

The foreign language classroom anxiety scale in the South African context: Drumming to a different beat?



North-West University, South Africa E-mail: carina.grobler@nwu.ac.za



North-West University, South Africa E-mail: erika.fourie@nwu.ac.za

ABSTRACT

Anxiety in language learning has been studied for decades. One of the most widely used instruments allowing researchers to investigate this phenomenon is the Foreign Language Classroom Anxiety Scale (FLCAS). Only a limited number of studies investigated psychometric have the properties of the questionnaire to gauge the validity (factor analyses) and reliability (Cronbach Alphas) of the indicated factors and items for their specific context. The present study adopted a wider approach by including Rasch analyses to adjust the FLCAS for use in the South African context. Before submitting the questionnaire to 124 first-year students in French a translated version was prepared to ensure cultural adaptability.

The investigation was iterative and was conducted in two cycles which included exploratory and confirmatory factor analyses, Cronbach's Alpha and Rasch analyses to arrive at a proposed model to ensure sound measure-ment of foreign language anxiety for the specific context.

Results point to three factors with 25 relevant items. The original FLCAS was thus improved for the specific context by eliminating problematic items highlighted by the analyses and regrouping the remaining items into the three uniquely South African factors. This represents a first analysis from a Rasch analysis perspective in this context.

Keywords: FLCAS; foreign language anxiety; speaking anxiety; psychometric properties of questionnaires; validation; translation

CITATION

Grobler, C., & Fourie, E. (2025). The foreign language classroom anxiety scale in the South African context: Drumming to a different beat? *Journal for Language Teaching, 59*(1), Article 6604. https://doi.org/10.56285/jltVol59iss1a6604

ISSN: 0259-9570 | eISSN: 2958-9320

[©] Copyright belongs to the authors under the Creative Commons License CC BY-NC-ND 4.0

The Foreign Language Classroom Anxiety Scale (FLCAS) compiled by Horwitz et al. (1986) is derived from a 33-item questionnaire that provides information about the levels of anxiety regarding communication apprehension, fear of negative evaluation, and test anxiety of the person completing it. The FLCAS has been widely used since its introduction in 1986. Previous studies on the scale used it to measure the levels of anxiety of learners of foreign languages [cf. for example Çağatay (2015); Chen (2023); Pan and Lou (2023); Philips (1992); Simons and Decoo (2007); Weiwei (2023)]. Others used the FLCAS in part [cf. Al-Saraj (2014); Lee (2011); Awan (2010); Duong and Nguyen (2023)], whereas some focused on the psychometric properties of the questionnaire and its validation [cf. amongst others Cheng et al. (1999); Hung and Swanto (2023); Matsuda and Gobel (2004); Tóth (2008)].

Investigations into its psychometric properties do not, however, propose clear-cut factors that indicate exactly what is tested by the FLCAS, nor is there agreement on the item composition of the different factors, a limitation which reinforces the importance of studying and validating the FLCAS when using it in a specific context. It is imperative to adopt a context-specific approach to ensure that the correct instrument is used to assess the levels of specific types of anxiety in learners.

Thus, the situation-specific approach to assessing FLA [foreign language anxiety] is both consistent with the definition of FLA and advantageous for researchers, facilitating the isolation of FLA from other types of anxiety and allowing for the identification of limited anxiety prompts within the learning context (Al-Saraj, 2014, p. 53).

The objective of the study reported here is to evaluate the psychometric properties of the FLCAS in a South African context with the aim of determining the psychometric quality of the questionnaire for use in this specific context. The processes described here could be valuable to teams consisting of teachers of foreign languages and statisticians who wish to investigate in a systematic manner learners' levels of foreign language anxiety within a specific context.

