


Rasch rating scale analysis of the survey of attitudes toward statistics

Mutasem Mohammad Akour^{1*} 

¹ Department of Educational Psychology, Faculty of Educational Sciences, The Hashemite University, Zarqa, JORDAN

Received 19 September 2022 • Accepted 08 November 2022

Abstract

Students in every discipline in higher education take at least one course in statistics. Therefore, it is necessary to enhance students' understanding of statistics and their achievement in such courses by considering several factors that might contribute to this enhancement. Students' attitudes toward statistics are a critical factor that influences their performance in statistics courses, and thus an accurate measurement of attitudes is needed. The survey of attitudes toward statistics (SATS-36) is widely used in measuring attitudes toward statistics; thus, it is important to ensure that its items accurately assess this construct. Therefore, the purpose of the current study was to validate this survey when administered to a convenience sample of 423 university students. Using the Rasch rating scale model, the current study examined the dimensionality, item fit to the Rasch model, item and person reliabilities, functionality of response categories, and distribution of the SATS-36 items along the attitudes toward statistics continuum.

The findings revealed excellent item and person reliabilities (greater than 0.90) and the unidimensionality of the survey. Additionally, all items were closely aligned with the respondents, and the response categories were well-functioning as each category had more than 10 observations and outfit statistics were all low. However, some improvements were suggested. All items on the effect subscale and some others from different subscales need to be altered in content, deleting three items (two from the value subscale and one from the difficulty subscale) and adding more items to have a better distribution of items along the continuum. Finally, the number of response categories is recommended to be reduced to five instead of seven to have a more efficient rating scale. The findings of the current study imply that even though great care has been taken in the development of this survey, examining the quality of its items and the utility of its rating scale in new settings, and using different validation approaches is necessary.

Keywords: statistics education research, validity, the SATS, psychometric properties, Rasch model

INTRODUCTION

In higher education and almost all disciplines, students are required to take at least one course in statistics or are exposed to some aspects of statistics. A wealth of past research has studied possible factors that could affect students' performance in statistics courses. Attitudes toward statistics were considered one of these key factors (Abbiati et al., 2021; Evans, 2007; Sesé et al., 2015; Tempelaar et al., 2007). Attitudes toward statistics refers to distinct but related dispositions that reflect favorable and unfavorable responses pertaining to statistics and statistics learning (Schau et al., 1995).

Students' performance in statistics courses was shown to have a positive relationship with attitudes toward statistics, indicating that students with more positive attitudes had higher achievement levels in these courses (Emmioglu & Capa-Aydin, 2012).

Negative attitudes toward statistics have not only an unfavorable impact on students' achievement in statistics courses but also an unpleasant influence on the adoption of statistical thinking and the likelihood of students using the knowledge acquired from statistics courses in their personal and professional lives. Students need to have positive attitudes toward statistics to be able to like, understand, and use statistics (Schau, 2003).

Contribution to the literature

- It is important to measure students' attitudes toward statistics given that attitudes influence students' performance in statistics courses.
- The current study contributed to the improvement of one of the most used surveys (the SATS-36) in measuring attitudes toward statistics.
- Using Rasch analysis, the current study provided new insights into the survey regarding the functionality of both the items and the rating scale.

Attitudes Toward Statistics

One of the most used instruments in measuring students' attitudes toward statistics is the survey of attitudes toward statistics (SATS) in its two versions: the four-factor SATS-28 (Schau et al., 1995) and the six-factor SATS-36 (Schau, 2003). The SATS-28 (Schau et al., 1995) consisted of 28 items distributed into four subscales:

- (a) Affect (six items): measures students' positive and negative feelings concerning statistics.
- (b) Cognitive competence (six items): measures students' attitudes toward intellectual knowledge and skills when applied to statistics.
- (c) Difficulty (seven items): measures students' attitudes toward the difficulty of statistics as a subject.
- (d) Value (nine items): measures students' attitudes toward the usefulness, relevance, and worth of statistics in their professional and personal life.

In 2003, Schau (2003) added two subscales to this instrument that was labeled SATS-36: interest (four items) measures students' levels of interest in statistics, and effort (four items) measures the amount of work that students spend on learning statistics.

The measurement quality of both versions of the survey was examined in a variety of research studies. In 2012, Nolan et al. (2012) reviewed past studies that presented validity and reliability evidence for the SATS-28 and the SATS-36. These studies (Bechrakis et al., 2011; Cashin & Elmore, 2005; Chiesi & Primi, 2009; Coetzee & Van der Merwe, 2010; Hilton et al., 2004; Tempelaar et al., 2007; Vanhoof et al., 2011) supported the four-factor structure of the SATS-28 and the six-factor structure of the SATS-36, using confirmatory factor analysis (CFA) and reported adequate to high internal consistency estimates for each subscale.

After that, additional studies were published, adapting the survey into different languages, and examining its factor structure and psychometric properties. Assarierh (2013) adapted the SATS-36 into Arabic. Data were collected from a sample of university students in an introductory course in statistics. Exploratory factor analysis (EFA) resulted in the same six factors obtained in the original survey, after the deletion of four items. Values of Cronbach's alpha for each subscale were adequate to good.

The psychometric properties of a Serbian version of the SATS-36 were explored by Stanisavljevic et al. (2014) using post-test data collected from a sample of medical students who were enrolled in an obligatory introductory statistics course. The findings of CFA confirmed the six-factor structure of the survey. Furthermore, Hommik and Luik (2017) adapted the SATS-36 in an Estonian context using data collected from a sample of secondary students at the beginning of a compulsory statistics course. The CFA findings did not support the six-factor model. However, the findings of EFA supported a four-factor structure of the scale, in which the three factors of affect, cognitive competence, and difficulty were combined into one.

The STAS-36 was also adapted into Turkish. Sarikaya et al. (2018) examined the factorial structure of this survey using the data collected from a sample of university students. CFA using item parceling approved the six-factor structure of the survey. All subscales had good reliability estimates, except for the difficulty subscale which showed an acceptable one. Moreover, Persson et al. (2019) conducted CFA on individual items on pre-test data from undergraduate students enrolled in an introductory course in statistics at a Swedish university. Their findings supported the six-factor structure of the survey. Three items in the difficulty factor were deleted to improve the survey. Additionally, they did not recommend combining the three components (affect, cognitive competence, and difficulty) into one.

Furthermore, Xu and Schau (2019) evaluated the fit of the six-factor model using the pre-test and post-test data obtained in three academic years from students enrolled in introductory statistics courses in the USA. The findings of this study showed that the six-factor model had a better acceptable fit with the post-test data than the pre-test data. Finally, Saidi and Siew (2019) investigated the factor structure of the SATS-36 using data obtained from a sample of tenth-grade students in Malaysia. The results of the second-order CFA on the item-level data confirmed the six-factor structure of the survey. Additionally, Cronbach's alpha, which was used to estimate reliability, showed good values.

