The Effectiveness of the Number of Extra Negative Items in Detecting Insufficient Effort Responses in Non-Cognitive Scales

Manar Hazaimeh, Mahmoud Alquraan

Yarmouk University, Irbid, Jordan
Email: manarhazaimeh0@gmail.com, mahmoud.q@yu.edu.jo

Abstract

This study attempted to determine the effectiveness of the number of extra negative items in identifying insufficient effort responses in Attitudes toward Statistics (SATS-36). The (SATS-36), which consists of 36 5-point Likert Scale items, was used to actually achieve the goal of this study. Furthermore, the researchers developed three forms, each with a different set of extra negative items (2, 4, and 6). The three forms were distributed to a sample of (750) students at Yarmouk University. The results revealed that form 1, which enclosed 6 extra negative items, had the lowest detected rate of insufficient effort responses (IERs) (7.20%), while form 3, which contained 2 extra negative items, had the highest detection rate (15.6%). The detection rate was discovered respectively among the Lie Detection Scale, Mahalanobois, l_z^p and the method of extra negative items. The highest detection rate was found in form 3 with two extra negative items, and data reliability decreased after the insufficient effort responses were removed (IERs). Furthermore, the results showed that the maximum changes in model-data fit indices after data filtering were in form 3, which contained two extra negative items. Moreover, the results indicate that the reliability of data after filtering those with insufficient effort responses (IERs) is reduced.

Keywords

Insufficient Effort Responses, Extra Negative Items, Attitudes Toward Statistics (SATS-36)

1. Introduction

The significance and quality of the data obtained from the research population,

and thus the value of the data collection techniques utilized, assess the value of scientific research; however, with the steadily rising use of studies that focus primarily on survey methods throughout all disciplines, scholarly and informational, it is vital to guarantee the validity of the data.

The significance of accuracy and its tools arises from the great significance of the results deliberately designed for the outcome; incorrect or finding is inconsistent would then eventually result in unfair choices, will not yield results and would not attain the goal of the research, as well as the attempt would be lost; thus, the researcher could perhaps verify the correlations of insufficient effort responses to data collection tools before trying to adopt these data, derive their results, and make them a reference for decision-making.

There were numerous distinctions in the terminology of the kinds of non-genuine reactions/responses in every analyst/researcher contrasted with different specialists. It has been named by responders, arbitrary (Random Responders [1] [2], and the people who are not making sufficient effort to answer (Insufficient Effort Responding [3], respondents who are non-genuine or non-intrigued during the reaction, Careless Responders [4] and different titles that discussed the respondents who are not inspired/interested in their responses while applying their scale or read the guidelines prior to doing as such [5].

[3] clearly delineated responders and Careless Responders as participants who don't seem interested in responding to questions about non-cognitive indicators and, therefore, could differentiate this group from selected people through several things, such as arbitrary answers, answers attitudes, or deliberately misinterpreting the item before responding; on the other hand, arbitrary answer defined by [6]: as they are the responses that the responder selects and recognizes lacking care or attentiveness, and the researcher finds such answers as a negative factor and difficulty for him because they have a significant impact on the research findings and conclusions.

The existence of insufficient effort responses influences statistical analysis values such as numerical mediums, standard deviations, reliability and constancy features, and so on, resulting in biased values that cannot be accepted or adopted [7]. Furthermore, due to their inaccuracy, such responses present many challenges to some researchers. These responses influence the results of assessments and research, as well as disease diagnosis, psychiatric reports, and the classification of individuals in psychiatric clinics and medical. [8] have recommended the need to reduce the negative effects of the existence of insufficient effort responses, to achieve more accurate and reliable results by recognizing and revealing them in specific methods, and then to eliminate the data of the individuals who submitted such responses.

This research attempts to provide a report on different approaches for identifying insufficient effort responses, including a concentration on the method of including negative items as a practical and easy-to-implement strategy, as well as its coherence with other ways.

2. The Lie Detection Scale from Eysenck's List of Personality

The lie scale from Eysenck's list of personalities was used (E.P.I), and the scale included (9) items, the answer to which is (yes) or (no), so that the high score indicates the respondent's desire to choose socially acceptable answers, Eysenck chose/picked (5) grades on the scale of lying as the highest threshold/criterion for acceptance of the answer, that is, a value greater than 5 indicates that the respondent is a liar [9].

3. Mahalanobois

This measure is typically used to analyze the influence of irregular values on the values of regression analysis, and values are considered ambiguous if the Mahalanobis distance Square is greater than the standard value of the Chi-square distribution; that is, this method is used and calculated by the value of the level of significance of the Chai range/method, and this value is an evidence of the existence of individuals who have caused discrepancies, and this issue is addressed [10].

[4] also used the method in their research where they evaluated the significance of Mahalanobis as an indicator to detect statistically significant differences and thus detect insufficient effort responses (IERs) individuals and realized that this approach is a good technique to observe them and gives accurate results, but the disadvantage of this method is that it is heavily dependent on the nature of the sample examined and the distribution of individuals' responses on the scale, which causes difficulties.

4. Statistical Index/Indicator l_z^p

Statistical indicator l_z^p is a natural extension of the index statistical l_z)a statistical algorithm or standard (Standardized Log-likelihood-Statistic), Which would be a typically used indicator that represents a mix of response possibility and competency evaluation [11] the statistic l_z based on l_o that proposed by [12], the statistic l_o is the experimentally measured and computed probability of the response for an item based on the response theoretical model for an item.

