

Statistics Anxiety Rating Scale (STARS) use in Psychology students: A review and analysis with an undergraduate sample

Rachel J. Nesbit & Victoria J. Bourne

Statistics anxiety is extremely common in undergraduate psychology students. The Statistics Anxiety Rating Scale (STARS) is at present the most widely used measure to assess statistics anxiety, measuring six distinct scales: test and class anxiety, interpretation anxiety, fear of asking for help, worth of statistics, fear of statistics teachers and computational self-concept. In this paper we first review the existing research that uses the STARS with psychology undergraduates. We then provide an analysis of the factor and reliability analysis of the STARS measure using a sample of undergraduate psychology students (N=315). Factor analysis of the STARS yielded nine factors, rather than the six it is intended to measure, with some items indicating low reliability, as demonstrated by low factor loadings. On the basis of these data, we consider the further development and refinement of measures of statistics anxiety in psychology students.

Keywords: STARS, psychology, statistics anxiety.

Introduction

IN THE UK, psychology is an increasingly popular subject for undergraduate students, with more than 80,000 undergraduate students across the UK in 2016–2017 (HESA). Despite the popularity of psychology in higher education many new entrants are unaware of the statistical components of their course (Ruggeri et al., 2008). Indeed, it has been reported that up to 80 per cent of psychology undergraduate students report experiencing statistics anxiety (Onwuegbuzie & Wilson, 2003), and rate statistics as the most challenging part of their psychology degree (Barry, 2012).

The Statistics Anxiety Rating scale (STARS; Cruise, Cash & Bolton, 1985) is currently the most widely used measure to assess statistics anxiety and has been revised for UK students (Hanna et al., 2008). The STARS contains 51-items split across six scales, arranged in two sections. The first section assesses statistics anxiety and consists of three scales (23 items): test and class anxiety, interpretation anxiety and fear of

asking for help. The second section of the STARS measures attitudes towards statistics and consists of three scales (28 items): worth of statistics, fear of statistics tutors and computational self-concept.

The initial factor structure of the STARS identified six distinct factors (Cruise et al., 1985), as did the UK adaption (Hanna et al., 2008), validating the six scales that were initially devised. Additionally, some researchers have found evidence of two second order superordinate factors (Papousek et al., 2010): statistics anxiety (test and class anxiety, interpretation anxiety and fear of asking for help) and attitudes towards statistics (worth of statistics, fear of statistics tutors and computational self-concept). This has led some researchers to select out only specific scales from the STARS to use in research. For example, some have used only the three scales relating to anxiety (e.g. Chew & Dillon, 2014; Macher et al., 2012), whilst other studies have used a composite score across the six scales (e.g. Chiou et al., 2014) or all six components of the STARS

(e.g. Chew & Dillon, 2014; Hanna & Dempster, 2009).

The use of the separate scales has been particularly useful when looking at individual differences in academic performance, attitudes and personality. For example, Hanna and Dempster (2009) found that only two of the STARS scales, computational self-concept and fear of asking for help, were associated with a student's own prediction of their performance in a statistics assessment, whereas the worth of statistics and interpretation anxiety scales of the STARS were significant predictors of student's actual performance. Further to this, group workshops to improve statistics anxiety and attitudes found improvements in the attitudes scales, but not in the anxiety scales (Hood & Neumann, 2013). Different scales of the STARS measure have also been differentially associated with personality. For example, agreeableness was found to be negatively correlated with worth of statistics, fear of asking for help and fear of statistics teachers, whereas extraversion was found to be positively related to interpretation anxiety, test and class anxiety and fear of asking for help (Chew & Dillon, 2014).

It is clear that the STARS is a useful research tool, both as an overall measure and divided into individual scales, and in this paper we aim to further explore the psychometric characteristics of STARS. First, this is achieved by reviewing the STARS scale scores and reliability in research that has used undergraduate psychology students. Second, we analyse a large dataset collected through the STARS measure, on an undergraduate psychology sample, where we consider both the factor structure and reliability of the STARS.

Review of published research

Selection criteria

A Google scholar search was carried out using the phrase 'Statistics anxiety rating scale', which produced 220 hits. These were then reduced in number by applying the selection criteria that papers should be

published empirical research papers, using solely undergraduate psychology students, with data collected through the use of the STARS measure, and where descriptive statistics were provided for at least some of the STARS scales (not solely composite scores).