All studies conducted on the SATS in its two versions to explore their factor structure and psychometric properties relied on using factor analysis (EFA or CFA) and Cronbach's alpha as an estimate of internal

consistency. No study, to the best of the author's knowledge, has used different approaches. Moreover, no study was conducted to examine the influence of both the number and the labeling of the response categories on the quality of this survey. Bond and Fox (2015) asserted that the quality of the items in any instrument and the utility of the rating scale should be assessed empirically, even though great care has been taken in the development process of that instrument. They stated that rather than confirming the optimal number of response categories for measuring a given construct, researchers should strive to empirically determine the optimal number of response categories whenever an existing measure is being used with a different population. Therefore, examining whether the items on the SATS-36 consistently collaborate to reflect the latent construct of attitudes toward statistics and that the response categories within all items are functioning as required is crucial.

The Rasch rating scale model (RSM) (Andrich, 1978) can show us how respondents used the rating scale and which response categorization would lead to a higher-quality measure. Thus, the current study examined the psychometric properties of the SATS-36, the functioning of its items, and the functioning of the rating scale used in this survey using the Rasch model on a post-test data of this survey.

It is hoped that the current study would provide researchers and practitioners with added information concerning the suitability of using the STAS-36 in measuring students' attitudes toward statistics. Additionally, the current study would provide researchers with new insight into the optimal number of response categories to use with this survey.

The Rasch Measurement Model

In evaluating psychometric properties of scales in psychology, education, and other disciplines, classical test theory (CTT) techniques, with factor analysis methods, are often used. This theory commonly sums scores on all items of a scale to compute a total score, and thus each item is considered to have the same difficulty. Putting it differently, each item is assumed to have the same amount of construct being measured. Therefore, regardless of the easiness or hardness of agreeing with an item, it carries the same amount of weight toward a respondent's total score. Additionally, all scores are assumed to be at the interval level. However, scores resulting from rating scales are at the ordinal level. That is, the difference between any two adjacent response categories (e.g., agree and strongly agree) is not the same as the difference between any other two adjacent categories. Moreover, CTT statistics are sample dependent. Therefore, the factor structure resulting from factor analysis might differ depending on the sample used.

The Rasch models are a family of measurement models that can be applied to different settings. The simplest model is the one-parameter logistic model, where the only parameter in the model is the location of the item on the latent trait continuum. This model is appropriate for dichotomous items (items scored as 0 and 1). However, for polytomous items (items scored 0, 1, 2, etc.) such as when dealing with ordinal data resulting from administering rating scales, RSM is applied.

In addition to the item location parameter, RSM includes a threshold parameter (τ_{ki}), which divides the distribution of responses into several ordered categories. The number of threshold values equals that of response categories (k) minus 1. In the SATS-36, a seven-point rating scale is used, and therefore, it would have six threshold values. These values cut the distribution of responses into seven ordered categories. Each threshold value represents a location on the latent continuum at which a person is equally likely to obtain one of two successive scores of two adjacent categories. For example, any item i on the SATS-36 has seven adjacent categories, and therefore, the first threshold of the item is the position on the attitudes toward statistics continuum at which a student is equally likely to obtain a score of 0 or 1. The second threshold is the position on the continuum at which a student is equally likely to obtain a score of 1 or 2, and so on through the sixth threshold value (DiStefano & Jiang, 2020).

Given the difficulty (δ_i) of the item i , τ_{ki} which is the same for all items, and (m) which is the maximum score, the probability that a category will be selected by a respondent with (β_n) level of attitudes toward statistics is provided by the following formula:

$$P(X_{ni} = x) = \frac{\exp \sum_{k=0}^x [\beta_n - (\delta_i - \tau_k)]}{\sum_{j=0}^m \exp \sum_{k=0}^j [\beta_n - (\delta_i - \tau_k)]} \quad (1)$$

This probability is then transformed into a logit score by taking the natural odds log value. If the probability for an item is computed across all persons, then the transformed probability is the item logit. However, a person logit results when the transformed probability is computed across all items for a given person (DiStefano & Jiang, 2020).

Therefore, instead of dealing with ordinal raw scores, RSM converts such data into logit units that are considered of interval level. One advantage of using Rasch models over CTT is that measurement indices are considered item- and sample-independent. RSM also provides numerous techniques to evaluate the functioning of each item, all items, and the rating scale used (Bond & Fox, 2015).

Purpose of the Current Study

Most students in higher education take only one introductory statistics course. Therefore, instructors should be aware of several factors that might affect

students' understanding of what they have learned in such a course and the application of this knowledge in their jobs and lives. Attitudes toward statistics are one of these crucial factors that can provide instructors with information about the effectiveness of different curricula and the approaches they use in teaching such courses, as a given approach might positively impact students' attitudes toward statistics (Vanhoof et al., 2011).

However, measuring students' attitudes toward statistics is only possible if instruments are available and if such instruments have multiple evidence of validity and reliability. One of the most used instruments is the SATS-36 (Schau, 2003). Several studies investigated the structure of this survey (Chiesi & Primi, 2009; Hommik & Luik, 2017; Persson et al., 2019; Saidi & Siew, 2019; Stanisavljevic et al., 2014; Tempelaar et al., 2007; Xu & Schau, 2019). In doing so, most of these studies relied on CFA, especially using parcels of items rather than individual items. Therefore, research using different approaches is needed to examine the structure and psychometric properties of this survey.

Rasch techniques have been used in the development and validation of several instruments that measure cognitive and affective constructs in science (Sabah et al., 2013; Sondergeld & Johnson, 2014), math education (Hidayat et al., 2021), statistics education (Teman, 2013), and other disciplines (Akour et al., 2021; Al-Thani et al., 2021; Hammouri et al., 2020). However, the Rasch model has not been used previously with SATS-36 data. Therefore, the current study aimed to confirm the psychometric properties (e.g., reliability of items and persons) and the dimensionality of the SATS-36 using RSM. More specifically, the objectives of the current study were, as follows:

1. Examining the structure and dimensionality of the SATS-36.
2. Examining the functioning of the SATS-36 items.
3. Examining the functioning of the rating scale utilized in the SATS-36.

METHODOLOGY

Research Design

This study was descriptive in nature; no given variable was manipulated. It described several aspects of the SATS-36, such as its dimensionality, its psychometric properties (person and item reliability), the spread of its items on the attitudes' continuum, and the functionality of the response categories used.