The index l_z^p can be simply defined, it is a statistical indicator based on the super-probability estimation approach that compares the predicted pattern of response from the item's responses theoretical model to the individual responses on a multi-scale measurement. [13] indicated that significantly negative values of the statistical indicator signify a high level of inconsistency. It should be mentioned that the statistical l_z^p necessitates detailed statistical analyses and specific software; consequently, the statistical l_z^p for multi-step items and quasi measurements was recently added to the setting "R-Package" [10].

5. Negative Items

Elements of the scale items are frequently switched to reduce the influence of

stereotypical responses, while there seems to be little agreement that this method is successful in reducing stereotyped responses. There are two methods for modifying items: the first is to use negative terms including the first is to add negative words such as no, without, otherwise, I'm not...and so forth. In this aspect, the manner of the new item is changed besides modifying the choice of words. While the second strategy is to use words with opposite meanings, including feeling adaptable vs. feeling exhausted, in which case the direction of the new item is improved by modifying the choice of words. As shown by [14], the first technique was adopted in the majority of research studies. While [15] feels that employing negative items diminishes questionnaire integrity, researchers continue to utilize negative items with the idea that they minimize stereotyped responses. [16] developed a scale with five pairs of items written in opposite ways (e.g., "the professor wasn't ready to separate" vs. "the professor was ready to separate"), and discovered that using opposing items produced noise known as misspecification. According to [17], it is possible to identify the pattern of response in the scales by adding items with a converse formulation, under which they developed a scale with five parts, each containing twenty items, so that each area has four items: two positive items (e.g., "I feel tired"), and two negative items (e.g., "I'm not tired at all", or "I feel very active"), where the influence of compliance is indicated by the opposite formulation.

[18] conducted research employing negative items to determine the influence of negative items on the identification of insufficient effort responses (IERs) answers. It was discovered that the presence of negative items does not reduce the presence of insufficient effort responses and that the results were inaccurate due to respondents' lack of attention and confusion, leading the researchers based on this study to propose the use of extra negative items as a method of detecting insufficient effort responses, by calculating the: (strongly agree = 1, agree = 2, neutral = 3, disagree = 4, strongly disagree = 5), as well as the average of the equivalent/extra negative items with the same content has been determined, leads to the following degrees of the hierarchy of answers: (strongly agree = 5, agree = 4, neutral = 3, disagree = 2, strongly disagree = 1). Whereas if mean absolute difference between positively and extra negatively-worded items (MAD) after reverse coding the negative items and the equivalent negative items (having the same content) is higher than (2), the individual's response is insufficient effort responses (IERs) [18] [19] [20], and indeed the extreme difference used in this research is (1.6), as calculated using the following formulae/equation:

$$\frac{\text{Upper limit of Likert scale}(5) - \text{lower limit of Likert scale}(1)}{\text{Number of levels for Likert scale}(5)}$$

$$\frac{1-5}{5} = 0.8 \text{ per item}$$

Therefore, two criteria were measured: 0.8 + 0.8 = 1.6. For example, if the average of an individual's response to the added negative items is approximately equivalent to (2) and the average of the same individual's response to the corresponding positive items with the same content is equal to (4), the absolute dif-

ference is equal to (2) and greater than (1.6), indicating that the individual is an insufficient effort response (IERs).

So many of the previous studies followed the tradition of many academics and professionals in equalizing the amount of positive and negative items on the scale without depending on a theoretical or scientific basis to support this viewpoint. In particular, there was also some agreement that the presence of negative items has a negative impact, which has led several more studies to propose moving further away from the inclusion of scales negative items as part of the scale because it may cause inequity in the measurement of the same characteristic, as well as it may reduce the validity of the scale due to the inconsistency of individual responses to the items. Many researchers have conclusively proved that negative items on the scale can be used to detect insufficient effort responses (IERs) replies; nevertheless, two problems arise: are negative items parts of the scale or supplementary items? How many items should be added to this item?

6. Problem Statement

A study of the concept of utilizing negative items to detect insufficient effort responses (IERs) reveals a deficiency of research or approaches demonstrating how many negative items should be added to the scale to detect insufficient effort responses. Even though the purpose of this study is to determine the validity of the approach of extra negative items (2, 4, and 6) in detecting insufficient effort responses and their effect on the measurement scale the scale and the statistical characteristics of the average rating on the scale, it explicitly attempts to answer the following study questions:

Q1: "Does the percentage of insufficient effort responses detected differ with the number of extra negative items (2, 4, and 6)?"

Q2: "What is the percentage of consistency between the categorization on the technique of extra negative items (2, 4, and 6) and other ways to detect insufficient effort responses (IERs) (lie detection test scale, Mahalanobois, method of l_z^p?

Q3: "How does removing respondents with insufficient effort responses affect reliability depending on the number of extra negative items (2, 4, and 6)?"

Q4: "How does a removing insufficient effort response from respondents affect confirmatory factor analysis model-data fit depending on the number of extra negative items (2, 4, and 6)?"

7. Definition of Terms

Insufficient Effort Responses: [21] defined insufficient effort responses as the respondent's incapacity or reluctance to react to items with interest.

