the items in the SATS-36 quite well. This was confirmed by analyzing pretest item responses from a sample of Swedish students using confirmatory factor analysis.

A few modifications of the original model are suggested in order to improve the instrument. First, three *Difficulty* items (item 22, item 30, and item 34) are excluded due to weak factor loadings. Two of the excluded items are associated with statistics being technical and involving massive computations. Today's students may not be so intimidated by such things. The third excluded item asks about *most* people's attitudes and not the attitudes of the student him/herself. VanHoof et al. (2011) also found weak loadings with three *Difficulty* items, two of which matched the items found in this study, indicating that the *Difficulty* component is difficult to capture with the originally suggested items.

*Table 4. Parameter estimates (loadings) of the five-component model: Affect and Cognitive Competence are combined into one component, Diffculty is reduced to three items (item 36 deleted)*

| | Item | Loading | | Item | Loading |
|---|---|---|---|---|---|
| *Affect/ Cognitive Competence* | 3 | 0.555 | *Value* | 7 | 0.559 |
| | 4 | 0.704 | | 9 | 0.544 |
| | 5 | 0.735 | | 10 | 0.565 |
| | 11 | 0.615 | | 13 | 0.601 |
| | 15 | 0.519 | | 16 | 0.681 |
| | 18 | 0.656 | | 17 | 0.544 |
| | 19 | 0.542 | | 21 | 0.461 |
| | 26 | 0.712 | | 25 | 0.705 |
| | 28 | 0.814 | | 33 | 0.765 |
| | 31 | 0.504 | | | |
| | 32 | 0.542 | *Interest* | 12 | 0.706 |
| | 35 | 0.777 | | 20 | 0.798 |
| | | | | 23 | 0.812 |
| *Difficulty* | 6 | 0.556 | | 29 | 0.871 |
| | 8 | 0.582 | | | |
| | 24 | 0.587 | *Effort* | 1 | 0.640 |
| | | | | 2 | 0.878 |
| | | | | 14 | 0.775 |
| | | | | 27 | 0.514 |
| Error covariance between items 3 and 19 (*p*-value) | | | | 0.421 (0.0348) | |
| Error covariance between items 31 and 32 (*p*-value) | | | | 0.394 (0.0392) | |

*Table 5. Estimated latent factor correlations (p-values) for the five-component model (item 36 deleted)*

| | *Affect/Cognitive competence* | *Difficulty* | *Value* | *Interest* | *Effort* |
|---|---|---|---|---|---|
| *Affect/Cognitive Competence* | 1.000 | | | | |
| *Difficulty* | 0.855 (0.035) | 1.000 | | | |
| *Value* | 0.451 (0.037) | 0. 087 (0.056) | 1.000 | | |
| *Interest* | 0.376 (0.043) | 0.051 (0.060) | 0.726 (0. 028) | 1.000 | |
| *Effort* | –0.101 (0.047) | –0.407 (0.051) | 0.276 (0.047) | 0.404 (0.047) | 1.000 |

*Table 6. Cronbach's alpha and Construct Reliability (CR) by attitude component for the modified six-component model*

| Component | Alpha | CR |
|---|---|---|
| *Affect* | 0.790 | 0.800 |
| *Cognitive competence* | 0.797 | 0.820 |
| *Difficulty* | 0.593 | 0.630 |
| *Value* | 0.801 | 0.839 |
| *Interest* | 0.840 | 0.875 |
| *Effort* | 0.711 | 0.802 |

Another suggested modification is to allow two *Affect* items to correlate, namely the only two positively worded items for this component, items with closely related meanings. This again supports the findings by VanHoof et al. (2011).

Further improvement of the structure is achieved by also letting two items within *Cognitive Competence* correlate, items that seem to have more in common than what can be captured by the original model structure.

All these modifications are reasonable and justifiable and lead to a model that fits the data better than the original model. The modified model shows internal reliability and consistency for all components except *Difficulty*, as measured by Cronbach's alpha and Construct Reliability (CR).

Focusing on the relationships between the six attitude components, strong correlations are found between *Affect*, *Cognitive Competence,* and *Difficulty*, again confirming the results by VanHoof et al. (2011). Combining these components into one, however, results in a worse model fit. A combination of *Affect* and *Cognitive Competence,* the two components with the strongest relationship, slightly improves the model fit indices when excluding item 36 based on its weak factor loading. This results in a five-component model with loadings and factor correlations quite similar to those of the six-component model. In other words, the relationship between the items and the factors as well as the relationship between the factors seem to be largely unaffected by whether *Affect* and *Cognitive Competence* are treated as one or two components. Whereas the five- and six-component models are largely comparable, with the modified five-component model exhibiting a slightly better set of fit statistics, we do not recommend combining any of the three components into one, because they all reflect quite different and theoretically reinforced aspects of students' attitudes towards statistics and combining them does not improve the model fit meaningfully.