2. Description and nature of the study

The FLCAS was investigated as part of a broader empirical study into the suitability of a design-model for a technology-enhanced out-of-class practice environment that focuses on the development of the oral communication skills of beginner language students (Grobler, 2020). The investigation into student characteristics formed part of a

multistrand design-based research project which adopted a pragmatic paradigm. The first strand was quantitative in nature and included the necessary steps to prepare different instruments for the local context. These instruments included the FLCAS, the computer-user self-efficacy scale (CUSE), and the self-directed learning inventory (SDLI), amongst others. The study was conducted with a group of beginners in French as a foreign language (FFL) within the South African higher education context. Applications were submitted to the North-West University's Arts research ethics committee, who granted permission for the study to be initiated and concluded (ethics clearance numbers NWU-00074-14-A7 and NWU-01494-19-A7). At the beginning of the research, each student also a signed consent form which informed them that participation would be voluntary, that they could withdraw at any time, and that their identity and the information they provided would be treated as confidential. Moreover, before completing each of the questionnaires, participants again granted the researcher permission to use the data collected for research purposes. In addition, at the beginning of the focus group interviews, students once again completed a consent form. The lecturer-researcher was never present when consent forms and questionnaires were distributed and collected, and data were processed and stored by independent parties.

The investigation into the FLCAS was not limited to doing a factor analysis and determining the reliability in respect of Cronbach's Alpha but included Rasch analyses. This was done to strengthen our understanding of and confidence in the results of the FLCAS in this study involving more than a hundred first-year students of FFL at the North-West University (NWU). The data collected were anonymised before being submitted for statistical analysis. Removing information that could divulge the identify of participants had no influence on the scholarly meaning of the research.

It is important (a) to investigate the uni/multidimensionality of the questionnaire to identify the constructs that are being tested by the instrument, and (b) to ensure that results of scores are correctly investigated and interpreted (Panayides & Walker, 2013, p. 2). More specifically, there is a need to (c) identify the items related to anxiety specifically when confronted with speaking the foreign language. This construct will play a central role in a broader study on ways to improve beginner students' oral communication skills in French, a language rarely encountered outside of the classroom in the context of this study. The value of this article lies in the steps followed to investigate the instrument. This methodology could be of value to foreign language teachers anywhere.

3. Method of investigation

Before administering the 33-item questionnaire to obtain the data needed to compile an investigation into its psychometric properties, it underwent a process to ensure a cultural adaptation to the local context (Sperber, 2004). This process included several steps, the first of which was to translate the original English version into the target language (Afrikaans) and then to have it back-translated into English by a different translator. The original source-language version was then compared with the back-translated source-language version by having each item ranked in terms of comparability of language and similarity of interpretability. This ranking was done by 30 raters who are fluent in the source language. Potentially problematic items that were identified by this process were reassessed and retranslated to ensure that these items would be interpreted in the same way in both languages (Sperber, 2004, p. 126). This was done after having received the consent of Dr Elaine Horwitz to use and translate the original FLCAS.

The final Afrikaans and English versions of the questionnaire were administered to a sample of 124 first-year French students at the North-West University. The data on the returned multiple-choice cards were run through a card reader in order to prepare the data in a format that could be used for the statistical analyses. Only the results from students who indicated that they had not done French in the past have been retained. This data set enabled further investigation into the validity and reliability of the FLCAS for the South African context.

Table 1 presents a summary of the results of some previous studies on the psychometric properties of the FLCAS showing the descriptive statistics and the number of factors identified with the labels assigned by the respective authors. Following each factor label, the items that are associated with each of the factors are listed. The results of the present study are shown in the last column.