Negative items (Negative items): it is items that measure the invert about what evaluates the measurement in regarding language formulation, *i.e.*, it denies positive measurement items and is linguistically anti-them.

Extra negative items: Negative items were designated as a strategy for identifying insufficient effort responses (IERs) since they are utilized after the scale has

been developed by adding extra items with negative formulation corresponding to positive items in the scale.

Attitudes Toward Statistics: is a mental and spiritual concept that serves to direct an individual's behavior so that he responds positively or negatively to statistical data; that is, it reflects the student's degree of inclination or lack of inclination towards statistics and is measured by the sum of the marks scored by the student on the scale of trends used in the study towards statistics according to the Pentateuch's Likert method.

8. Methods

8.1. Participants

The research sample was chosen from 750 students from all scientific and humanitarian/academic faculties at Yarmouk University, where 460 questionnaires were distributed electronically by attaching three links, with each student choosing just one. As well as 290 throughout hard copies on Yarmouk University students within faculties which demand direct implementation using random Equivalent Samples, so that members of one division answer all three models in a typical way in which the first student answers the first model, the second student on the second model, the third student on the third model, and then repeat the pattern and the fourth student answers the first model, and so on for the rest of the division, and after collecting the participants' responses mostly on the scale. There were 250 students and students who decided to respond to the first model, 250 students and students who answered to the second model, and 250 students and students who agreed with the third model.

8.2. Measure

The metrics towards statistics scale (36-SATS) was adopted for the current study, which was done by [22], where the scale became six-dimensional-after it was made up of four variables. And (36) items in its final version are distributed on the six dimensions (Affect), and consists of (6) items focused on feelings towards statistics and after (Difficulty), and consists of (7) items measuring the feeling of students related to their treatment with symbols, formulas, and statistical equations and after (Value), and consists of (9) items include the benefit of statistics in practice, increasing employment opportunities, and the development of statistical equations and formulas. After cognitive efficiency (Competence): the dimension consists of (6) items and represents the students' metrics towards their self-competence and knowledge and mental skills during the application of Statistics and after the tilt (Interest): the dimension consists of (4) items and measures the student's degree of attitude towards the topics of Statistics and enjoy using it and appreciating after the effort (Effort): it consists of (4) items.

8.3. Validity and Reliability

[23] interpreted the scale (SATS-36) on the Jordanian environment, and he iso-

lated his passages into five stages. The scale was directed to an example of (445) students, and the consequences of the examination uncovered that the scale comprises six significant perspectives, and he viewed the consolidated factors (48. 41%) of the standard deviation of the scale, and he determined the reliability co-efficacy. [24] separated the exchanges of virtual trustworthiness/reliability by learning the respectability of the phrasing of the conditions of the scale and its appropriateness for the Arab environment, the age phase of the review test, and above all its reasonableness for the reason for which it was ready, where he introduced the scale in Arabic structure to eight judges from teachers spent significant time in estimation, assessment and instructive brain science, and it was observed that every one of the expressions of the scale is appropriate and sound and fit the Arab environment, and there were no extreme perceptions of any of its expressions. He also conducted experiments also with the scale on a survey sample of (56) students to ensure the appropriateness of the language for the scale phrases and to determine the appropriate application time of the scale, and observed that there were no problems in understanding the scale phrases, and revealed that the appropriate time for the application is (35) minutes, the same time that the scale was developed. The scale's dimensions, as well as the relationships between the aspects and the scale's overall score, are statistically significant at the level of ($0.05 = \alpha$). He also checked the instrument's stability, internal reliability, and size, all of which were satisfactory.

8.4. Explanation of Extra Negative Items/Adding Extra Negative Items (2, 4, and 6)

[19] highlighted the insertion of a negative item to each of the scale's (5 - 8) positive items. However, according to [25], regardless of the length of the original questionnaire, the usage of conceptual equivalents or contextual antonyms should be limited (two to four) pairs of synonymous or linguistically contrasting items. The statistical aptitude is assessed using (36) items, with a negative item added for every (5 - 8) positive items. As a consequence, six negative items were inserted, and two items were skipped. Three models are used:

Model one has six extra negative items that correlate with the scale's six positive items.

- The second model has four more negative items that correlate with roughly four positive items on the scale.
- The third model has two more negative items that correlate with approximately two positive items on the scale.

9. Data Analysis

To explain the research question(s), the package of (SPSS V26) (Statistical package for Social Sciences software Amos (AMOS V26) (Analysis of Moment Structures) of the statistical analysis of the data collected was utilized, as follows:

- *The percentage of non-responses that were effectively valued.*

Insufficient Effort Responses were classified as shown by the difference in the

number of Extra Negative Items (2, 4, 6) according to the criterion of calculating mean absolute difference between positively and extra negatively-worded items (MAD) after reverse coding the negative items, which contain the same content; if the difference is greater than (1.6), the individual's attitude was classified as insufficient effort responses (IERs). Chi-square (χ^2) has also been used in bilateral comparison of insufficient effort responses' frequencies/ratios connected with the number of negative items provided to identify them (2, 4, 6), where the approach that identifies the most insufficient effort responses (IERs) is recommended [26].