Participants

Participants in the current study were 423 students of counseling psychology, enrolled in an obligatory introductory course in statistics in the faculty of educational sciences at a large university in Jordan. The

sample included 380 (90%) female students and 43 male students. Data were collected towards the end of five consecutive semesters, starting from the second semester of the academic year 2017/2018. In all semesters, this course was taught by the same instructor, the author of the current study.

Participation was voluntary (convenience sample), which was the sole inclusion criteria set in identifying the study respondents. Students were free to withdraw at any time. Before participation, students were informed of the purpose of the study and the instructions for responding to the scale. They were also informed that their responses were confidential and would be used only for research purposes. All participants gave written informed consent.

Instrument: SATS-36

This instrument is composed of 36 items that measure six dimensions of attitudes toward statistics: affect (six items), cognitive competence (six items), value (nine items), difficulty (seven items), interest (four items), and effort (four items). All items incorporate a seven-point Likert response format (1="strongly disagree," 4="neither disagree nor agree," and 7="strongly agree," in which responses to negatively worded items are reverse coded. Students' responses on each dimension are combined to form component or subscale scores, such that higher scores on each subscale reflect more positive attitudes toward statistics. Higher scores on the difficulty subscale indicate that students believe that statistics is easier compared to those students with lower scores.

Sample items on the affect subscale are "I like statistics" and "I enjoy taking statistics courses." On the cognitive competence subscale, "I understand statistics equations" and "I can learn statistics." On the value subscale, "statistics is worthless" and "statistics is irrelevant in my life." On the difficulty subscale, "statistics is a complicated subject" and "statistics is highly technical." On the interest subscale, "I am interested in using statistics" and "I am interested in learning statistics." On the effort subscale, "I studied hard for every statistics test" and "I completed all of my statistics assignments."

The Cronbach's alpha for scores on the affect subscale ranged between 0.80 to 0.89. On the cognitive competence subscale, between 0.77 to 0.88. On the value subscale, between 0.74 to 0.90. On the difficulty subscale, between 0.64 to 0.81 (Schau, 2003). On the interest subscale, between 0.80 to 0.84. On the effort subscale, between 0.76 to 0.81 (Emmioglu & Capa-Aydin, 2012).

Two versions of the SATS-36 are available. A pre-test version that can be administered at the beginning of a statistics course and a post-test version that can be administered towards the end. Both versions share the same questions, but with some changes in verb tense.

The current study utilizes the post-test Arabic version of this survey (Assarierh, 2013). Assarierh (2013) translated the SATS-36 into Arabic. Then, four experts evaluated the translation. The resulting Arabic version was then submitted to another translator for back translation into English. The original version was then compared with the back-translated version for possible discrepancies. The last version of the scale was pre-tested for clarity on a sample of 75 students, and no problems were reported in answering all items. In the current study, the Cronbach's alpha for the affect subscale was 0.79, for the cognitive competence subscale 0.77, for the value subscale 0.79, for the difficulty subscale 0.68, for the interest subscale 0.84, and for the effort subscale 0.80.

Data Analysis

Linacre (2022) recommended starting with one combined unidimensional analysis. The dimensions on the SATS-36 might not be six substantively different dimensions; rather, they might be strands. Therefore, RSM analysis was conducted via the Winsteps (v. 3.57.2) computer program (Linacre, 2005) on the scores from items on all six dimensions, as follows:

1. To evaluate the dimensionality of the SATS-36, principal component analysis of the residuals (PCAR) and fit statistics were used. When using PCAR, we look at the size of the first contrast (the first component in the correlation matrix of the residuals after conducting principal component analysis). If the eigenvalue of the first contrast is less than three, then this contrast is just random noise. If not, the patterns of the loadings of this contrast might indicate the presence of a second dimension. Fit statistics (infit and outfit mean squares) are used to examine the fit of the data to the Rasch model. If all values range between 0.5 and 1.5, the uni-dimensionality of the SATS-36 would be supported (Boone et al., 2014).
2. To evaluate the functioning of the items, Wright maps, separation indices, and reliability indices were used. Wright maps display item and person measures on the same logit scale to evaluate item placement and targeting. On the other hand, person separation is used to classify respondents, while using item separation to verify the item hierarchy. Separation can range from zero to infinity, with higher values better than lower ones. Person and item reliabilities vary from 0.00 to 1.00, with higher values considered better. Person separation less than two and person reliability less than 0.80 indicate that the instrument may not be sensitive enough to distinguish between high and low performers, and therefore more items are needed. Item separation less than three and item reliability less than 0.90 indicate that the respondents are not

enough to confirm the item difficulty hierarchy (Linacre, 2022).

3. To evaluate the functioning of the rating scale, the following four guidelines (Linacre, 2002) were used:

- (1) at least 10 observations for each rating scale category,
- (2) outfit mean squares less than 2.0,
- (3) average measures increasing monotonically with each rating scale category, and
- (4) step calibrations advancing monotonically with the categories.

RESULTS

This study examined the efficiency and psychometric properties of the SATS-36 using RSM. The findings are presented below in three parts according to the objectives of the study.

Evaluating the Dimensionality of the SATS-36

First: Infit and outfit statistics

To evaluate the dimensionality of the SATS-36, infit and outfit statistics for each item were computed and presented in **Table 1**.

Table 1 shows that item difficulty measures ranged between -1.41 for item E14 "I studied hard for every statistics test," which was the easiest item to agree with, to 1.73 for item D24 "learning statistics requires a great deal of discipline," which was the hardest item to agree with. All infit and outfit statistics fell within the acceptable range (0.5-1.5), except for item 11 "I have no idea of what's going on in this statistics course," where the infit statistics was 1.54. This indicates that there exists 54% more variation in the observed data than the Rasch model predicted. This could happen when a person with a higher level of attitudes toward statistics did not agree with items that are easy to agree with, or when a person with a lower level of attitudes toward statistics agrees with items that are hard to agree with. Given that this value was slightly above 1.5, this item was kept for further analyses. All values of infit and outfit statistics supported that this scale measures a unidimensional construct, which is attitudes toward statistics.

Second: Principal component analysis of the residuals

A PCAR was also conducted as another method of assessing dimensionality. If data fit the Rasch model, then all variance in the data is explained by the latent factor of attitudes toward statistics and what is left in the data, or the residuals, are just random noise. However, if a substantial dimension is identified using PCAR a separate measure for that dimension should be created and Rasch analysis should be done separately for that dimension (Linacre, 2022).