Summary

Eleven articles were identified that matched our criteria. The descriptive statistics and reliability analysis can be found in Table 1, where reported. It is evident from the table that scale scores differed across samples, with some samples demonstrating higher levels of statistics anxiety and more negative attitudes towards statistics than others. Whilst in the majority of studies using the STARS scale scores are computed as the mean of the total for that scale, in some instances researchers have calculated a mean score as participant's average score on that specific scale (Hood & Neumann; Paechter et al., 2017).

Looking at the descriptive statistics, there is some variability in the scales across studies. However, due to the varying number of items per scale, it is difficult to compare scores across scales within studies. Considering scores within scales and across studies, some scales appear to have similar mean scores across the studies (e.g. test and class anxiety having scores from 26.53–30.01), whereas other have more variable scores (e.g. worth of statistics having scores from 34.89–57.12).

Importantly, some of the scale reliabilities are quite variable. Across all studies, the test and class anxiety scale consistently has alphas that indicate a good level of internal consistency and the worth of statistics scale was found to have excellent levels of internal consistency. However, other scales varied more in their consistency. For example, the interpretation anxiety scale had alpha levels indicating levels of internal consistency ranging from acceptable to excellent.

Table 1: Means, Standard Deviations and reliability statistics for STARS scales for all available published research papers using psychology undergraduate students

	Number of items:	Test and class anxiety		Interpretation anxiety		Fear of asking for help		Worth of statistics		Fear of statistics teachers		Computational self-concept	
		M (SD)	α	M (SD)	α	M (SD)	α	M (SD)	α	M (SD)	α	M (SD)	α
		8		11		4		16		5		7	
	Scale scores:	8-40		11-55		4-20		16-80		5-25		7-35	
	N												
1. Hanna, Shevlin & Dempster (2008)	650	27.07 (6.69)	.87	30.26 (8.42)	.87	10.39 (3.87)	.83	56.98 (14.13)	.94	18.20 (4.72)	.83	23.91 (6.60)	.87
2. Hanna & Dempster (2009)	52	26.53 (6.31)	-	31.50 (11.26)	-	11.27 (4.10)	-	57.12 (10.05)	-	19.43 (3.55)	-	22.87 (6.17)	-
3. Hamid & Sulaiman (2014)	139	30.01 (5.39)	.81	35.96 (5.57)	.75	13.32 (3.94)	.84	34.89 (11.29)	.91	20.48 (3.36)	.73	21.85 (5.97)	.82
4. Chew & Dillon (2014)	83	27.02 (6.34)	.88	29.82 (8.12)	.90	9.53 (3.32)	.83	39.39 (12.50)	.94	11.24 (4.45)	.86	18.05 (6.26)	.88
5. Ruggeri, Dempster, Hanna & Cleary (2008)+	158	27.08 (6.10)	-	31.48 (7.63)	-	8.80 (3.75)	-	41.76 (10.95)	-	11.48 (3.52)	-	20.75 (6.05)	-
6. Lester (2016)	93	28.40 (6.80)	.87	31.60 (9.10)	.91	10.10 (4.00)	.84	54.40 (13.50)	.94	19.70 (3.50)	.74	22.40 (7.00)	.88
7. Hood & Neumann (2013)**	113	2.93	.86	2.40	.88	2.33	.91	2.08	.94	1.81	.76	2.07	.88
8. Chew, Swinbourne & Dillon (2014)	76	28.12 (6.35)	.88	30.17 (8.49)	.90	9.93 (3.93)	.88	-	-	-	-	-	-
9. Chew, Dillon & Swinbourne (2018)	202	28.71 (6.51)	.89	31.08 (8.94)	.91	9.93 (4.07)	.88	39.09 (13.49)	.94	11.41 (4.20)	.81	18.77 (7.06)	.90
10. Paechter, Macher, Martskvishvili, Wimmer & Papousek (2017) **	224	2.57 (0.76)	-	1.72 (0.54)	-	2.10 (0.82)	-	-	-	-	-	-	-
11. Macher, Paechter, Papousek & Ruggeri (2012)	147	-	.85	-	.90	-	.85	-	-	-	-	-	-
Mean scores		27.86		31.48		10.41		46.23		15.99		21.23	

Note: + Statistics were reported separately for Year 1 (N = 158) and Year 2 (N = 38) students, Year 1 statistics reported but no significant differences reported p>.50. ** Scores calculated as the average of all items in this scale, instead of mean of scale total. Mean scale scores calculated for all papers using total scores only.