- *The percentage of consistency between the technique of extra negative items (2, 4, 6) and other ways (Lie detection scale/the lie scale, Mahalanobois, method Of l_z^p).*

Insufficient Effort Responses were assessed on the statistical metrics scale (SATS-36) based on the number of extra negative items (2, 4, 6) to identify careless respondents, as well as whether respondents were classified as insufficient effort responses (IERs) or not insufficient effort responses (IERs). Other techniques' responses were classified (a metric of lie detection, a technique of (Mahalanobois), the way l_z^p) in addition to utilizing Chi-square for stability and to compare the findings of compatibility between the categorization using the way of extra negative items (2, 4, and 6) against the other ways (a measure of lie detection, Mahalanobois, the method l_z^p .

- *Cronbach Alpha approach consistency coefficients:*

The Cronbach method alpha consistency coefficients of the items of the measure of metrics towards statistics (SATS-36) were determined before and after the elimination of Insufficient Effort Responses by altering the number of negative items provided (2, 4, 6).

- *Goodness-of-Fit Index of conformity factor analysis.*

Conformity Factor Analysis (CFA) was performed using the (AMOS V26) for the method of extra negative items (2, 4, and 6) in order to diagnose a model that gives the best value matching of three models—throughout the following indicators (Chi-Square brothers, absolute fit indicators, incremental conformity indicators) with different number of extra negative items in order to detect the response of insufficient effort responses (IERs) Furthermore.

According [27], the researcher made the following observations on model matching acceptability values in confirmatory factor analysis.

9.1. Chi-Square Criteria's/Standard Value (Chi-Square/DF = Cmin)

It is the value of the Chi-square derived from the model divided by the number of variables; if this ratio is less than (5), the model is accepted; if it is below (2), the recommended model is equal to the presumed model of the sample data.

9.2. Goodness of Fit Index (GFI)

The indicator of goodness consistency has a value ranging from zero to one,

and it determines the amount of variance in the Matrix resulting from the correlational statistical approach, which again is similar to the coefficient (F2) in data analysis, and the value (0.90) is the lowest acceptable value in this index.

9.3. Comparative Fit Index (CFI)

According to some researchers, the best value that reflects the model's effectiveness and compatibility with the data of the study sample is (0.90), and the closer the value to the correct one is, the better, as well as the value of the indicator (TKR - Los), which should be (0.90) or more/greater.

9.4. Root Mean Square Error of Approximation (RMSEA)

One such criterion is one of the most key criteria of comparability value; it illustrates the amount of error in the model as well as the rates of variance from the optimal standards; structural modelling studies have demonstrated its superiority and good performance; a value less than (0.05) indicates a good match, values ranging from (0.05 - 0.08) indicate a reasonable convergence error in society, and values ranging from (0.10 - 0.08) indicate a totally inadequate fit. If the indicator values surpass (0.10), it indicates a lack of conformance. We can deduce that a value of zero indicates the best possible match, whereas a value of greater than zero suggests a lesser quality and even worse fit.

10. Results and Discussion

To answer this question, responses were classified as insufficient effort responses (IERs) or not insufficient effort responses (IERs) using the extra negative items technique, with the criterion of calculating the mean absolute difference between positively and extra negatively-worded items (MAD) after reverse coding the negative items, which contain the same content, if the difference is greater/more than (1.6). Furthermore, after modifying the degree of tangible significance in light of the strategy for Bonferroni (Bonferroni Correction), which demonstrated that the degree of significance changes while making bilateral correlations the level of statistical significance was divided by the total number of bilateral variations, yielding the value of the level of statistical significance after correction (0.0166), and the number of bilateral comparisons in light of the number of percentages; to retain with the entire type inaccuracy; so the degree of factual significance was isolated by the quantity of Chi-square (χ^2) for relevant correlation of quasi reaction ratios according to the extra negative parts Technique (2, 4, 6), as shown in the table below:

Table 1 clearly demonstrates that the rates of insufficient effort responses identified differ significantly regardless of the method of extra negative items (2, 4, 6), with the first model including six extra negative items (7.20%), the second model including four extra negative items (7.60%), and the third model including two extra negative items (15.60%).

Table 1. Shows the degree of participants' responses based on the strategy of extra negative items (6, 4, 2).

Model	Classification	N	Percentage %
1st Model **(6 extra negative items)**	Not IER	232	92.80%
	IER	18	7.20%
	Total	250	100.00%
Model 2 **(4 extra negative items)**	Not IER	231	92.40%
	IER	19	7.60%
	Total	250	100%
Model 3 **(2 extra negative items)**	Not IER	211	84.40%
	IER	39	15.60%
	Total	250	100%

Model 1, which enclosed 6 extra negative items, had the lowest detected rate of insufficient effort responses (IERs) (7.20%), while Model 3, which contained 2 extra negative items, had the highest detection rate (15.6%). These findings are consistent with the earlier findings [28]. This study found that (0.8% - 20.3%) of insufficient effort responses (IERs) were valued using various methods, and we note that the ratios provided by the three models would be within the range of the ratios found in the study.

The percentage submitted by the third model (15.60%), which included two more negative items, corresponded with the findings of the study [29], which revealed that (16.2%) of respondents had a insufficient effort responses (IERs) reaction, which are two close/equally ratios.

Table 2 shows the measurement percentages of insufficient effort responses using the strategy of extra negative items (2, 4, 6). It involves the comparison of the probability value with a value of (0.0166); there was statistical significance between the first model, which includes six extra negative items, the third model, which includes two extra negative items, and the second model, which includes four extra negative items, and the third model, which includes two extra negative items, indicating that the highest detection rate for insufficient effort responses was on the second model. This is in line with [26], which found that an approach that conveys a higher number of insufficient effort responses is desirable.

