The Internal Structure of New Attitudes towards Statistics Inventory among Postgraduate Students in Medical and Health Sciences

Hazwan Mat Din [1,2], Wan Nor Arifin[1], Sarimah Abdullah[1], Muhammad Saiful Bahri Yusoff[3], Aniza Abd Aziz[4]

[1]Unit of Biostatistics and Research Methodology, School of Medical Sciences, Universiti Sains Malaysia, Kelantan, MALAYSIA

[2]Malaysian Research Institute on Ageing, Universiti Putra Malaysia, Selangor, MALAYSIA

[3]Medical Education Department, School of Medical Sciences, Universiti Sains Malaysia, Kelantan, MALAYSIA

[4]Unit of Community Medicine, Faculty of Medicine, Universiti Sultan Zainal Abidin, Terengganu, MALAYSIA

ABSTRACT

Universiti Sains Malaysia Attitudes toward Statistics (USM-AS) is a newly developed, self-administered inventory for measuring attitudes toward statistics among postgraduate students. The USM-AS consists of three factors (affect, cognition, and behaviour) and 58 items. This study was aimed to validate the USM-AS and provide evidence of its construct validity by investigating its internal structure. This study consists of four validation studies, which were conducted consecutively among postgraduate students (Master's degree and PhD students) in the medical and health sciences. The revised USM-AS, consisting of three factors and 12 items, was found to have good construct validity, which was demonstrated by good model fit, high factor loadings (0.62 to 0.92), high construct reliability (0.84 to 0.89), good discrimination between factors and good test-retest reliability (intraclass correlation coefficients = 0.63 to 0.82). However, validity was limited to postgraduate students in the medical and health sciences in Malaysia. Further validation studies among postgraduate students in other scientific fields are recommended to provide additional construct validity evidence of the USM-AS.

Keywords: *Attitudes toward statistics, Construct validity, Postgraduate students, Self-administered inventory, USM-AS*

CORRESPONDING AUTHOR Wan Nor Arifin, Unit of Biostatistics and Research Methodology, School of Medical Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia | E-mail: wnarifin@usm.my

INTRODUCTION

Background

Statistics is an integral component of Universiti Sains Malaysia Attitudes toward Statistics (USM-AS) postgraduate studies because of its importance in the analysis and interpretation of data (1, 2). Despite its importance in research, previous studies have shown that statistics is often viewed negatively (3–5). Negative attitudes toward statistics hinder the effective learning of this subject (3, 6) although it is deemed an important part of postgraduate courses. A positive attitude toward statistics is important in the process of learning and understanding statistical methods (3, 5–7).

Many inventories have been developed to measure attitudes toward statistics. These inventories vary greatly in term of the number of factors, theoretical background, and quality of validation evidence (1, 8). Nolan et al. (8) found that only four inventories had substantial validity evidence: Statistics Attitude Survey (SAS) (9), Attitudes toward Statistics Scale (ATS) (10), the Survey of Attitudes Toward Statistics–28 (SATS-28) (11) and the Survey of Attitudes Toward Statistics–36 (SATS-36) (12). Of these four, the SATS-28 and SATS-36 are the most notable for their accuracy in determining the validity evidence in confirmatory factor analyses (CFA).

SATS-28 consists of four factors (affect, cognitive competence, value, and difficulty), which are measured by 28 items (11). The items are rated on a seven-point Likert-type scale (from 1 = strongly disagree to 7 = strongly agree). Two additional factors, interest and effort, and eight items, were added to the SATS-28, which created the SATS-36 (12). Previous studies showed that both the SATS-28 and SATS-36 demonstrated good model fit and internal consistency reliability (8). However, in previous validation studies, several issues were noted in the use of the SATS-28 and SATS-36 (11, 13–16): (i) item parcelling

was used in most of the analyses; and (ii) multicollinearity was notable between several factors (affect, cognitive competence, and difficulty, with or without parcelling). The use of item parcelling is controversial (16) because it may obscure the latent structure of the data (17). Multicollinearity between factors also indicates the need to combine factors (17). Additionally, the SATS-28 and SATS-36 were developed and validated for use among undergraduate students although some authors have utilised the inventories among postgraduate students (5, 18).

USM-AS Inventory

In view of the limitations of the available inventories, a self-administered inventory named USM-AS was developed to measure attitudes toward statistics among postgraduate students (1). It was specifically developed for Master's degree and PhD students who had previous exposure to statistics courses (1). The USM-AS was developed using the Delphi method based on the tripartite theory of attitudes (19).

The Delphi method is a structured group communication process, involving four phases, which are: (i) exploration of a subject by a group; (ii) reaching an understanding of how the group views the subject; (iii) resolving disagreement; and (iv) final evaluation (20). Based on these phases, Arifin et al. (1) conducted the group communication in five phases as follows:

1. Identification of factors of attitude toward statistics: The early four phases involved three experts, who were applied statisticians. One of the experts coordinated the communication process by email. The coordinator emailed the experts, requesting several possible factors of attitude toward statistics. Every expert provided possible factors based on their experience, previous research, and literature reviews. The factors provided were continuously discussed until all experts agreed on a number of candidate factors.

2. Verification of the factors: The experts were then asked to further verify the appropriateness and applicability of each factor to postgraduate students. Only factors that were considered appropriate and applicable to the students were considered in the next phase.