Table 1. Item logit measures, fit statistics, and point measure correlations for the SATS-36 items

Subscale	Item number	M	IMSNQ	OMNSQ	PMC
Affect	3. I like statistics.	-0.19	0.91	0.90	0.65
	4. *I feel insecure when I have to do statistics problems.	0.13	1.07	1.12	0.54
	15. *I get frustrated going over statistics tests in class.	0.63	1.40	1.40	0.60
	18. *I am under stress during statistics class.	-0.75	1.31	1.24	0.52
	19. I enjoy taking statistics courses.	-0.22	0.82	0.78	0.71
Cognitive competence	28. *I am scared by statistics	0.28	1.23	1.21	0.70
	5. *I have trouble understanding statistics because of how I think.	-0.28	0.95	0.91	0.64
	11. *I have no idea of what's going on in this statistics course.	-1.37	1.54	1.41	0.41
	26. *I make a lot of math errors in statistics.	0.76	1.01	1.02	0.58
	31. I can learn statistics.	-0.73	0.63	0.60	0.63
Value	32. I understand statistics equations.	-0.80	0.82	0.79	0.60
	35. *I find it difficult to understand statistical concepts.	0.12	0.83	0.84	0.70
	7. *Statistics is worthless.	-0.81	1.07	1.03	0.52
	9. Statistics should be a required part of my professional training.	0.24	1.05	1.03	0.54
	10. Statistical skills will make me more employable.	-0.01	0.89	0.89	0.49
	13. *Statistics is not useful to the typical professional.	-0.02	0.83	0.85	0.53
	16. *Statistical thinking is not applicable in my life outside my job.	0.41	0.90	0.96	0.40
	17. I use statistics in my everyday life.	0.95	1.07	1.08	0.39
	21. *Statistics conclusions are rarely presented in everyday life.	0.79	0.93	1.00	0.33
	25. *I will have no application for statistics in my profession.	-0.14	1.08	1.09	0.49
Difficulty	33. *Statistics is irrelevant in my life.	0.49	0.87	0.90	0.45
	6. Statistics formulas are easy to understand.	0.02	0.80	0.82	0.60
	8. *Statistics is a complicated subject.	-0.05	0.84	0.83	0.67
	22. Statistics is a subject quickly learned by most people.	0.64	1.23	1.27	0.29
	24. *Learning statistics requires a great deal of discipline.	1.73	1.31	1.38	0.18
	30. *Statistics involves massive computations.	0.55	0.89	0.89	0.54
Interest	34. *Statistics is highly technical.	1.06	1.20	1.26	0.24
	36. *Most people have to learn a new way of thinking to do statistics.	1.24	1.13	1.21	0.18
	12. I am interested in being able to communicate statistical information to others.	-0.07	0.92	0.93	0.52
	20. I am interested in using statistics.	0.24	0.80	0.82	0.62
	23. I am interested in understanding statistical information.	-0.08	0.71	0.69	0.70
Effort	29. I am interested in learning statistics.	-0.08	0.94	0.89	0.62
	1. I completed all of my statistics assignments.	-0.99	1.05	0.98	0.42
	2. I worked hard in my statistics course.	-1.00	1.10	1.06	0.33
	14. I studied hard for every statistics test.	-1.41	1.33	1.23	0.31
	27. I attended every statistics class session.	-1.32	1.37	1.36	0.24

Note. M: Measure; IMSNQ: Infit MSNQ; OMSNQ: Outfit MSNQ; PMC: Point measure correlation; & *Negatively worded items

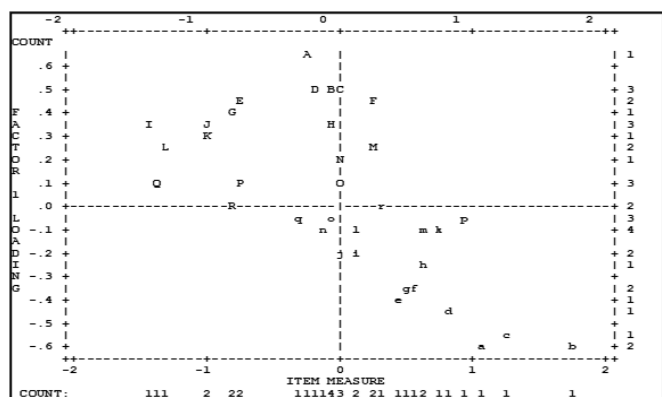


Figure 1. Location of each item of SATS-36 as a function of residual factor loadings (*y*-axis) and item measures (*x*-axis) (Source: Author's own elaboration)

The first contrast had an eigenvalue of 4.1, which indicates that all items on the SATS-36 may not define a single trait. Therefore, the next step was to examine the plot that provides the location of each item as a function of residual factor loadings (*y* axis) and item measures (*x* axis), as depicted in **Figure 1**.

In inspecting **Figure 1** we look for item groups that share the patterns of unexpectedness. If item groups do share patterns, then these items might be a second variable (Boone & Staver, 2020). Therefore, the next step was to look at the content of the items at the top part of **Figure 1** labeled with capital letters, and those at the bottom part labeled with small letters. If these two groups of items have different content, then they might be considered as composing different variables, and thus they should be split into separate analyses. However, if these two groups of items share common content then they are considered as parts of the same dimension or are located at two different ends of one continuum or one variable, and thus all items should go into one analysis (Linacre, 2022).

Items in the top part of **Figure 1** labeled with letters A to Q were items 19, 23, 29, 3, 31, 20, 32, 12, 14, 1, 2, 27, 9, 10, 6, 18, and 11. When examining the content and type of these items, it was found that three of these items (19, 3, and 18) belong to the affect subscale, all items of the interest subscale (items 23, 29, 20, and 12), all items of the

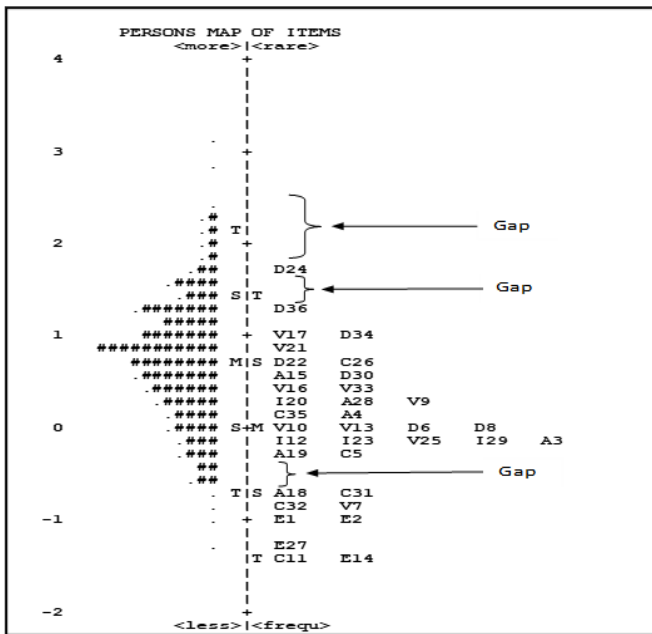


Figure 2. Wright (person-item) map of the SATS-36 (D: Difficulty, V: Value, C: Cognitive competence, A: Affect, I: Interest, and E: Effort) (Source: Author’s own elaboration)

effort subscale (items 14, 1, 2, and 27), three items (31, 32, and 11) belong to the cognitive competence subscale, two items (9 and 10) belong to the value subscale, and one item (item 6) belongs to the difficulty subscale.