Factor and reliability analysis of STARS

Methods

Participants

The sample consisted of 315 undergraduate psychology students ($M_{age} = 19.4$, $SD = 3.2$), primarily in their first year of study ($N = 215$; Year 2 $N = 26$, Year 3 $N = 74$). The majority of the sample were female (87.9%, $N = 277$), representing the typical intake of psychology undergraduate students across the UK. This study was granted approval by the Royal Holloway College Ethics Committee.

Materials

Students completed a paper version of the STARS (Hanna et al., 2008). The STARS contains 51 items, across six scales. The first three scales measure statistics anxiety, (test and class anxiety, interpretation anxiety, fear of asking for help) with participants asked to rate the extent to which a given situation makes them feel anxious on a 5-point Likert scale, from not at all anxious through to extremely anxious. Within this section, test and class anxiety consisted of eight items, interpretation anxiety consisted of 11 items and fear of asking for help contained four

items. In the second section, participants were asked about their attitudes towards statistics through three scales (worth of statistics, fear of statistics teachers and computational self-concept). Students responded on a 5-point Likert scale from strongly disagree through to strongly agree. Worth of statistics contained 16 items, fear of statistics teachers consisted of five items and computational self-concept contained seven items. Descriptive statistics and reliability analysis for the 6 scales can be found in Table 2. In all instances higher scores reflected higher levels of statistics anxiety and more negative attitudes towards statistics.

Results

Descriptive statistics and correlations between all six STARS scales are presented in Table 2. In comparing these scores to those reported in previously published papers using Psychology undergraduate students (see Table 1), our scores are relatively low, showing that our sample are typically less anxious and have more positive attitudes than those reported elsewhere. Reliability statistics are comparable to those found in

Table 2: Descriptive statistics and correlations between all six STARS scales

	Mean (SD)	Cronbach's alpha	IA	AfH	WoS	FST	CSC
TCA	23.25 (7.07)	.89	.59**	.47**	.48**	.38**	.61**
IA	24.65 (7.38)	.87		.47**	.35**	.32**	.38**
AfH	8.17 (3.51)	.85			.18**	.26**	.26**
WoS	30.55 (10.19)	.92				.65**	.72**
FST	9.16 (3.33)	.73					.55**
CSC	16.47 (6.47)	.87					

Note: IA = interpretation anxiety, AfH = fear of asking for help, WoS = worth of statistics, FST = fear of statistics teachers, CSC = computational self-concept.

previous research. Cronbach’s alpha showed that internal consistency was excellent for the worth of statistics scale, acceptable for the fear of statistics teachers scale, and good for the remaining four scales.

Factor analysis

A factor analysis was performed using Principle Component Analysis (as used in the original STARS analysis; Cruise et al., 1985), with varimax rotation. Bartlett’s test of Sphericity is significant, $\chi^2(1275) = 8824.6$, $p < .001$, indicating that factor analysis is appropriate for this data set. The Kaiser-Meyer-Olkin statistic was high indicating sampling adequacy (KMO= .93). The determinant was $1.15E^{-13}$, therefore there is no

evidence of singularity in the data set. The assumptions suggest the data are suitable for factor analysis. In the presentation of our findings, factor loadings below .4 have been suppressed.