3. Defining the factors: The experts were asked to provide provisional definitions for each of the factors, keeping the context in postgraduate settings. These provisional definitions were discussed until they agreed on a suitable definition for each factor.

4. Identification of relevant and representative items: In this phase, the experts were asked to construct as many items as possible that were relevant and representative of the factors. The item construction was guided by the definitions provided in the preceding phase. The constructed items included negative and positively worded items.

5. Final verification of the factors and items: A face-to-face meeting was organised, involving all the three experts and an invited expert (also an applied statistician) who was not involved in the preceding four phases. The factors, definitions and items were presented and evaluated in the presence of the new expert. Suitable response options to the items were discussed in view of the advantages and disadvantages of several different response options. The final draft of USM-AS was produced from this meeting.

The newly developed USM-AS consists of three factors and 58 items: Affect (22 items) measures "emotional reaction and feeling toward statistics"; Cognition (19 items) measures "thought and belief about statistics"; and Behaviour (17 items) measures "actions resulting from affect and cognition toward statistics" (1). Each item is rated on a Likert-type scale of 1 to 5 (1 = strongly disagree; 2 = disagree; 3 = neutral; 4 = agree; 5 = strongly agree).

The evidence of the construct validity of the USM-AS as measured by content was previously provided during its development as described and reported by Arifin et al. (1). The present study aims to continue the validation of the USM-AS and provide evidence for its construct validity by investigating its internal structure.

METHODS

General Methods for the Validation Studies

This research consisted of four validation studies, which were conducted from September 2013 to December 2014 among postgraduate students on the Health Campus of Universiti Sains Malaysia (USM). Postgraduate students comprised of PhD, Master of Science (MSc) and Master of Medicine (MMed) students, mainly from the School of Medical Sciences, the School of Health Sciences and the School of Dental Sciences. Only registered postgraduate students who attended intensive statistics courses were eligible to participate in this study. This requirement fit the target population of the USM-AS (1), thus ensuring that an accurate assessment of the attitudes toward statistics could be conducted.

The intensive statistics courses were a basic statistics course and an intermediate statistics course. Both were five days in duration, and they covered covering basic statistical analyses and multivariable statistical analyses, respectively. The data collection was conducted on the third or fourth day of the basic statistics course, assuming that it was the first statistics course the students had attended. In contrast, data collection took place on any of the days during the intermediate statistics course because the students should have previously attended the basic statistics course. This approach to data collection was applied to ensure that the students had been exposed to statistics before they responded to the

USM-AS. In addition, different cohorts of students were invited to participate in Validation Study 1 (basic statistics course in 2013), Validation Study 2 (intermediate statistics course in 2013), and Validation Study 3 (basic statistics course in 2014). The involvement of students from different cohorts were ensured because the students were required to attend only one statistics course per year. For Validation Study 4 (intermediate statistics course in 2014), it might involve the same cohort of students from the basic statistics course in 2013.

Ethical approval to conduct the study was obtained from the USM Research Ethics Committee (Human). Questionnaire forms consisting of anonymous demographic information and the USM-AS inventory were distributed to all students who were attending the courses. The purpose and expected outcome of the study were explained to the students. All students were assured that it was not compulsory to participate in the study. Their implied consent to participate in the study was obtained when the students returned the forms to the researchers.

The data management and analyses were performed using the software IBM SPSS Statistics Version 20 and SPSS Amos Version 19.

Validation Study 1: Exploratory

Validation Study 1 was conducted to find evidence of the construct validity by internal structure of the USM-AS. The measurement instruments were exploratory factor analysis (EFA) and internal consistency reliability (Cronbach's alpha).

The questionnaire forms containing the 58-item USM-AS inventory were pretested among 10 postgraduate students, all of whom were in the Master of Science degree programme. They found the items easy to understand; the response options were self-explanatory. Following the pre-test, the questionnaire forms were distributed on the third day of the basic statistics course.

The EFA was performed on data using the principal axis factoring extraction method and the promax rotation method. Principal axis factoring was chosen because it is less prone to improper solutions and free of distributional assumptions (21). Oblique rotation methods, one of which is promax, are preferred because oblique rotation allows correlations between factors, thus provides more realistic representation of how factors are correlated (17). The Kaiser Mayer Olkin (KMO) ≥ 0.7 and Bartlett's test of sphericity ($p < 0.05$) were used to indicate the suitability of the data for the EFA. The number of extracted factors was determined by eigenvalues > 1.0 and scree plot inspection (17). A factor loading (FL) ≥ 0.5 was considered acceptable (22). For internal consistency reliability, a Cronbach's alpha coefficient ≥ 0.7 was satisfactory (23).

For the EFA, a minimum sample size of 150 was required whenever 10 or more items were expected to have factor loadings of 0.4 (24).

Validation Study 2: Confirmatory

Validation Study 2 was conducted to confirm the evidence of construct validity by internal structure based on the findings in Validation Study 1. Validation Study 2 was performed using a new sample of CFA and construct reliability. The questionnaire forms containing the 27-item USM-AS inventory were distributed on the first day of the intermediate statistics course.