Similarly, items appeared in the bottom part of **Figure 1** labeled with letters a to q were items 34, 24, 36, 21, 16, 30, 33, 22, 4, 13, 26, 35, 15, 25, 8, 17, and 5. When examining the type of these items, it was found that six of the items (34, 24, 36, 30, 22, and 8) belong to the difficulty subscale, six items (21, 16, 33, 13, 25, and 17) belong to the value subscale, three items (26, 35, and 5) belong to the cognitive competence subscale, and two items (4 and 15) belong to the affect subscale.

Looking back at the type and content of those items at the two parts of **Figure 1**, no distinct type of items were clustered in either part of the graph. However, we can see that almost all positively worded items were at the top part of **Figure 1** and almost all negatively worded items were at the bottom part of this figure. No clear difference in the item themes was found. After that, the analysis was conducted four times using simulated data. Eigenvalues of the first contrast were 1.6, 1.5, 1.6, and 1.7 respectively. All these eigenvalues were below two. This indicates that all items are part of the same dimension (Linacre, 2022) and thus, define a single trait which is attitudes toward statistics.

Evaluating the Functioning of the SATS-36 Items

First: Wright maps

A Wright map is a graphical representation of item difficulties and person abilities showing on a common logit scale.

Figure 2 displays this map for the SATS-36 data. The distribution of students’ attitudes toward statistics levels is displayed on the left side of the map, while the items’ difficulty levels are displayed on the right side of the map.

Students appeared toward the top of the map represent more agreeable students or students with higher levels of attitudes toward statistics, while those at the bottom represent students with lower levels of attitudes. Items appeared toward the top of the map represent items with higher logit values which are harder to agree with, while those appeared at the bottom are the easiest to agree with.

On the item side of the map, the “M” represents the location of the average difficulty of all items, while the “M” plotted on the person side represents the location of the average person ability of all respondents. The “S” represents one standard deviation, and the “T” represents two standard deviations (Boone & Staver, 2020). Items that are at the same location as a person have a 50% chance of being agreed with by that respondent. Items below a person’s measure have a greater probability of being agreed with, while those items above a person’s measure have a lower probability of being agreed with by that respondent.

Various parts of the Wright map can be useful in examining the measurement functioning of the SATS-36. One such benefit is to compare the location of the average difficulty of all SATS-36 items (M=0.0 logits) and the average person ability of all respondents (M=0.76 logits). Given that the mean person measures is higher than the mean item measures, the survey items were, generally, easier to agree with. Next, the difference between mean item measures and mean person measures is computed. Since there is less than 1.00 logit difference between these two means (difference=0.76 logits), this indicates a good survey item targeting or that items were relatively closely aligned with the respondents (Boone et al., 2014). This means that there are not too many items that were easy for students to agree with, and there are not too many items that were hard for students to agree with.

Another thing to look at in the Wright map is the distribution of the survey items along the continuum. When the items place on distinct locations of the trait continuum, or there are limited regions on the continuum where there are no items, this is an indication of a well-functioning instrument. The Wright map for the SATS-36 proved that this survey is a well-functioning one. However, some items (nine items) were located below person measures of many students. Those items are all items of the effort subscale (items 1, 2, 14, and 27), one item from the value subscale (item 7: Statistics is worthless), one item from the affect subscale (item 18: I am under stress during statistics class), and three items from the cognitive competence subscale

(item 11: I have no idea of what's going on in this statistics course; item 31: I can learn statistics; and item 32: I understand statistics equations). These items were extremely easy to agree with by students.

Moreover, several items (items V33 and V16, items V10 and V13, items D6 and D8, items I12, I23, and I29, and items E1 and E2) shared the same location on the continuum, indicating that these items may be redundant. One could delete one or two of these items or revise the content and wording of each item to improve the efficiency of this instrument. Based on the Wright map, items V16 "statistical thinking is not applicable in my life outside my job" and item V33 "statistics is irrelevant in my life" were placed on the same location. From the point of view of students, it seems that they did not make any distinction between statistics and statistical thinking, and thus both items share common content, which is the irrelevance of statistics in student's life. Therefore, deletion of one of these two items is recommended. Given that item V33 is shorter, and it asks about attitudes toward statistics in general, then it is recommended to retain this item.

Moreover, two items that belong to the value subscale were also located at the same position. That is, item V10 "statistical skills will make me more employable" and item V13 "statistics is not useful to the typical professional." Both items consult the importance of statistics in future career. However, what is more important in any profession is the skills that students acquire from any course. Given that item V10 taps this issue and that item V13 is more general in content and are negatively worded, it is recommended to delete item V13 and retain item V10.

Regarding the difficulty subscale, two items (item D6 and item D8) shared the same location on the continuum. Item D6 "statistics formulas are easy to understand" measures attitudes toward statistics formulas, whereas item D8 "statistics is a complicated subject" measures attitudes toward statistics in general. Even though item D8 is negatively worded, retaining this item and deletion of item D6 is recommended, given that the content of item D6 is part of the content of item D8.

Furthermore, three items from the interest subscale shared the same location: Item I12 "I am interested in being able to communicate statistical information to others," item I23 "I am interested in understanding statistical information," and item I29 "I am interested in learning statistics." Items I12 and I23 are related to statistical information, and thus students had the same chance of agreeing to both items. It is recommended to either delete item I12 or, due to the small number of items in this subscale, to rephrase the three items to remove the common content, and thus to make them more distinguishable in tapping distinct parts of the interest component.

The last two items that shared the same location were items E1 "I completed all of my statistics assignments" and E2 "I worked hard in my statistics course" that belong to the effort subscale. It seems that students viewed the completion of all assignments as evidence of working hard for the course. However, the other two items in this subscale (studying hard for the exams and attending classes) are also indicators of working hard for the course. Therefore, it is recommended to remove the more general item (item E2) and retain the remaining items. Given that this subscale has only four items, altering the content and wording of each item is recommended so that students would have different degrees of agreeability to them.

Finally, when looking back at the Wright map several parts of the map do not have items. These gaps should be filled with items. Therefore, it is recommended to fill the gaps (above item D24, between items D24 and D36, and between items A18 and A19) with new items to improve the overall precision of the SATS-36.