The extra negative items approach could be used to identify insufficient effort responses in order to obtain more accurate outcomes, and it can serve as a support method for other methods of detecting insufficient effort responses as a method that is simple to use and utilize, especially if the researcher's attention is focused on identifying substantial amounts of insufficient effort responses.

To answer the second question, the percentage of insufficient effort responses who are compatible with the way of extra negative items (2, 4, and 6) and other procedures (Lie detection scale, Mahalanobois, method of l_z^p), **Table 3** shows that.

Table 2. Chi-square test findings (χ^2) for binary analyses of insufficient effort responses (IERs) ratios discovered using the extra negative items approach (6, 4, 2).

Model	1st Model (6 extra negative items)		Model 2 (4 extra negative items)	
	χ^2	Significance	χ^2	Significance
Model 2 (4 extra negative items)	0.029	0.8645		
Model 3 (2 extra negative items)	8.715	0.0032	7.786	0.0053

Table 3. Percentage of insufficient effort responses who are compatible with the way of extra negative items (2, 4, 6) and other procedures (Lie detection scale, Mahalanobois, method of l_z^p).

Method of negative items extra/model	Lie detection scale	Mahalanobois	Method of l_z^p
Model 1 (6 extra negative items)	10%	6.1%	9.8%
Model 2 (4 extra negative items)	9.8%	17.5%	13%
Model 3 (2 extra negative items)	20.7%	19.6%	23.2%

Table 3 revealed that the rate of compatibility for the classification of insufficient effort responses (IERs) respondents between the method of extra negative items (2, 4, 6) and the lie detection scale ranged from (9.8%) on the second model, which includes four extra negative items, to (20.7%) on the third model, which includes two extra negative items, and the percentage of compatibility for the classification of insufficient effort responses (IERs) respondents includes six extra negative items. As well as (19.6%) on the third model, which includes two extra negative items, and the percentage of agreement for the classification of insufficient effort responses (IERs) respondents between the method of extra negative items (2, 4, 6) and the method of in ranged from (9.8%) on the first model, which includes six extra negative items, to (23.2%) on the third model, which includes two extra negative items.

The latter values indicate that the method of extra negative items (2, 4, and 6) has the greatest consistency of the rating of insufficient effort responses (IERs) respondents with other techniques (Lie detection scale, Mahalanobois, method of l_z^p). The third model-which involved two extra negative items. This study implies that adding two negative items increases responsiveness to the negative items approaches more than adding four or six negative items. This could be because the arithmetic mean was used to calculate the difference between individuals' answers to the extra negative and positive items, and as is widely recognized, the arithmetic mean impacted by outliers, especially for small sample sizes. The usage of four and six items narrowed the gap between the two formats, which may have helped to keep some of the less insufficient effort responses (IERs) responses hidden/concealed.

To answer the third question, Cronbach Alpha coefficients for statistical items (SATS-36) before and after removing insufficient effort responses depend on the number of extra negative items (2, 4, and 6). Table 4 shows that.

Table 4 clearly demonstrates that the Cronbach Alpha coefficient values before removing the insufficient effort responses ranged from (0.930) on the second model, which includes four extra negative items to (0.957) on the first model—which includes six extra negative items and that the Cronbach Alpha coefficient values after eliminating the insufficient effort responses varied from (0.921) on the third model, which includes two extra negative items to (0.947) on the first model.

The above findings demonstrate that the Cronbach Alpha coefficient values declined on the items of the statistical trend scale (SATS-36) after the insufficient effort responses were eliminated by altering the number of negative items provided (2, 4, 6).

To answer the fourth question, model consistency indicators were obtained before and after the deletion of insufficient effort responses using the extra negative items approach" (2, 4, 6). Table 5 and Figure 1 shows that.

Table 5 demonstrates that the data fitting rating scales for the three models utilizing the extra negative items Method (2, 4, and 6) were rather satisfactory, presumably due to the short sample size and uneven response behavior on the scale. The variation in model fit values before and after eliminating insufficient effort responses using the extra negative items Method is displayed in Table 5. (2, 4, 6). Overall, the results show that the relevance of the data for the three models varies slightly once insufficient effort responses are eliminated using the extra negative items Method (2, 4, and 6). The Comparative Fit Index (CFI) decreased very fractionally in the second model, which involves four extra negative items, by (0.009), and increased in the first model, which includes six extra negative items, and the third model, which includes two extra negative items, respectively, by (0.030 and 0.006). According to [30], a (0.01) difference in the Comparative Fit Index (CFI) is a substantial difference.

Table 5 also illustrates that the standard χ^2/DF grew by (0.285) on the third

Table 4. Cronbach Alpha coefficients for statistical items (SATS-36) before and after removing insufficient effort responses depending on the number of extra negative items (2, 4, 6).

Model	Alpha Cronbach coefficient before removing (IERs)	Alpha Cronbach coefficient after removing (IERs)
Model 1 (6 extra negative items)	0.957	0.947
Model 2 (4 extra negative items)	0.930	0.927
Model 3 (2 extra negative items)	0.940	0.921

Table 5. Differences in Model fit indices before and after removing insufficient effort responses using the extra negative item's method (2, 4, and 6).