Prior to performing the CFA by maximum likelihood (ML) estimation method, the multivariate normality of the data was assessed. The multivariate normality assessment was done by Mardia's normalised estimate of multivariate kurtosis (25) and the chi-square versus the Mahalanobis distance plot (26). A critical ratio of kurtosis < 5.0 (27) and a fairly straight line on the chi-square versus the Mahalanobis distance plot (26, 28) indicated multivariate normality. Multivariate outliers were identified by the

Mahalanobis distance values and the visual inspection of the chi-square versus the Mahalanobis distance plot.

CFA was performed using Amos, and the model fit assessment was done by five fit indices recommended by Brown (17): chi-square goodness-of-fit (χ^2), standardised root mean square residual (SRMR), root mean square error of approximation (RMSEA), comparative fit index (CFI), and Tucker-Lewis fit index (TLI). The following cut-off values were applied to indicate model fit: χ^2 p > 0.05, CFI and TLI ≥ 0.95, RMSEA and its upper 90% confidence limit ≤ 0.08, RMSEA's close fit (CFit) p > 0.05 and SRMR ≤ 0.08 (17, 29). FLs ≥ 0.5 were considered acceptable (22). Factor correlations < 0.85 indicated good discrimination between the factors (17).

When a model fits poorly, it can be improved by removing problematic items (17). In this study, the model improvement was based on the assessment of standardised residuals (SR) and modification indices (MI). In order of magnitude, items with standardised residuals (SRs) > |2.58| followed by |1.96|, and those with relatively high MIs, were inspected as the possible sources of poor model fit (17). Construct reliability (CR) was calculated using the formula given in Fornell and Larker (30). A CR value ≥ 0.7 was considered acceptable (22).

For CFA, a minimum sample size of 150 was required for normally distributed data using ML estimation method (31).

Validation Study 3: Cross-validation

Validation Study 3 was conducted to cross-validate USM-AS in a new sample and to provide evidence of its construct validity according to internal structure. The evidence was provided by CFA and CR following the same methods used in Validation Study 2. For CFA, a minimum sample size of 150 was recommended (31).

Validation Study 4: Test-retest reliability

Validation Study 4 provided additional validity evidence of the internal structure in the form of test-retest reliability by intraclass correlation. Based on the cross-validation CFA findings in Validation Study 3, 12 items remained in the USM-AS inventory. The questionnaire forms containing the 12-item USM-AS inventory were distributed on the second day (Time 1) and the fifth day (Time 2) of the intermediate statistics course, which was a gap of three days between the test and retest days.

A two-way mixed intraclass correlation (ICC) model of consistency type was used as suggested by Weir (32) in a test-retest situation. The ICC was analysed for the total scores by factor at Time 1 and Time 2. The ICC results were interpreted according to Cicchetti (33), in which ICC values of less than 0.40 are poor, between 0.40 and 0.59 are fair, between 0.60 and 0.74 are good, and equal to or more than 0.75 are excellent. A sample size of 26 was required for the test-retest (two repetitions) using ICC whenever the minimum acceptable ICC = 0.6 and the expected ICC = 0.85, at a two-tailed significance level = 0.05 and power = 0.8 (34).

RESULTS

Validation Study 1: Exploratory

A total of 156 postgraduate students in the campus were involved. The sample consisted of 49 (31.4%) male and 107 (68.6%) female students, with a mean age of 29.0 years (SD = 3.16). There were 93 (59.6%) MMed, 49 (31.4%) MSc, and 14 (9.0%) PhD students. Of these, 136 (87.2%) were from the School of Medical Sciences, and 20 (12.8%) were from the School of Health Sciences, the School of Dental Sciences and other schools and institutes in USM. Most of them were Year Two students (n = 118, 75.6%), followed by Year One students (n = 31, 19.9%), and Year Three students (n = 7, 4.5%).

The data included three cases in which one response was missing for each case (0.03% of the overall data points). The missing values were imputed with the means of the respective items (i.e., the single imputation method). Given the negligible percentage of missing values, this method of imputation was chosen because of its simplicity. At this percentage of missingness, most imputation methods will yield similar results with minimum bias in estimates, regardless of the mechanisms of missingness (35, 36).

KMO was 0.84 and Bartlett's test of sphericity was significant ($p < 0.001$), which confirmed that the data were suitable for EFA. Twelve eigenvalues were above 1.0, which suggested 12 factors, whereas the scree plot inspection suggested between three and four factors. The EFA was continued by fixing the number of factors to three, which corresponded to the three attitudinal factors in the USM-AS. Items that had FLs < 0.5 (19 items) and relatively low communalities (2 items) were deleted. Items that were extracted into different factors (6 items) and items that cross-loaded into multiple factors (3 items) were also deleted. In addition, an item was deleted because the item statement was considered too general according to expert opinion. In total 31 items were deleted following the EFA.

The EFA resulted in a three-factor solution and 27 items (Affect: 9 items, Cognition: 12 items, Behaviour: 6 items), explaining 55.0% of the variance in the items. The FLs, communalities, and Cronbach's alpha values are presented in Table 1.