Second: Item and person reliabilities

Item and person reliabilities were excellent. Item separation was 11.22, and item reliability was 0.99. These high values of item separation (>3) and item reliability (>0.90) indicate that the sample of participants in the current study confirmed the item difficulty hierarchy (or construct validity) of the SATS-36. In other words, high item reliability means that there is a high probability that items classified as hard to agree were, actually, harder to agree with than those items that were classified as easy to agree with.

Additionally, person separation was 3.02 and person reliability was 0.90. These high values of person separation (>2) and person reliability (>0.80) indicate that the SATS-36 is sensitive enough to distinguish between students with higher and lower levels of attitudes toward statistics. This means that there is a high probability that students classified with higher levels of attitudes toward statistics were, in fact, had higher levels of attitudes than students classified with low levels of attitudes toward statistics. Therefore, no more items are needed.

Evaluating the Functioning of the Rating Scale

Linacre (2002) provided several guidelines to judge the effectiveness of rating scales. The first guideline is the existence of at least ten observations for each rating scale category. The existence of categories with few responses do not allow for stable estimation of the parameters. This may indicate unneeded categories, and thus, collapsing adjacent categories into one category with higher number of responses is proposed. Percentages of observations for each rating scale category are presented in **Table 2**.

Table 2. Summary of the rating scale functioning of the SATS-36

Category label	Observed percentage	Infit MNSQ	Outfit MNSQ	Observed average	Step calibration
1=SD	4	1.40	1.46	-0.19	None
2	6	1.13	1.17	-0.20*	-0.67
3	12	0.96	0.94	-0.01	-0.77*
4=N	20	0.93	0.87	0.15	-0.41
5	21	0.97	0.88	0.41	0.28
6	15	1.03	1.04	0.77	1.00
7=SA	22	0.88	0.91	1.26	0.57*

Note. SD: Strongly disagree; N: Neither disagree nor agree, SA: Strongly agree; observed percentage is the percent of all respondents who selected a given category; infit and outfit MNSQ is the average of the infit, or outfit mean squares for the responses in each category; observed average is the average of measures across all observation in each category; step calibration is calibrated measure of transition from category below to this category; & gap is difference between adjacent step calibrations

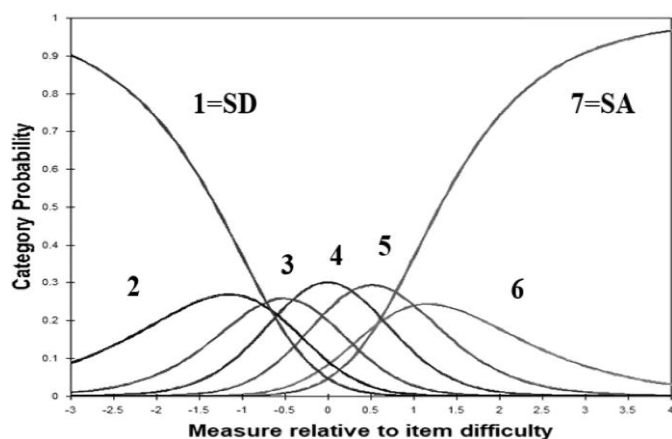


Figure 3. Category probability curve for the SATS-36 (Source: Author's own elaboration)

Table 2 shows that the response category "strongly disagree" had the least observed percentage with 4% of total observations (total observations or total observed count=number of items*number of students). This count far exceeds the required 10 observations.

Another guideline used to assess the functionality of the SATS-36 response categories is that outfit MNSQ should be less than two. Any category with outfit measure exceeding two indicates that this category is introducing more misinformation than information, or more noise into the analysis (Linacre, 2022). **Table 1** illustrates that outfit MNSQ values for all response categories fell below two, with the largest value of 1.46 corresponding to the "strongly disagree" category.

The third guideline is that average measures increase monotonically with each rating scale category. This means that students with higher levels of attitudes toward statistics are expected to agree with items that are hard to agree with. **Table 2** shows that observations in higher categories are produced by higher measures, except for category 2 with average measure of -0.20 which does not manifest higher levels of attitudes than category 1 with average measure of -0.19. Linacre (2002) suggested combining non- or barely advancing categories with those below them. Therefore, it is suggested to combine category 2 with category 1 to obtain a clearly monotonic structure.

The final guideline is that step calibrations advance monotonically with the categories. Step ordering means that as measures increase, each response category must have higher probability of being chosen. Failure of adherence to this guideline is referred to as "step disordering." In **Table 2**, disordering of step calibrations are indicated by "*". The findings in **Table 2** related to this guideline can be better understood when combined with the findings from **Figure 3**, which represents the probability curve for each category. When the rating scale is well functioning, the probability curve would have distinct hills or peaks where each response category is the most probable or modal.

Table 2 shows that the step calibration from category 2 to category 3 was -0.77 logits, which is the point in **Figure 3**, where the probability curves for categories 2 and 3 cross at the left side of the plot. The peak of the curve of category 2 does not appear as a distinct hill. That is, category 2 was never the most likely category to be observed at any point on the variable.

Similarly, the step calibration from category 6 to category 7 was 0.57 logits, which corresponds to the point in **Figure 3**, where the probability curves for categories 6 and 7 cross at the right side of the plot. The peak of the curve of category 6 does not appear as a distinct hill. Thus, distinct hills and ordering of step calibrations would occur if categories 2 and 3 were combined in one category, and categories 6 and 7 were also combined in one category. Thus, the proposed number of categories would be five instead of seven, which labeled as strongly disagree, disagree, neither disagree nor agree, agree, and strongly agree.

DISCUSSION

Introductory statistics courses are a vital part of every discipline in higher education due to the importance of statistical thinking and statistical skills in the professional life of university students. Therefore, it is necessary to foster students' understanding of statistics and, consequently, their achievement in such courses, and consider several factors that might contribute to this enhancement.

Past research (Chiesi & Primi, 2010; Ramirez et al., 2012) revealed that students' attitudes toward statistics are a critical factor that influences students' performance in statistics courses, and thus they need accurate measurement. SATS-36 is widely used in measuring attitudes toward statistics, as such, it is important to ensure that its items accurately assess the attitudes.

The purpose of the current study was to contribute to the improvement of this survey using RSM. The data in the current study fitted RSM. Fit indices fell within the acceptable range (0.5-1.5), and the results of the PCAR supported that this survey measured one construct, which is attitudes toward statistics. Additionally, high values of item reliability (0.99) and person reliability (0.90) reflected the excellent psychometric properties of this survey. Moreover, the Wright map showed that the SATS-36 is functioning well. All items were closely aligned with the respondents, meaning that there were not too many items that were easy for students to agree with, and there were not too many items that were hard to agree with. Additionally, response categories were well-functioning as each category had more than 10 observations, and outfit statistics were all low. However, an in-depth examination of the functionality of the SATS items and the response categories revealed some issues that need remedy.