Model	Goodness-of-Fit Index	Value before removing (IERs)	Value after removing (IERs)	CFA Model-Data Fit Difference	ranges values of Fit Index	Threshold/ fitting value
		Chi-square = \cmin				
	χ^2/DF	2.706	2.529	0.177	0 to 5	<5
		Absolute Fit indexes				
	RMSEA	0.083	0.081	0.002	0.05 > RMSEA > 0.08	<0.05
Model 1 (6 extra negative items)	GFI	0.754	0.751	−0.003	0 > GFI > 1	>0.90
		Incremental Fit Indexes				
	NFI	0.757	0.758	0.001	0 > NFI > 1	>0.90
	CFI	0.830	0.836	0.006	0 > CFI > 1	>0.90
	IFI	0.832	0.838	0.006	0 > IFI > 1	>0.90
		Chi-square = \cmin				
	χ^2/DF	3.530	3.478	0.052	0 to 5	<5
		Absolute Fit indexes				
	RMSEA	0.101	0.104	−0.003	0.05 > RMSEA > 0.08	<0.05
Model 2 (4 extra negative items)	GFI	0.744	0.735	−0.009	0 > GFI > 1	>0.90
		Incremental Fit Indexes				
	NFI	0.646	0.635	−0.011	0 > NFI > 1	>0.90
	CFI	0.715	0.706	−0.009	0 > CFI > 1	> 0.90
	IFI	0.718	0.710	−0.008	0 > IFI > 1	>0.90
		Chi-square = \cmin				
	χ2/DF	1.981	2.266	−0.285	0 to 5	<5
		Absolute Fit indexes				
	RMSEA	0.090	0.078	0.012	0.05 > RMSEA > 0.08	<0.05
Model 3 (2 extra negative items)	GFI	0.692	0.763	0.071	0 > GFI > 1	>0.90
		Incremental Fit Indexes				
	NFI	0.582	0.645	0.063	0 > NFI > 1	>0.90
	CFI	0.730	0.760	0.030	0 > CFI > 1	>0.90
	IFI	0.738	0.765	0.027	0 > IFI > 1	>0.90

model, which includes two negative extra items, and that, however significant, it remains within the acceptable range of conformity. It is reduced by (0.052, 0.177) on the first model, which includes six extra negative items, and by (0.052, 0.177) on the second model, which includes four extra negative items. The standard chi-square values on the three models fitted adequately. As shown in **Table 5**, the value of the index of the root of the squares of residuals (RMSEA) increased very slightly for the second model, which includes four extra items by
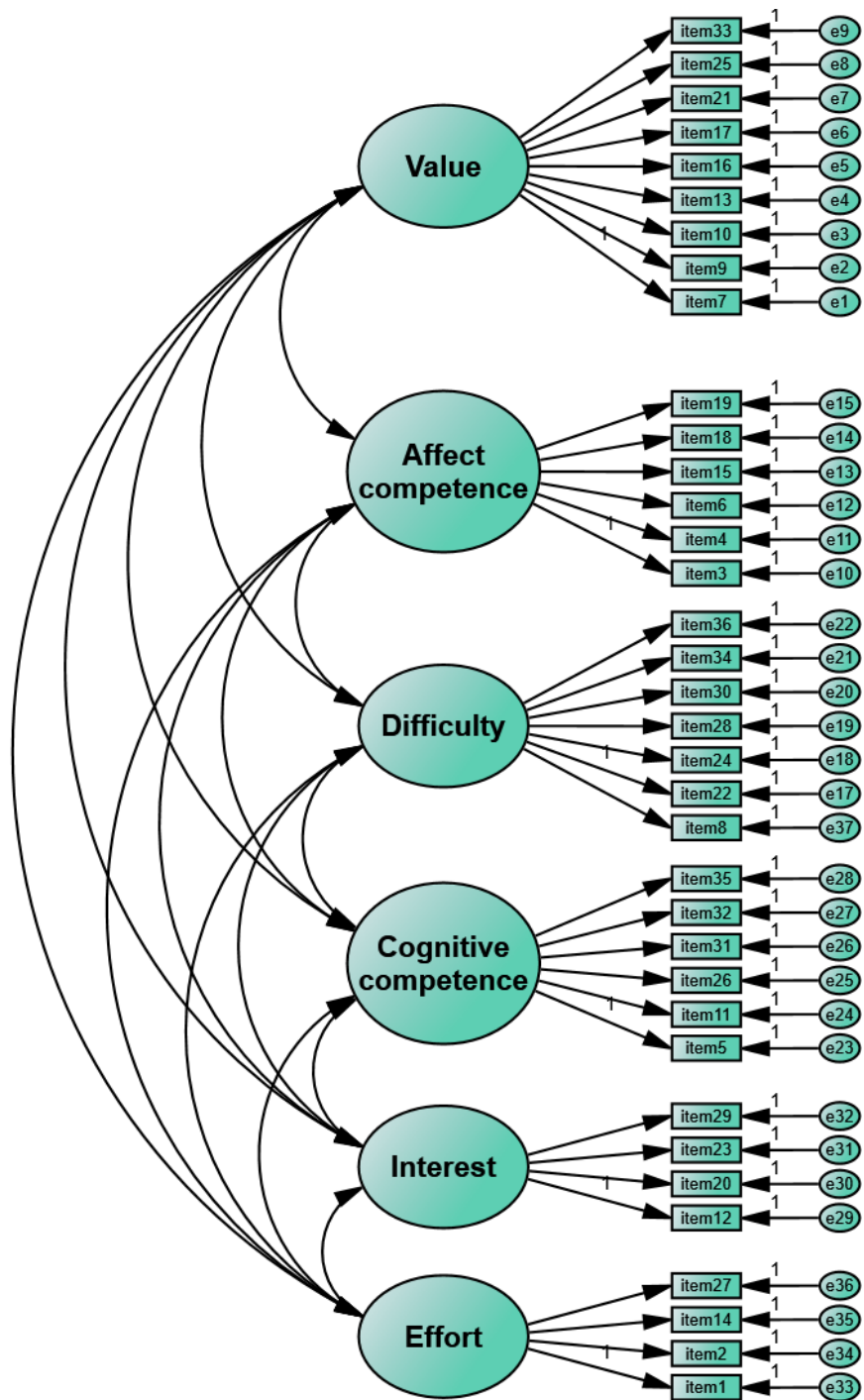
Figure 1. Conceptual model (Developed by Researcher).