In contrast to the six factors extracted in previous studies, nine factors were extracted, explaining a total of 61.1 per cent of the variance in the STARS (See Table 3). Factor 1 contained 13 items – all of which belonged to the worth of statistics scale, three of the 16 worth of statistics items did not load on to this factor, with one item not loading on to any factor. Factor 2 contained 11 items, with 10 of the 11 interpretation anxiety items. The eleventh interpretation anxiety item loaded on to its own factor (9). Factor 3

Table 3: Factor analysis of the STARS measure (N = 315)

		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9
	Eigenvalue	15.15	5.23	2.53	2.03	1.49	1.40	1.21	1.08	1.03
	% variance explained	7.42	5.28	4.63	4.38	3.36	1.98	1.45	1.35	1.29
WoS_50		.776								
WoS_42		.747								
WoS_29		.713								
WoS_26		.713								
WoS_27		.701								
WoS_33		.699								
WoS_28		.689								
WoS_49		.663								
WoS_41		.618								
WoS_35		.574								
WoS_40		.522								
WoS_37		.482								
WoS_47		.473								
WoS_24										
WoS_36				.404						
WoS_45									.489	

Table 3: Factor analysis of the STARS measure (N = 315) *continued*

		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9
	Eigenvalue	15.15	5.23	2.53	2.03	1.49	1.40	1.21	1.08	1.03
	% variance explained	7.42	5.28	4.63	4.38	3.36	1.98	1.45	1.35	1.29
IA_06			.753							
IA_05			.728							
IA_20			.685							
IA_02			.638							
IA_11			.634							
IA_14			.568							
IA_07			.567							
IA_12			.564							
IA_09			.557							
IA_18			.436							
IA_17										.652
TCA_10				.764						
TCA_15				.759						
TCA_01				.718						
TCA_08				.716						
TCA_22				.560						
TCA_04			.466	.527						
TCA_13				.482						
TCA_21				.434						
CSC_31					.675					
CSC_39					.628					
CSC_25					.627					
CSC_38					.611					
CSC_48					.593					
CSC_51					.589					
CSC_34		.472			.573					

Table 3: Factor analysis of the STARS measure (N = 315) continued

		Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8	Factor 9
	Eigenvalue	15.15	5.23	2.53	2.03	1.49	1.40	1.21	1.08	1.03
	% variance explained	7.42	5.28	4.63	4.38	3.36	1.98	1.45	1.35	1.29
AfH_16						.823				
AfH_19						.787				
AfH_23						.752				
AfH_03						.740				
FoST_32							.723			
FoST_30							.695			
FoST_46					.408		.444			
FoST_44								.657		
FoST_43					.657					

contained nine items, with all eight of the test and class anxiety grouped together, and one item from the worth of statistics scale. Factor 4 contained nine items, including all seven of the computational self-concept items and two items from the fear of statistics teachers scale. Factor 5 contained four items, all of which belonged to the fear of asking for help scale. Factor 6 contained three items, all of which belonged to the fear of statistics teachers scale, with the remaining of the fear of statistics teachers items loading on to other factors. Factors 7, 8 and 9 each contained single items.

Discussion

The aims of this paper were to summarise the descriptive statistics and reliability analysis where available from previously published research articles using STARS with undergraduate psychology students. The summary in Table 1, demonstrates that there is variability in the scores on the six STARS scales depending on the sample used. Further, the reliability statistics range from acceptable to excellent, with some variability across studies. We also analysed the STARS scales

in our sample of undergraduate psychology students. In comparison to previous studies, our sample were typically less anxious and had more positive attitudes towards statistics. The reliability statistics were comparable to those reported in previous studies. Contrary to previous literature, our analysis produced nine factors in contrast to the six factors extracted elsewhere (Hanna et al., 2008). Whilst items within scales tended to factor together, there were a number of items that loaded onto different scales, and a number of items had low factor loadings between .4 – .5.

Test and class anxiety has eight items within the scale, and it is typically found to have a good level of internal consistency in both previous research and our own analysis. Our factor analysis showed that these eight items all factored together; however one item (Doing the coursework for a statistics course) also loaded onto the factor that primarily included the interpretation anxiety items. Generally, this scale seems robust.

Interpretation anxiety has eleven items within the scale, and whilst we found it to have a good level of internal consistency, the

previous research has found more variable statistics, ranging from acceptable to good levels of consistency. Ten of the eleven items factored together, with one item (Trying to understand the odds in a lottery) being placed in a factor of its own. Despite the large number of items contained within this scale, its reliability seems variable across samples.

The fear of asking for help scale contains only four items, however this scale tends to have good or excellent reliability and all four items factored together.