Possible Scale Improvements

The first improvement deals with all items located below the person measures of, almost, all respondents. The SATS-36 was designed to assess six components of attitudes toward statistics. The items for each component were developed to assess the attitudes toward statistics that are easier and harder to agree with. Items on the effort subscale were anticipated to be easier for students to agree with as they tap routine tasks that every student does in every course (attending class, studying hard, and doing assignments). This anticipation was supported by the findings from the Wright map. However, these items were much easier than anticipated because they fall far below the ability levels of students with low levels of attitudes toward statistics. Moreover, some other items were located below the person measures of many students: three items from the cognitive competence subscale, one item from the value subscale, and one item from the affect subscale. These items were extremely easy for students to agree with. It is suggested to revise the wording of all these items to make them harder to agree with to help shift the mean items upward closer to the mean persons, and to allow these items to tap distinct locations on the trait continuum.

The second possible improvement to the scale is to maximize parsimony by removing items redundant in both content and difficulty (i.e., items placed on the same location on the item difficulty continuum). Xu and Schau (2019) asserted that more research is needed to decide

about item deletion. According to the findings of the current study, deleting the following three items is recommended: two items from the value subscale (items V13 and V16) and one from the difficulty subscale (item D6). As a result, the SATS-36 would contain 33 items instead of 36. This item deletion is not supposed to affect the content validity of each subscale due to the considerable number of items in each one (nine items in the value subscale and seven in the difficulty subscale).

Furthermore, three items of the interest subscale were located at the same position on the trait continuum. One probable reason for this is what Xu and Schau (2019) referred to as method effects. Each item in this subscale contains the word "interest." However, instead of deleting any item from this subscale, due to the small number of items in this subscale, revising the wording of each item is suggested to spread out these items on the continuum on different ability levels (low, medium, and high).

The third improvement is to add more items to fill the gaps in the trait continuum to ensure that all parts of the attitudes toward statistics are well-assessed. More items are needed toward the upper part of the continuum (items that are harder to agree with) to tap the locations against respondents who have higher levels of attitudes toward statistics.

The fourth and final improvement is related to the number of response categories utilized in this survey. The functionality analysis of the seven-point rating scale supported the combination of the two upper and the two lower response categories resulting in a five-point rating scale, instead. This would help students better decide which response for a given item to select and reduce the time needed to complete this survey, which is a vital goal in the administration of self-reported scales.

CONCLUSIONS

The current study evaluated the SATS-36 using RSM. Specifically, the structure and dimensionality of the survey, reliability of persons and items, quality of the items, and quality of the response categorizations were evaluated. The findings revealed that the survey is unidimensional and has excellent item and person reliabilities. However, using RSM helped in proposing some modifications to the SATS-36. These empirically driven improvements were modifying some items, deleting some others, and changing the number of response categories.

The findings have some implications for researchers and practitioners. Researchers and anticipated users of the SATS-36 can feel more confident in using this survey in measuring attitudes toward statistics, in terms of the excellent psychometric properties of this survey. On the other hand, the findings of the current study imply that no instrument is immune from criticism. The number and wording of the SATS-36 items, in addition to the

rating scale utilized, might be susceptible to change in future research. Therefore, users of the survey should keep track of the possible modifications that could take place on this survey to be able to use its latest version, which would have cumulative evidence of validity and reliability, to measure the construct of attitudes toward statistics more accurately.

Limitations and Future Research

One limitation of the current study is the administration of the SATS-36 to students from only one discipline. Even though RSM provides item parameters independent of the specific sample of respondents, it is recommended to replicate this study with different and larger samples from a wider range of disciplines to obtain more generalizable results. Another limitation in examining the functionality of the SATS-36 is the small number of male students who participated in the current study. Therefore, it is recommended to examine the functionality of this survey when administered to a larger sample containing comparable percentages of male and female students. Moreover, future research could assess the measurement invariance of this instrument across multiple important variables such as the gender of respondents or the time of administration (pre- vs. post-test data). Finally, it is recommended to collect data using two different numbers of response categories, five and seven, to empirically compare the effectiveness of the seven-point rating scale against that with a smaller number of categorizations.

Author notes: The author has agreed with the results and conclusions.

Funding: No funding source is reported for this study.

Ethical statement: The authors stated that the study was approved by the Psychological Research Ethics Subcommittee on February 4, 2018 with project number 18-02-11. Participation was strictly voluntary, and students were free to withdraw at any time. Authors further stated that the students were informed of the purpose of the study and the instructions for responding to the scale before participation. Their responses were kept confidential and would be used only for research purposes. Written informed consents were obtained and documented from all participants.

Declaration of interest: No conflict of interest is declared by the author.

Data sharing statement: Data supporting the findings and conclusions are available upon request from the author.

REFERENCES

- Abbiati, N., Fabrizio, M., Lopez, M., Perez, A., Plencovich, M., & Cueto, G. (2021). Attitudes related to students' performance in statistics in university programs in Argentina. *Statistics Education Research Journal*, 20(2), 8. <https://doi.org/10.52041/serj.v20i2.356>
- Akour, M. M., Hammouri, H., Sabah, S., & Alomari, H. (2021). Is it suitable to use the same categorization in rating scales when applied to students with distinctive levels of achievement? *Practical Assessment, Research, and Evaluation*, 26, 18. <https://doi.org/10.7275/21246182>
- Al-Thani, A. M., Akour, M. M., & Al-Kohji, S. M. (2021). Psychometric evaluation of an Arabic red flag screening checklist of autism spectrum disorders in infants and toddlers. *Research in Autism Spectrum Disorders*, 82, 101727. <https://doi.org/10.1016/j.rasd.2021.101727>
- Andrich, D. (1978). A rating formulation for ordered response categories. *Psychometrika*, 43(4), 561-573. <https://doi.org/10.1007/BF02293814>
- Assarierh, R. (2013). Construct validity of an Arabic version of the survey of attitudes toward statistics SATS-36. *Journal of College of Education: Ain Shams University*, 37(3), 650-672.
- Bechrakis, T., Gialamas, V., & Barkatsas, A. N. (2011). Survey of attitudes toward statistics (SATS): An investigation of its construct validity and its factor structure invariance by gender. *International Journal of Theoretical Educational Practice*, 1(1), 1-15.
- Bond, T. G., & Fox, C. M. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences*. Routledge. <https://doi.org/10.4324/9781315814698>
- Boone, W. J., & Staver, J. R. (2020). *Advances in Rasch analyses in the human sciences*. Springer. https://doi.org/10.1007/978-3-030-43420-5_21
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). *Rasch analysis in the human sciences*. Springer. <https://doi.org/10.1007/978-94-007-6857-4>
- Cashin, S. E., & Elmore, P. B. (2005). The survey of attitudes toward statistics scale: A construct validity study. *Educational and Psychological Measurement*, 65(3), 509-524. <https://doi.org/10.1177/0013164404272488>
- 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. <https://doi.org/10.1016/j.lindif.2008.10.008>
- Chiesi, F., & Primi, C. (2010). Cognitive and non-cognitive factors related to students' statistics achievement. *Statistics Education Research Journal*, 9(1), 6-26. <https://doi.org/10.52041/serj.v9i1.385>
- Coetzee, S., & Van der Merwe, P. (2010). Industrial psychology students' attitudes towards statistics. *South African Journal of Industrial Psychology*, 36(1), 1-9. <https://doi.org/10.4102/sajip.v36i1.843>
- DiStefano, C., & Jiang, N. (2020). Applying the Rasch rating scale method to questionnaire data. In M. S. Khine (Ed.), *Rasch measurement: Applications in quantitative educational research* (pp. 31-46). Springer. https://doi.org/10.1007/978-981-15-1800-3_3