(0.003), as [31] points out that (RMSEA) tends to simply ignore the most simple models, and has conducted well in the most complex larger items, dropped on the first model, which includes six extra items negative, and the (0.012 and 0.002) respectively, the value of the index of the root of the squares of residuals indicated an acceptable fit, and this result was similar to the findings of [32] and the study of [29], which revealed a slight improvement in the values of consis-

tency indicators when insufficient effort responses were eliminated. As shown in the results presented in Table 5, the third model, which had two negative addition items, had the best modifications to the data appropriateness ratios for the three models after the elimination of responses.

11. Conclusion

The study was concluded as the use of extra negative items in the detection of insufficient effort responses based on the results revealed, which proved their indicators to be better after the elimination of insufficient effort responses, and it also enters directly into the layout of the scale, which facilitates the process of their implementation and extraction of their results. In addition, the approach of extra negative items is studied in the detection of insufficient effort responses using a variable number of extra negative items. Furthermore, the approach of extra negative items is studied in the detection of insufficient effort responses based on different levels of the Likert scale, as the current study does.

Recommendations

There is a need for further research into the method of extra negative items in the detection of insufficient effort responses by using different levels of the Likert scale, the knowledge of much more than an exceed for each level of the different levels of Likert, and the development of several criteria; to ensure that the data is completely accurate, which ensures its quality and allows us to generalize the findings.

Researchers also encourage continual study to compare the various approaches to detecting insufficient effort responses that can be utilized by researchers who are not experts in measurement, assessment, and statistics.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

[1] Kountur, R. (2016) Detecting Careless Responses to Self-Reported Questionnaires. *Eurasian Journal of Educational Research*, **64**, 307-318. https://doi.org/10.14689/ejer.2016.64.17

[2] Meijer, R.R. and Nering, M.L. (1997) Trait Level Estimation for Nonfitting Response Vectors. *Applied Psychological Measurement*, **21**, 321-336. https://doi.org/10.1177/01466216970214003

[3] Huang, J.L., Curran, P.G., Keeney, J., Poposki, E.M. and DeShon, R.P. (2012) Detecting and Deterring Insufficient Effort Responding to Surveys. *Journal of Business and Psychology*, **27**, 99-114. https://doi.org/10.1007/s10869-011-9231-8

[4] Meade, A.W. and Craig, S.B. (2012) Identifying Careless Responses in Survey Data. *Psychological Methods*, **17**, 437-455. https://doi.org/10.1037/a0028085

[5] McKay, A.S. (2014) Improving Data Quality with Four Short Sentences: How an

Honor Code Can Make the Difference during Data Collection. M.A. Thesis, California State University, San Bernardino.

[6] Thompson, A.H. (1975) Random Responding and the Questionnaire Measurement of Psychoticism. *Social Behavior and Personality: An International Journal*, **3**, 111-115. https://doi.org/10.2224/sbp.1975.3.2.111

[7] Zijlstra, W.P., van der Ark, L.A. and Sijtsma, K. (2011) Outliers in Questionnaire Data: Can They Be Detected and Should They Be Removed? *Journal of Educational and Behavioral Statistics*, **36**, 186-212. https://doi.org/10.3102/1076998610366263

[8] Rios, J.A., Guo, H., Mao, L. and Liu, O.L. (2017) Evaluating the Impact of Careless Responding on Aggregated-Scores: To Filter Unmotivated Examinees or Not? *International Journal of Testing*, **17**, 74-104. https://doi.org/10.1080/15305058.2016.1231193

[9] Eysenck, H.J. and Eysenck, S.G.B. (1965) The Eysenck Personality Inventory. *British Journal of Educational Studies*, **14**, Article No. 140.