Table 1: Factor loadings and Cronbach's alpha in the EFA of USM-AS

Factor	Item statement	Factor loadings	Communalities	Cronbach's alpha
Affect	**Statistical analysis makes me feel bored (R)**	0.63	0.35	0.87
	I feel hopeless when doing statistical analysis (R) (SR, 5)	0.64	0.39	
	I feel sleepy while reading statistical materials (R)	0.53	0.46	
	I think statistics is easy to learn (SR, 3)	0.65	0.55	
	Interpreting statistical result is a burden for me (R) (MI)	0.69	0.46	
	Statistical interpretation makes me feel bored [Interpreting statistics makes me feel bored] (R)	0.58	0.44	
	I feel tired to learn statistics [I feel tired when I learn statistics] (R)	0.64	0.50	
	I am not confident to interpret statistical test [I am not confident in interpreting statistical tests] (R) (SR, 2)	0.72	0.43	
	I hate calculation required in statistics [I hate doing the calculations required in statistics] (R)	0.64	0.45	

(Continued on next page)

Table 1: (*Continued*)

Factor	Item statement	Factor loadings	Communalities	Cronbach's alpha
Cognition	Statistics is helpful to generate new research questions [Statistics is helpful in generating new research questions] (FL)	0.54	0.39	0.92
	Statistical method is important to solve scientific problems [The statistical method is important in solving scientific problems] (SR, 9)	0.68	0.51	
	Application of statistical knowledge is helpful to critically appraise a research article [The application of statistical knowledge is helpful in critically appraising a research article]	0.64	0.42	
	Statistics is helpful for evidence-based practice (SR, 1)	0.64	0.40	
	Understanding statistics is helpful in research decision making [Understanding statistics is helpful in making research decisions]	0.72	0.50	
	Statistics is important in research (SR, 10)	0.77	0.57	
	Statistics is helpful to understand research findings [Statistics is helpful in understanding research findings] (FL)	0.81	0.67	
	Statistics lecture is important [Statistics lectures are important] (FL)	0.79	0.67	
	Sample size calculation is important (SR, 11)	0.75	0.57	
	Statistics is helpful to discuss research findings [Statistics is helpful in discussing research results] (SR, 12)	0.75	0.59	
	Statistics is helpful to write research results [Statistics is helpful in writing research results]	0.76	0.61	
	Learning statistics for research project[s] is vital (SR, 8)	0.61	0.39	
Behaviour	**I read statistics before the lecture**	0.76	0.49	0.84
	I do additional reading on statistics	0.80	0.61	
	I buy statistics books [I buy books on statistics]	0.70	0.49	
	I discuss about statistics with friends [I discuss statistics with friends] (SR, 6)	0.50	0.37	

(*Continued on next page*)

Table 1: (*Continued*)

Factor	Item statement	Factor loadings	Communalities	Cronbach's alpha
	I guide other students on data analysis [I guide other students in doing data analyses] (SR, 4)	0.57	0.47	
	I attend extra statistics class[es] (SR, 7)	0.70	0.61	

Notes: (FL) = deleted because factor loadings < 0.5; (MI) = deleted because of having relatively high MI; (R) = reverse-scored items; (SR, order of deletion) = deleted because standardised residuals > |1.96|. Items that remained in the CFA (Validation Study 2) are highlighted in bold. Item statements with corrections by a native speaker of English are provided within the statements or after the statements, surrounded by square brackets.

Table 2: Fit indices of USM-AS models

Model	χ^2 (df), *p*	SRMR	RMSEA (90%CI), CFit *p*	CFI	TLI
USM-AS EFA	761.18 (321), < 0.001	0.09	0.09 (0.08, 0.10), < 0.001	0.80	0.78
USM-AS CFA	67.51 (41), < 0.001	0.06	0.06 (0.03, 0.09), 0.228	0.97	0.95
USM-AS CV	86.27 (51), < 0.001	0.06	0.07 (0.04, 0.10), 0.089	0.96	0.95

Notes: USM-AS EFA = three-factor model of USM-AS based on EFA; USM-AS CFA = three-factor model of USM-AS based on CFA; USM-AS CV = three-factor model of USM-AS based on cross-validation study; SRMR = standardised root mean square residual; RMSEA = root mean square error of approximation; CFit = close fit; CFI = comparative fit index; TLI = Tucker-Lewis fit index.

Validation Study 2: Confirmatory

The study sample comprised 174 postgraduate students. The sample consisted of 64 (36.8%) male and 110 (63.2%) female students, with a mean age of 31.7 years (*SD* = 3.89). The majority of them came from the School of Medical Sciences (n = 165; 94.8%), while the remainder came from the School of Health Sciences and the School of Dental Sciences (n = 9; 5.2%). The sample comprised of 118 (67.8%) MMed, 41 (23.6%) MSc, and 15 (8.6%) PhD students. There were 17 (9.8%) Year One students, 26 (14.9%) Year Two students, 122 (70.7%) Year Three students, and 8 (4.6%) Year Four students.

The data included two cases with one missing response each (0.04% of the overall data points). Similar to Validation Study 1, the missing values were imputed using the means of the respective items (i.e., single imputation method).

The multivariate normality assessment resulted in a critical ratio of kurtosis (10.80)

and a fairly straight line on the chi-square versus the Mahalanobis distance plot. The data were considered multivariate normal because based on the critical ratio of kurtosis, the violation of multivariate normality was small, and the data appeared normally distributed on the chi-square versus the Mahalanobis distance plot. Thus, the CFA was performed using ML estimation method in Amos.