Worth of statistics contains the most items within STARS, with 16 items representing this scale. Both in the previous research and our analysis, this scale was found to have consistently excellent reliability. However, our factor analysis only placed 13 of the items together. One item (I am a subjective person, so the objectivity of statistics is inappropriate for me) did not load onto any of the nine factors. Another loaded onto the factor that mainly contained test and class anxiety items (Statistics is for people who have a natural leaning toward maths) and the third item not contained within the main worth of statistics factor was placed in a factor on its own (I cannot tell you why, but I just do not like statistics).

Fear of statistics teachers contains just five items, and this was the least reliable of the scales within our analysis, with the reliability being on the lower end of the acceptable criterion. In the previous research this scale also tended to find that its reliability is either on the lower end of the good criterion, or within acceptable levels. Our factor analysis also showed that these five items factored across three distinct factors. Three items were placed together, but one of these was also placed into the factor that mainly included computational self-concept items (Statistics teachers talk so fast you cannot logically follow them), along with one other item (Statistics teachers speak a different language). One further item loaded onto its own factor (Statisticians are more number oriented than they are people oriented).

Taking together these various sources of evidence, the fear of statistics teachers scale appears to be the most problematic from the six STARS scales.

Finally, computational self-concept is a scale that contains seven items, and it typically has a good level of internal consistency from the previous research and our own analysis. Further, all seven items loaded onto a single factor, although one item (Since I have never enjoyed maths I do not see how I can enjoy statistics) also loaded onto another factor that mainly contained items from the worth of statistics scale.

Looking across all six of the STARS scales, both the previous research and our own analysis has shown variability in the reliability and cohesiveness of each scale. Whilst some appear to have excellent reliability and factor well together (test and class anxiety, asking for help), others have lower levels of reliability and the factoring of items is less well defined (fear of statistics teachers). It therefore seems that, whilst the STARS has a number of strengths, some scales may benefit from further development and refinement.

With this in mind, it is interesting to consider further ways in which a measure of statistics anxiety and attitudes towards studying statistics, could be developed and improved. One difficulty in using STARS is the greatly varying number of items per scale, ranging from four to sixteen items. Typically scale scores are computed by summing the scores, but this can make the interpretation of student's anxieties and attitudes across the six scales of STARS difficult to interpret as the means and standard deviations vary a great deal across the scales simply by virtue of the summing of scores. Some papers have attempted to resolve this by using mean scale scores (e.g. Hood & Neumann, 2013; Paechter et al., 2017), but this then makes it difficult to compare scores across different studies. In future refinements of STARS, it may be beneficial to resolve this issue by ensuring that each scale contains the same number of items.

Increasingly, research studies around statistics anxiety are moving away from primarily measuring data at one time point (for example, looking at its relationship with academic performance or individual differences) towards using it as a measure of efficacy for interventions that are designed to alleviate anxiety and improve performance on statistical assessments. In studies where more than one measure of statistics anxiety is needed, there may be some issues with repeatedly using the same measure of statistics anxiety due to potential cross over (practice or fatigue) effects. As a result, if future work were to further develop the STARS measure, it may be helpful to develop two parallel forms of the questionnaire to enable interventions to be effectively assessed.

Finally, all of the items are marked in the same direction. That is, selecting the option to the far right of the Likert scale always indicates being extremely anxious or having negative attitudes towards statistics. Having all of the items scored in the same way may result in higher levels of acquiescence bias (Cheung & Rensvold, 2000), potentially biasing the findings towards increased levels of anxiety and more negative attitudes towards learning about statistics.

Whilst the STARS is the most frequently used measure of statistics anxiety and attitudes towards learning about statistics within the population of psychology undergraduate students, it clearly has a number of strengths, but also some weaknesses. Whilst some scales have been repeatedly found to have a high or excellent level of internal consistency, this cannot be said for all six scales. Additionally,

the variable number of items across the six scales can make the interpretation of differences across scales more challenging, and the lack of negatively scored items may lead to levels of statistics anxiety being overestimated in some studies. Finally, the lack of a second parallel form of the measure may be problematic for longitudinal studies or studies assessing the efficacy of interventions. Future work that aims to develop, refine and improve the measurement of statistics anxiety and attitudes towards learning about statistics should consider these factors to enable a more valid and reliable measurement of these variables.

Note

We refer readers to the Hanna et al., (2008) paper for the STARS items that have been used and correspond to the item numbers used in this paper.