[10] Tendeiro, J.N., Meijer, R.R., Niessen, A.S.M., *et al.* (2016) PerFit: An R Package for Person-Fit Analysis in IRT. *Journal of Statistical Software*, **74**, 1-27. https://doi.org/10.18637/jss.v074.i05

[11] Snijders, T.A.B. (2001) Asymptotic Null Distribution of Person Fit Statistics with Estimated Person Parameter. *Psychometrika*, **66**, 331-342. https://doi.org/10.1007/BF02294437

[12] Levine, M.V. and Rubin, D.B. (1979) Measuring the Appropriateness of Multiple-Choice Test Scores. *Journal of Educational and Behavioral Statistics*, **4**, 269-290. https://doi.org/10.3102/10769986004004269

[13] Conijn, J.M., Emons, W.H.M. and Sijtsma, K. (2014) Statistic l_z-Based Person-Fit Methods for Noncognitive Multiscale Measures. *Applied Psychological Measurement*, **38**, 122-136. https://doi.org/10.1177/0146621613497568

[14] Swain, S.D., Weathers, D. and Niedrich, R.W. (2008) Assessing Three Sources of Misresponse to Reversed Likert Items. *Journal of Marketing Research*, **45**, 116-131. https://doi.org/10.1509/jmkr.45.1.116

[15] Schriesheim, C.A. and Hill, K.D. (1981) Controlling Acquiescence Response Bias by Item Reversals: The Effect on Questionnaire Validity. *Educational and Psychological Measurement*, **41**, 1101-1114. https://doi.org/10.1177/001316448104100420

[16] Bradley, K.D., Royal, K.D. and Bradley, J.W. (2008) An Investigation of 'Honesty Check' Items in Higher Education Course Evaluations. *Journal of College Teaching & Learning*, **5**, 39-48. https://doi.org/10.19030/tlc.v5i8.1240

[17] Hinz, A., Michalski, D., Schwarz, R. and Herzberg, P.Y. (2007) The Acquiescence Effect in Responding to a Questionnaire. *Psycho-Social-Medicine*, **4**, 1-9.

[18] Van Sonderen, E., Sanderman, R. and Coyne, J.C. (2013) Ineffectiveness of Reverse Wording of Questionnaire Items: Let's Learn from Cows in the Rain. *PLOS ONE*, **8**, e68967. https://doi.org/10.1371/journal.pone.0068967

[19] Józsa, K. and Morgan, G.A. (2017) Reversed Items in Likert Scales: Filtering Out Invalid Responders. *Journal of Psychological and Educational Research*, **25**, 7-25.

[20] Schwarz, R. and Hinz, A. (2001) Reference Data for the Quality of Life Questionnaire EORTC QLQC30 in the General German Population. *European Journal of Cancer*, **37**, 1345-1351. https://doi.org/10.1016/S0959-8049(00)00447-0

[21] Baer, R.A., Ballenger, J., Berry, D.T.R. and Wetter, M.W. (1997) Detection of Random Responding on the MMPI-A. *Journal of Personality Assessment*, **68**, 139-151. https://doi.org/10.1207/s15327752jpa6801_11

[22] Schau, C. (2003) Students' Attitudes: The "Other" Important Outcome in Statistics Education. 2003 *Joint Statistical Meetings-Section on Statistical Education*, San Francisco, 3673-3681

[23] Saraierh, R. (2013) Construct Validity of the Arabic Version of Arabic version of the Measure in Attitudes toward Statistics Scale (SATS-36). *Faculty of Education-Ain Shams University*, **3**, 651-672.

[24] Al-Sharim, A. (2015) Invariance of Factor Structure to Survey Attitudes toward Statistics (SATS-36) by Administration Time of Scale. *The International Journal of Interdisciplinary Educational Studies*, **4**, 14-31.

[25] DeSimone, J.A., Harms, P.D. and DeSimone, A.J. (2015) Best Practice Recommendations for Data Screening. *Journal of Organizational Behavior*, **36**, 171-181. https://doi.org/10.1002/job.1962

[26] Hendrawan, I., Glas, C.A.W. and Meijer, R.R. (2005) The Effect of Person Misfit on Classification Decisions. *Applied Psychological Measurement*, **29**, 26-44. https://doi.org/10.1177/0146621604270902

[27] Hult, G.T.M., Hair Jr, J.F., Proksch, D., Sarstedt, M., Pinkwart, A. and Ringle, C.M. (2018) Addressing Endogeneity in International Marketing Applications of Partial Least Squares Structural Equation Modeling. *Journal of International Marketing*, **26**, 1-21. https://doi.org/10.1509/jim.17.0151

[28] Steedle, J.T., Hong, M. and Cheng, Y. (2019) The Effects of Inattentive Responding on Construct Validity Evidence When Measuring Social-Emotional Learning Competencies. *Educational Measurement Issues and Practice*, **38**, 101-111. https://doi.org/10.1111/emip.12256

[29] Al Quraan, M. (2019) The Effect of Insufficient Effort Responding on the Validity of Student Evaluation of Teaching. *Journal of Applied Research in Higher Education*, **11**, 604-615. https://doi.org/10.1108/JARHE-03-2018-0034

[30] Cheung, G.W. and Rensvold, R.B. (2009) Structural Equation Modeling: A Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, **9**, 233-255. https://doi.org/10.1207/S15328007SEM0902_5

[31] Breivik, E. and Olsson, U.H. (2001) Adding Variables to Improve Fit: The Effect of Model Size on Fit Assessment in LISREL. In: Cudeck, R., Du Toit, S., Sorbom, D., Eds., *Structural Equation Modeling: Present and Future*, Scientific Software International, Lincolnwood, IL, 169-194.

[32] Steedle, J. (2018) Detecting Inattentive Responding on a Psychosocial Measure of College Readiness. Research Report 2018-5, ACT, Inc., Iowa City.