- Emmioglu, E., & Capa-Aydin, Y. (2012). Attitudes and achievement in statistics: A meta-analysis study. *Statistics Education Research Journal*, 11(2), 95-102. <https://doi.org/10.52041/serj.v20i2.356>
- Evans, B. (2007). Student attitudes, conceptions, and achievement in introductory undergraduate college statistics. *The Mathematics Educator*, 17(2), 24-30.
- Hammouri, H., Akour, M. M., & Sabah, S. (2020). Evaluation of students' noncognitive large scale assessments measures: Problem solving experiences scale as a case study. *Opcion*, 36(26), 3115-3141. <https://produccioncientificaluz.org/index.php/opcion/article/view/32635/34123>
- Hidayat, R., Idris, W. I. W., Qudratuddarsi, H., & Rahman, M. N. A. (2021). Validation of the mathematical modeling attitude scale for Malaysian mathematics teachers. *EURASIA Journal of Mathematics, Science and Technology Education*, 17(12), em2047. <https://doi.org/10.29333/EJMSTE/11375>
- 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. https://doi.org/10.1207/S15328007SEM1101_7
- Hommik, C., & Luik, P. (2017). Adapting the survey of attitudes towards statistics (SATS-36) for Estonian secondary school students. *Statistics Education Research Journal*, 16(1), 228-239. <https://doi.org/10.52041/serj.v16i1.229>
- Linacre, J. M. (2002). Understanding Rasch measurement: Optimizing rating scale category effectiveness. *Journal of Applied Measurement*, 3(1), 85-106.
- Linacre, J. M. (2005). *Winsteps Rasch measurement computer program* (3.57.2). <https://www.winsteps.com/>
- Linacre, J. M. (2022). *A user's guide to WINSTEPS MINISTEP Rasch-model computer programs*. <https://archive.org/details/B-001-003-730>
- Nolan, M. M., Beran, T., & Hecker, K. G. (2012). Surveys assessing students' attitudes toward statistics: A systematic review of validity and reliability. *Statistics Education Research Journal*, 11(2), 103-123. <https://doi.org/10.52041/serj.v11i2.333>
- Persson, I., Kraus, K., Hansson, L., & Wallentin, F. Y. (2019). Confirming the structure of the survey of attitudes toward statistics (SATS-36) by Swedish students. *Statistics Education Research Journal*, 18(1), 83-93. <https://doi.org/10.52041/serj.v18i1.151>
- Ramirez, C., Schau, C., & Emmioglu, E. (2012). The importance of attitudes in statistics education. *Statistics Education Research Journal*, 11(2), 57-71. <https://doi.org/10.52041/serj.v11i2.329>
- Sabah, S., Hammouri, H., & Akour, M. (2013). Validation of a scale of attitudes toward science across countries using Rasch model: Findings from TIMSS. *Journal of Baltic Science Education*, 12(5), 692-702. <https://doi.org/10.33225/jbse/13.12.692>
- Saidi, S. S., & Siew, N. M. (2019). Investigating the validity and reliability of survey attitude towards statistics instrument among rural secondary school students. *International Journal of Educational Methodology*, 5(4), 651-661. <https://doi.org/10.12973/ijem.5.4.651>
- Sarikaya, E. E., Ok, A., Capa-Aydin, Y., & Schau, C. (2018). Turkish version of the survey of attitudes toward statistics: Factorial structure invariance by gender. *International Journal of Higher Education*, 7(2), 121-127. <https://doi.org/10.5430/ijhe.v7n2p121>
- Schau, C. (2003). Students attitudes: The "other" important outcome in statistics education. In *Proceedings of the Joint Statistical Meetings* (pp. 3673-3683).
- Schau, C., Stevens, J., Dauphinee, T. L., & Vecchio, A. Del. (1995). The development and validation of the survey of attitudes toward statistics. *Educational and Psychological Measurement*, 55(5), 868-875. <https://doi.org/10.1177/0013164495055005022>
- Sesé, A., Jiménez, R., Montaña, J. J., & Palmer, A. (2015). Can attitudes toward statistics and statistics anxiety explain students' performance? *Revista de Psicodidáctica [Psychodidactic Magazine]*, 20(2), 285-304. <https://doi.org/10.1387/RevPsicodidact.13080>
- Sondergeld, T. A., & Johnson, C. C. (2014). Using Rasch measurement for the development and use of affective assessments in science education research. *Science Education*, 98(4), 581-613. <https://doi.org/10.1002/sce.21118>
- 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), e112567. <https://doi.org/10.1371/journal.pone.0112567>
- Teman, E. D. (2013). A Rasch analysis of the statistical anxiety rating scale. *Journal of Applied Measurement*, 14(4), 414-434.
- Tempelaar, D. T., van der Loeff, S. S., & Gijsselaers, W. H. (2007). A structural equation model analyzing the relationship of students' attitudes toward statistics, prior reasoning abilities and course performance. *Statistics Education Research Journal*, 6(2), 78-102. <https://doi.org/10.52041/serj.v6i2.486>
- Vanhoof, S., Kuppens, S., Sotos, A. E. C., Verschaffel, L., & Onghena, P. (2011). Measuring statistics

attitudes: Structure of the survey of attitudes toward statistics (SATS-36). *Statistics Education Research Journal*, 10(1), 35-51. <https://doi.org/10.52041/serj.v10i1.354>

Xu, C., & Schau, C. (2019). Exploring method effects in the six-factor structure of the survey of attitudes toward statistics (SATS-36) 4. *Statistics Education Research Journal*, 18(2), 39-53. <https://doi.org/10.52041/serj.v18i2.139>

<https://www.ejmste.com>