The CFA of the EFA-derived three-factor model with 27 items (USM-AS EFA) showed that the model did not fit the data (Table 2). To improve the model fit, the items were deleted iteratively. Three items of the cognition factor with FLs of less than 0.5 were deleted ("Statistics is helpful in generating new research questions" [FL = 0.32], "Statistics is helpful in understanding research findings" [FL = 0.45], and "Statistics lectures are important" [FL = 0.47]).

Six items with SRs of more than |2.58| were deleted sequentially in the following order: "Statistics is helpful for evidence-

based practice" (SRs = −3.19, −2.60; FL = 0.61), "I am not confident in interpreting statistical tests" (SRs = −2.89, −3.00; FL = 0.66), "I think statistics is easy to learn" (SR = 2.76, 2.68; FL = 0.56), "I guide other students in doing data analyses" (SR = 2.72, 3.10; FL = 0.59), "I feel hopeless when doing statistical analysis" (SR = −2.84; FL = 0.69), "I discuss statistics with friends" (SR = 2.72; FL = 0.64). At this point, the model did not fulfil any of the fit indices, except for SRMR = 0.08. Six items with SRs of more than |1.96| iteratively were deleted in the following order: "I attend extra statistics classes" (SR = 2.15, 2.22; FL = 0.51), "Learning statistics for research projects is vital" (SR = −2.32; FL = 0.54), "The statistical method is important in solving scientific problems" (SR = −2.35; FL = 0.60), "Statistics is important in research" (SR = 2.38; FL = 0.67), "Sample size calculation is important" (SR = 2.57; FL = 0.66), "Statistics is helpful in discussing research results" (SR = −2.06, −2.21, −2.36; FL = 0.72). Although there were no more items with SR > 1.96, the model still did not fit (based on CFI = 0.93, RMSEA = 0.08 [90% CI: 0.05, 0.10], CFit p = 0.03). Thus, an item in the affect factor ("Interpreting statistical result is a burden for me") was deleted because it showed a relatively high MI (highest MI = 11.62; FL = 0.72).

Finally, after deleting 16 problematic items, the CFA resulted in a three-factor model with 11 items (Affect: 5 items, Cognition: 3 items, Behaviour: 3 items), which showed good model fit (USM-AS CFA, Table 2), high FLs (0.58 to 0.80), high construct reliability, and good discrimination between the factors (r = −0.37 to 0.50) (Table 3), which indicated good validity according to internal structure.

Validation Study 3: Cross-validation

The study sample comprised 139 postgraduate students. The multivariate normality assessment indicated four outliers. Following the exclusion of these four outliers, the results showed acceptable

multivariate normal data, with a critical ratio of kurtosis = 10.46 (mild violation to normality) and a fairly straight line on the chi-square versus the Mahalanobis distance plot. Thus, assuming multivariate normality, the CFA was performed on the data of the remaining 135 students, which consisted of 40 (29.6%) male and 95 (70.4%) female students, with mean age of 31.3 years (SD = 3.12). There were 104 (77.0%) MMed, 17 (12.6%) MSc, and 8 (5.9%) PhD students, while the remainder (n = 6, 4.4%) were undisclosed. Of these, 109 (80.7%) were from the School of Medical Sciences and 20 (14.3%) were from the School of Health Sciences, the School of Dental Sciences and other schools and institutes in USM, while the rest (n = 6, 4.4%) were undisclosed. There were 11 (8.1%) Year One students, 111 (82.2%) Year Two students, 9 (6.7%) Year Three students, and 4 (3.0%) Year Four students.

Based on the CFA findings in Validation Study 2, only 11 items remained in the USM-AS inventory. The cognition and behaviour factors had only three items each. After comparing the items in the EFA (Validation Study 1) and CFA (Validation Study 2), some deleted items had good content, which were possibly not clearly worded. In addition, according to expert opinion, the item "I read the statistics before the lecture" in the behaviour factor was not clear despite its high FL of 0.79. Hence, it was recommended to replace the item to improve its meaning. Following these recommendations, five items were added to the USM-AS after the revision. Three items were added to the cognition factor: "Statistics is important in drawing conclusions based on research results" (rephrased from "Statistics is helpful in discussing research findings"); "Statistics is important in understanding research results" (rephrased from "Statistics is helpful in understanding research findings"); "Statistics is important in planning research" (rephrased from "Statistics is important in research"). Two items were added to the behaviour factor; "I discuss statistics with other students" (rephrased

Table 3: Factor loadings and construct reliabilities of USM-AS CFA and USM-AS CV

Factor	Item statement	USM-AS CFA[a]		USM-AS CV[b]	
		Factor loadings	Construct reliabilities	Factor loadings	Construct reliabilities
Affect	I feel sleepy while reading statistical materials (R)	0.73	0.88	0.67	0.89
	I feel tired to learn statistics [I feel tired when I learn statistics] (R)	0.78		0.81	
	Statistical analysis makes me feel bored (R)	0.78		0.84	
	Statistical interpretation makes me feel bored [Interpreting statistics makes me feel bored] (R)	0.79		0.88	
	I hate calculation required in statistics [I hate doing the calculations required in statistics] (R)	0.80		0.71	
Cognition	Statistics is helpful to write research results [Statistics is helpful in writing research results]	0.62	0.75	0.86	0.87
	Statistics is important to understand research results [Statistics is important in understanding research results]	–		0.62	
	Understanding statistics is helpful in research decision making [Understanding statistics is helpful in making research decisions]	0.74		0.90	
	Application of statistical knowledge is helpful to critically appraise a research article [The application of statistical knowledge is helpful in critically appraising a research article]	0.75		0.77	
Behaviour	I buy statistics books	0.58	0.73	0.63	0.84
	I do additional reading on statistics	0.70		0.92	
	I read statistics before the lecture	0.79		–	
	I do preliminary reading before statistics lecture[s]	–		0.82	