Acknowledgements

This paper is in association with a Division of Academics, Researchers and Teachers in Psychology, Early-Career Bursary Prize to support attendance to the DART-P Inaugural conference, Birmingham, UK.

The authors

Rachel J. Nesbit & Victoria J. Bourne

Department of Psychology, Royal Holloway, University of London

Correspondence

Rachel Nesbit, Department of Psychology, Royal Holloway, University of London
Email: rachel.nesbit@rhul.ac.uk

References

- Barry, J.A. (2012). Do students who get low grades only in research methods need the same help as students who get low grades in all topics in psychology? *Psychology Teaching Review*, 18(2), 116–125.
- Cheung, G.W. & Rensvold, R.B. (2000). Assessing extreme and acquiescence response sets in cross-cultural research using structural equations modeling. *Journal of Cross-Cultural Psychology*, 31(2), 187–212.
- Chew, P.K.H., Dillon, D.B. & Swinbourne, A.L. (2018). An examination of the internal consistency and structure of the Statistical Anxiety Rating Scale (STARS). *PLoS ONE*, 13(3). <https://doi.org/10.1371/journal.pone.0194195>
- Chew, P.K., Swinbourne, A. & Dillon, D.B. (2014). No evidence of attentional bias in statistics anxiety. *The European Journal of Social & Behavioural Sciences*, 10, 1451–1465.
- Chiou, C.C., Wang, Y.M. & Lee, L.T. (2014). Reducing statistics anxiety and enhancing statistics learning achievement: Effectiveness of a one-minute strategy. *Psychological reports*, 115(1), 297–310.
- Cruise, R.J., Cash, R.W. & Bolton, D.L. (1985, August). Development and validation of an instrument to measure statistical anxiety. In *American Statistical Association Proceedings of the Section on Statistics Education* (Vol. 4, pp.92–97).
- Druggeri, K., Dempster, M., Hanna, D. & Cleary, C. (2008). Experiences and expectations: The real reason nobody likes stats. *Psychology Teaching Review*, 14(2), 75–83.
- Hamid, H.S.A. & Sulaiman, M.K. (2014). Statistics anxiety and achievement in a statistics course among psychology students. *International Journal of Behavioral Science*, 9(1), 55–66.
- Hanna, D. & Dempster, M. (2009). The effect of statistics anxiety on students' predicted and actual test scores. *The Irish Journal of Psychology*, 30(3-4), 201–209.
- Hanna, D., Shevlin, M. & Dempster, M. (2008). The structure of the statistics anxiety rating scale: A confirmatory factor analysis using UK psychology students. *Personality and Individual Differences*, 45(1), 68–74.
- HESA. HE student enrolments by subject of study and domicile. Retrieved from 2016/17, <https://www.hesa.ac.uk/data-and-analysis/students/what-study>
- Hood, M. & Neumann, D.L. (2013). Evaluation of a workshop to reduce negative perceptions of statistics in undergraduate psychology students. *Psychology Learning & Teaching*, 12(2), 115–125.
- Lester, D. (2016). Predicting success in psychological statistics courses. *Psychological reports*, 118(3), 772–777.
- Macher, D., Paechter, M., Papousek, I., & Ruggeri, K. (2012). Statistics anxiety, trait anxiety, learning behavior, and academic performance. *European journal of psychology of education*, 27(4), 483–498.
- Onwuegbuzie, A.J. & Wilson, V.A. (2003). Statistics anxiety: Nature, etiology, antecedents, effects, and treatments—a comprehensive review of the literature. *Teaching in Higher Education*, 8(2), 195–209.
- Paechter, M., Macher, D., Martskvishvili, K., Wimmer, S. & Papousek, I. (2017). Mathematics anxiety and statistics anxiety. Shared but also unshared components and antagonistic contributions to performance in statistics. *Frontiers in psychology*, 8, 1196.
- Papousek, I., Ruggeri, K., Macher, D. et al. (2012). Psychometric evaluation and experimental validation of the statistics anxiety rating scale. *Journal of Personality Assessment*, 94(1), 82–91.
- Ruggeri, K., Dempster, M., Hanna, D. & Cleary, C. (2008). Experiences and expectations: The real reason nobody likes stats. *Psychology Teaching Review*, 14(2), 75–83.