These results are based on a relatively large sample of Swedish students, with an approximately equal number of female and male students, including students with a number of different majors. The results and conclusions based on this heterogeneous sample from other academic fields and other statistics courses than previously studied add to the findings of previous research by confirming the six-factor structure with some modifications. These suggested modifications should be taken seriously by statistics researchers in the future.

## 5.2. SUGGESTIONS FOR FURTHER RESEARCH

As shown in this study, the *Difficulty* component seems difficult to capture with the original items. Here we excluded three out of seven items. Further research is needed to either (1) confirm that four items are adequately capturing the *Difficulty* component, or (2) find new items that could be better indicators of how difficult the students perceive statistics to be. It is clear that the SATS-36 needs some tweaking to meet its full potential. Furthermore, because only the pretest version of the instrument is analyzed in this study, research is needed on the posttest version. In addition, it would be very interesting to investigate any causal relationships between the attitude components. Finally, as the meaning of the items is very important and students of today might not interpret the items in the same way as it was intended when the SATS-36 was constructed, it would be valuable to conduct future research in which students are asked what they believe the different items mean.

## ACKNOWLEDGEMENTS

This research was partially supported by funding from the KoF11 research evaluation at Uppsala University.

## REFERENCES

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, *16*(3), 297–334.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). *Multivariate data analysis* (7th ed, pp. 679–680). Prentice-Hall, Inc. Upper Saddle River, NJ.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2014). *Multivariate data analysis* (pp. 578–584). Pearson New International Edition. Harlow: Pearson Education Limited.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural equation modelling: Guidelines for determining model fit. *The Electronic Journal of Business Research Methods, 6*(1), 53–60. [Online: http://www.ejbrm.com/issue/download.html?idArticle=183 ]

Jöreskog, K. G. (1993). Testing structural equation models. In K. A. Bollen & J. S. Long (Eds.), *Testing structural equation models* (pp. 294–316). Thousand Oaks, CA: Sage.

Jöreskog, K. G., Olsson, U. H., & Wallentin, F. Y. (2016). *Multivariate analysis with LISREL*. New York: Springer.

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

Satorra, A., & Bentler, P. M. (1988). Scaling corrections for chi-square statistics in covariance structure analysis. *Proceedings of the American Statistical Association, Section on Business and Economic Statistics Section* (pp. 308–313). Alexandria, VA: American Statistical Association.

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.

Schau, C. (2003, August). *Students' attitudes: The "other" important outcome in statistics education*. Paper presented at the Joint Statistical Meetings, San Francisco.

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

Tempelaar, D. T., & Nijhuis. J. F. H. (2007). Commonalities in attitudes and beliefs toward different academic subjects. In M. K. McCuddy, H. van den Bosch, J. W. B. Martz, A. V. Matveev, & K. O. Morse (Eds.), *Educational innovation in economics and business X: The challenges of educating people to lead in a challenging world* (pp. 225–250). Berlin: Springer.

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

VanHoof, S., Kuppens, S., Castro Sotos, A. E., Verschaffel, L., & Onghena, P. (2011). Measuring statistics attitudes: Structure of the Survey of Attitudes Toward Statistics (SATS-36). *Statistics Education Research Journal, 10*(1), 35–51. [Online: https://iase-web.org/documents/SERJ/SERJ10(1)_Vanhoof.pdf ]

Wentzel, K. R., & Wigfield, A. (2009). *Handbook of motivation at school.* New York: Routledge.

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

INGER PERSSON
Department of Statistics, Uppsala University
P.O. Box 513, SE 751 20
Uppsala, SWEDEN

## APPENDIX: SUMMARY STATISTICS FOR MISSING VALUES

*Table A1. Number of students with missing responses (n = 704)*

| Number of items with missing response | Number of students | Percent of students |
|---|---|---|
| 0 | 625 | 88.78 |
| 1 | 65 | 9.23 |
| 2 | 11 | 1.56 |
| 3 | 1 | 0.14 |
| 4 | 1 | 0.14 |
| 6 | 1 | 0.14 |

*Table A2. Missing value percentages per item*

| Item | % missing | Item | % missing | Item | % missing | Item | % missing |
|---|---|---|---|---|---|---|---|
| 1 | 0.14 | 10 | 0.43 | 19 | 0.00 | 28 | 0.28 |
| 2 | 0.43 | 11 | 0.14 | 20 | 0.14 | 29 | 0.99 |
| 3 | 1.28 | 12 | 0.28 | 21 | 0.28 | 30 | 0.99 |
| 4 | 0.43 | 13 | 0.14 | 22 | 0.00 | 31 | 0.28 |
| 5 | 0.00 | 14 | 0.14 | 23 | 0.14 | 32 | 0.43 |
| 6 | 0.43 | 15 | 1.28 | 24 | 0.57 | 33 | 0.43 |
| 7 | 0.28 | 16 | 0.28 | 25 | 0.71 | 34 | 0.14 |
| 8 | 0.99 | 17 | 0.00 | 26 | 0.14 | 35 | 0.14 |
| 9 | 0.99 | 18 | 0.00 | 27 | 0.57 | 36 | 0.28 |