Notes: (R) = reverse-scored items. Newly added items in USM-AS CV are highlighted in bold.
[a]Factor correlations: Affect ↔ Cognition $r = 0.48$, Affect ↔ Behaviour $r = 0.50$, Cognition ↔ Behaviour $r = -0.37$;
[b]Factor correlations: Affect ↔ Cognition $r = -0.04$, Affect ↔ Behaviour $r = 0.54$, Cognition ↔ Behaviour $r = 0.08$.
Item statements with corrections by a native speaker of English are provided within the statements or after the statements, surrounded by square brackets.

from "I discuss statistics with friends") and "I do preliminary reading before statistics lectures" (rephrased from "I read statistics before the lecture"). In total, 15 items on the USM-AS were tested in Validation Study 3 after "I read statistics before the lecture" was deleted from the behaviour factor. The questionnaire forms containing the 15-item USM-AS inventory were distributed on the fourth and fifth days of the basic statistics course.

In the CFA, the addition of the five new items resulted in a poor model fit. Three problematic items were identified: "I discuss statistics with other students" (low FL = 0.47); "Statistics is important in drawing conclusions based on research results" (SR = 2.22 with "Statistics is important in understanding research results" and a relatively low FL = 0.63); "Statistics is important in planning research" (SR = −2.16 with "Statistical analysis makes me feel bored" and a relatively low FL = 0.66). After deleting these three items, the CFA resulted in a three-factor model with 12 items (Affect: 5 items, Cognition: 4 items, Behaviour: 3 items), which showed good model fit (USM-AS CV, Table 2), high FLs (0.62 to 0.92), high construct reliability, and good discrimination between the factors (r = −0.04 to .54) (Table 3), which indicated good validity according to internal structure. A repeat CFA on the original sample (n = 139) before the removal of outlying observations did not show any remarkable differences in the factor loadings and factor correlations that would change any conclusions made on the CFA on the data of 135 students. The descriptive statistics of the total scores by factor based on USM-AS CV are presented in Table 4.

Validation Study 4: Test-retest reliability

In Validation Study 4, the sample comprised 27 postgraduate students, consisting of 5 (18.5%) male and 22 (81.5%) female students, with a mean age of 30.7 years (SD = 4.87). The ICC values (single measures) of the affect, cognition, and behaviour factors were 0.82, 0.63, and 0.79 respectively. These results indicated that the test-retest reliability was good in the cognition factor and excellent in the affect and behaviour factors.

DISCUSSION

The present study was conducted to find evidence to supports the validity of the USM-AS following the development of USM-AS (1). It describes construct validity in the form of evidence that supports validity instead of types of validity, which is in line with the unitary concept of validity proposed by Messick (37). This concept of validity was further refined and simplified by the American Educational Research Association (AERA), American Psychological Association (APA), and National Council on Measurement in Education (NCME) (37). The five sources of evidence are content, response process, internal structure, relations to other variables, and consequences (38, 39).

This research consisted of four validation studies, which were conducted consecutively to provide evidence of construct validity according to internal structure and relations with other variables. One EFA and two CFAs were performed to verify the internal structure of the USM-AS. It was verified

Table 4: Descriptive statistics of total scores by factor (N = 129)

Factor	Mean (SD)	Minimum	Maximum	Lowest possible	Highest possible
Affect	14.5 (3.87)	6	24	5	25
Cognition	17.1 (1.72)	12	20	4	20
Behaviour	7.6 (2.60)	3	15	3	15

Note: SD = standard deviation.

that USM-AS consisted of three factors and 12 items; five items for Affect, four items for Cognition, and three items for Behaviour. These findings were demonstrated by the good fit of the USM-AS CV model. Additionally, the FLs ranged from 0.62–0.92, and the internal consistency reliability was high (CRs = 0.84–0.89). The test-retest reliability of all factors was also good (ICCs = 0.63–0.82). The evidence of the validity of the internal structure of the USM-AS found in this study was not compared with other existing inventories because of the differences in the theoretical aspect of the inventories and the target population that was assessed.

LIMITATIONS

In the present study, the respondents were postgraduate students in the medical and health sciences. Therefore, the validity of the USM-AS in measuring attitudes toward statistics is limited to only postgraduate students in these fields. However, the general applicability of the USM-AS to all postgraduate students is possible. Further cross-validation studies of postgraduate students in other scientific fields are needed to establish the validity of the inventory in other populations.

At present, there are no comparable inventories that could provide additional evidence of the construct validity of USM-AS. The evidence of validity according to internal structure and the relations with other variables (or factor scores in other inventories) of USM-AS could not be compared to other existing inventories because of differences in the theoretical background and the statistical approach used in the development of the inventories. These differences were clearly described in a systematic review by Nolan et al. (8), in which the theoretical number of factors, the factor definitions, and the statistical approach used in the evaluation of internal structure differed significantly. Hence, the fit of a three-factor model of USM-AS is not

comparable to, for example, a four-factor model of SATS-28 with or without item parcelling. The differences in the numbers and definition of factors do not allow an assessment of the correlations between the factor scores of USM-AS and the factor scores of other inventories because the factors measure entirely different concepts. Furthermore, in this study, the focus of USM-AS on postgraduate students required a different set of items that were relevant to postgraduate students who are familiar with research and self-learning. In contrast, because the most commonly used SATS-28 and SATS-36 inventories were developed and validated for use among undergraduate students, the items are relevant to that population.

During the development of USM-AS (1), there was no native speaker of English among the experts involved in the research. This draft version of USM-AS was not submitted for editing to any native speaker of English during the development. There was also no native speaker of English among the students involved in the pre-test (Validation Study 1). This unfortunately resulted in item statements that were not grammatically correct (in Tables 1 and 3, grammatically sound item statements are provided within the square brackets). The item statements were only corrected by a professional editor, who was a native speaker of English at the end of the study. Therefore, it raised an issue with the validity of the content of the statements. The students involved in this study were mainly non-native speakers of English and they did not raise any issue with regards to the wording of the statements. It was assumed they were able to grasp the intended meaning in each statement despite the grammatical inaccuracy of the statements. However, this issue with the grammatical inaccuracy cannot be overlooked, which might limit the use of USM-AS only in Malaysia. Therefore, it is recommended that further validation of USM-AS should use the corrected statements as provided in Tables 1 and 3.

The sample size in Validation Study 3 was smaller (n = 135) than what was expected (n = 150). However, the average item communalities for each factor for the model with 12 items were 0.62, 0.63 and 0.64 for affect, cognition and behaviour, respectively. According to Hair et al. (22), whenever the model has less than six factors and item communality of more than 0.6, a sample size of at least 100 is adequate. It was thus justified to rely on the CFA results in Validation Study 3, because the sample size was between the range of 100 to 150.

CONCLUSION

The 12-item USM-AS was found to have good construct validity in measuring the affect, cognitive, and behaviour factors in attitudes toward statistics. The validity was demonstrated by a good model fit, high FLs, high construct reliability, good discrimination between factors and good test-retest reliability. However, this validity is limited to postgraduate students in the medical and health sciences in Malaysia. Thus, in future research, further validation studies should be conducted among postgraduate students in other scientific fields to establish the construct validity of USM-AS in other postgraduate populations.

ACKNOWLEDGEMENTS

REFERENCES

1. Arifin WN, Yusoff MSB, Aziz AA, Din HM, Abdullah S. Development of a new inventory of attitude towards statistics among postgraduate students. In: International Academic Forum. The Asian Conference on Education 2014. Aichi, Japan: The International Academic Forum (IAFOR); 2015 [cited 4 June 2017]; p. 577–84. Available from: http://papers.iafor.org/conference-proceedings/ACE/ACE2014_proceedings.pdf

2. Perepiczka M, Chandler N, Becerra M. Relationship between graduate students' statistics self-efficacy, statistics anxiety, attitude toward statistics, and social support. Professional Counselor. 2011;1(2):99–108. https://doi.org/10.15241/mpa.1.2.99

3. Ashaari NS, Judi HM, Mohamed H, Wook TMT. Student's attitude towards statistics course. Procedia Soc Behav Sci. 2011;18:287–94. https://doi.org/10.1016/j.sbspro.2011.05.041

4. Sami W. Biostatistics education for undergraduate medical students. Biomedica. 2010;26:80–4.

5. Zhang Y, Shang L, Wang R, Zhao Q, Li C, Xu Y, Su H. Attitudes toward statistics in medical postgraduates: measuring, evaluating and monitoring. BMC Med Educ. 2012;12(1):117. https://doi.org/10.1186/1472-6920-12-117

6. Daher AM, Amin F. Assessing the perceptions of a biostatistics and epidemiology module: views of year 2 medical students from a Malaysian university, a cross-sectional survey. BMC Med Educ. 2010;10(1):34. https://doi.org/10.1186/1472-6920-10-34

7. Judi HM, Ashaari NS, Mohamed H, Wook TMT. Student's profile based on attitude towards statistics. Procedia Soc Behav Sci. 2011;18:266–72. https://doi.org/10.1016/j.sbspro.2011.05.038

8. Nolan MM, Beran T, Hecker KG. Surveys assessing student' attitudes toward statistics: a systematic review of validity and reliability. Statistics Education Research Journal. 2012;11(2):103–23.

9. Roberts DM, Bilderback EW. Reliability and validity of a statistics attitude survey. Educ Psychol Meas. 1980;40(1):235–38. https://doi.org/10.1177/001316448004000138

10. Wise SL. The development and validation of a scale measuring attitudes toward statistics. Educ Psychol Meas. 1985;45(2):401–5. https://doi.org/10.1177/001316448504500226

11. Schau C, Stevens J, Dauphinee TL, Del Vecchio A. The development and validation of the survey of attitudes toward statistics. Educ Psychol Meas. 1995;55(5):868–75. https://doi.org/10.1177/0013164495055005022

12. Schau C. Students' attitudes: The "other" important outcome in statistics education. In: Proceedings from 2003 Joint Statistical Meetings. San Francisco, CA: American Statistical Association; 2003 [cited 4 June 2017]; p. 3673–83. Available from: http://statlit.org/pdf/2003SchauASA.pdf

13. Bechrakis T, Gialamas V, Barkatsas AN. 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. 2011;1(1):1–15.

14. Dauphinee TL, Schau C, Stevens JJ. Survey of attitudes toward statistics: factor structure and factorial invariance for women and men. Struct Equ Modeling. 1997;4(2):129–41. https://doi.org/10.1080/10705519709540066

15. Tempelaar DT, Van DerLoeff SS, Gijselaers WH. A structural equation model analyzing the relationship of students' attitudes toward statistics, prior reasoning abilities and course performance. Statistics Education Research Journal. 2007;6(2):78–102.

16. Vanhoof S, Kuppens S, Sotos AEC, Verschaffel L, Onghena P. Measuring statistics attitudes: structure of the survey of attitudes toward statistics (SATS-36). Statistics Education Research Journal. 2011;10(1):35–51.

17. Brown TA. Confirmatory factor analysis for applied research. 2nd ed. New York: The Guilford Press; 2017.

18. Harshe DG, Abraham DA. A study of attitudes of teaching faculty and postgraduate residents at a tertiary care teaching hospital toward biostatistics. Muller J Med Sci Res. 2017;8(1):10–4. https://doi.org/10.4103/0975-9727.199362

19. Lawrence NK. Attitudes and attitude change. In: Davis SF, Buskist W. 21st century psychology: a reference handbook (vol. 2). California: SAGE Publications; 2008. https://doi.org/10.4135/9781412956321.n63

20. Linstone HA, Turoff M. The Delphi method: techniques and applications; 2002 [cited 4 June 2017]. Available from: https://web.njit.edu/~turoff/pubs/delphibook/delphibook.pdf

21. Fabrigar LR, Wegener DT, MacCallum RC, Strahan EJ. Evaluating the use of exploratory factor analysis in psychological research. Psychol Methods. 1999;4(3):272–99. https://doi.org/10.1037/1082-989X.4.3.272

22. Hair Jr. JF, Black WC, Babin BJ, Anderson RE. Multivariate data analysis. 7th ed. Upper Saddle River, NJ: Pearson Prentice-Hall; 2009.

23. Bland JM, Altman DG. Cronbach's alpha. Br Med J. 1997;314:572. https://doi.org/10.1136/bmj.314.7080.572

24. Guadagnoli E, Velicer WF. Relation to sample size to the stability of component patterns. Psychol Bull. 1988;103(2):265–75. https://doi.org/10.1037/0033-2909.103.2.265

25. Harrington D. Confirmatory factor analysis. New York: Oxford University Press; 2008. https://doi.org/10.1093/acprof:oso/9780195339888.001.0001

26. Arifin WN. The graphical assessment of multivariate normality using SPSS. Education in Medicine Journal. 2015;7(2):e71–5.

27. Bentler PM. EQS 6 structural equations program manual. Encino, CA: Multivariate Software Inc; 2006.

28. Burdenski T. Evaluating univariate, bivariate, and multivariate normality using graphical and statistical procedures. Multiple linear regression viewpoints. 2000;26(2):15–28.

29. Schreiber JB, Nora A, Stage FK, Barlow EA, King J. Reporting structural equation modelling and confirmatory factor analysis results: a review. The Journal of Educational Research. 2006;99(6):323–38. https://doi.org/10.3200/JOER.99.6.323-338

30. Fornell C, Larcker DF. Evaluating structural equation models with unobservable variables and measurement error. J Mark Res. 1981;18(1):39–50. https://doi.org/10.1177/002224378101800104

31. Muthén LK, Muthén BO. How to use a Monte Carlo study to decide on sample size and determine power. Struct Equ Modelling. 2002;9(4):599–620. https://doi.org/10.1207/S15328007SEM0904_8

32. Weir JP. Quantifying test-retest reliability using the intraclass correlation coefficient and the SEM. J Strength Cond Res. 2005;19(1):231–40.

33. Cicchetti DV. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. Psychol Assess. 1994;6(4):284–90. https://doi.org/10.1037/1040-3590.6.4.284

34. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. Stat Med. 1998;17:101–10. https://doi.org/10.1002/(SICI)1097-0258(19980115)17:1<101::AID-SIM727>3.0.CO;2-E

35. Scheffer J. Dealing with missing data. Research Letters in the Information and Mathematical Sciences. 2002;3:153–60.

36. Tabachnick BG, Fidell LS. Using multivariate statistics. 5th ed. Boston: Pearson Education; 2007.

37. Messick S. Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. Am Psychol. 1995;50(9):741–9. https://doi.org/10.1037/0003-066X.50.9.741

38. American Educational Research Association, American Psychological Association, and National Council on Measurement in Education. Standards for educational and psychological testing. Washington, DC: American Educational Research Association; 1999.

39. Cook DA, Beckman TJ. Current concepts in validity and reliability for psychometric instruments: theory and application. Am J Med. 2006;119(2):166.e7–16.