	Aida (1994)	Cheng et al. (1999)	Pérez-Paredes et al. (2000-2001)	Matsuda et al. (2004)	Tóth (2008)	Cao (2011)	Mak (2011)	Panayides et al. (2013)	Hung & Swanto (2023)	This study
Sample size	96	423	198	252	117 (66)	300	313	304	100	124
Language	Japanese	English	English	English	English majors (non-English majors)	English	English	English	English	French
Sample age	University students	University students	14-65	University students	18-24 (18-39)	University students	University students	16-18 years	Vocational college students	University students
Country of study	USA	Taiwan	Spain	Japan	Hungary	China	China	Cyprus	Malysia	South Africa
Overall Alpha	0.94	0.95	0.89	0.78	0.93 (0.92)	0.95	0.91	0.96	0.92	0.93
Range of scores	47–146	Not reported	49–140	Not reported	Not reported	Not reported	Not reported	34–153	Not reported	26–123
Mean score	96.7/165 (58.6%)	Not reported	89.07/165 (54%)	100.75/165 (61%)	Not reported	Not reported	80.09/132 (60.6%)	76.9/165 (46.6%)	Not reported	76/125 (60.8%)
Standard deviation	22.1	Not reported	18.98	7.3	Not reported	Not reported	Not reported	25.9	Not reported	17
Resulting factors and associated items (Cf. Horwitz et al. 1986 p.129 for a full description of the items)	1. Speech anxiety and fear of negative evaluation: 1, 3, 4, 7, 8, 9, 12, 13, 16, 18, 20, 21, 23, 24, 27, 29, 31, 33	1. Low confidence in speaking English: 1, 2, 7, 13, 14, 18, 23, 24, 27, 31	2. Communication apprehension: 1, 3, 9, 12, 13, 18, 20, 24, 27, 31, 33	1. General English classroom performance anxiety: 2, 3, 4, 8, 9, 10, 12, 13, 14, 16, 19, 20, 22, 24, 25, 26, 27, 29, 30, 33	2. Communication apprehension: 1, 2, 3, 4, 9, 12, 13, 14, 18, 20, 24, 26, 27, 28, 29, 31, 32, 33	Communication apprehension: 1, 4, 9, 14, 15, 18, 24, 27, 29, 30, 32	1. Speech anxiety and fear of negative evaluation: 27, 3, 9, 31, 4, 33, 12, 13, 19, 24, 26, 29, 16, 1, 20	Instrument was found to be uni- dimensional. However, the following items were thematically regrouped by the authors: 1. Test anxiety: 8, 10, 19, 21	1. Communication apprehension: 1, 2, 3, 6, 8, 13, 14, 15	1. Speech anxiety in the classroom: 1, 3, 9, 13, 18, 20, 23, 24, 27, 28, 31, 33

Table 1. Summary of results of studies on the psychometric properties of the FLCAS: descriptive statistics, factors identified and its associated item numbers

Resulting factors and associated items	2. Fear of failing: 10, 22, 25, 26	2. General English classroom performance anxiety / Negative thoughts and emotional responses in stressful and formal English- learning situations: 4, 10, 15, 19, 20, 21, 25, 30, 29, 33	2. Anxiety about foreign language learning processes and situations: 4, 7, 15, 16, 23, 25, 29, 30	2. Low self- confidence in speaking English: 1, 5, 6, 7, 10, 17, 18, 21, 23, 28, 32	2. Fear of inadequate performance: 5, 7, 8, 10, 11, 16, 21, 23, 25, 30	2. Test anxiety: 3, 5, 6, 8, 10, 11, 12, 16, 17, 20, 21, 22, 25, 26, 28	2. Uncomfortable- ness when speaking with native speakers: 8, 11, 14, 32	2. Uneasiness while speaking in a foreign language classroom: 1, 9, 18, 20, 24, 27	2. Fear of negative evaluation: 1, 3, 4, 5, 6, 7	2. Communication with native speakers: 14, 32
Resulting factors and associated items	3. Comfortable- ness with Japanese: 11, 14, 32		3. Comfortable- ness in using English inside and outside the classroom: 8, 14, 32			3. Fear of negative evaluation: 2, 7, 13, 19, 23, 31, 33	3. Negative attitudes towards the English class: 5, 6, 17	3. Comfort/ uneasiness of students in speaking to native English speakers: 14, 32	3. Test anxiety: 1, 3, 5, 6, 8	3. Anxiety about foreign language learning processes and situations (Pérez-Parades, 2000-2001): 4, 7, 8, 12, 15, 16, 19, 21, 25, 29, 30
Resulting factors and associated items	4. Negative attitudes: 5, 17		4. Negative attitudes toward learning English: 6, 17				4. Negative self- evaluation: 7, 23			
Resulting factors and associated items							5. Fear of failing the class/Conseque nces of personal failure: 10, 15, 22			
Items from original FLCAS not retained	2, 6, 15, 19, 28, 30	3, 5, 6, 8, 9, 11, 12, 16, 17, 22, 26, 28, 32	2, 5, 10, 11, 19, 21, 22, 26, 28	11, 15	6, 15, 17, 19, 22	(All items retained)	2, 18, 21, 25, 28, 30	2, 5, 6, 17, 22	11, 13, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29 30, 31, 32, 33	2, 5, 6, 10, 11, 17, 22, 26

Journal for Language Teaching | Ijenali Yekufundzisa Lulwimi | Tydskrif vir Taalonderrig

ISSN: 0259-9570 | eISSN: 2958-9320

To obtain the results for this study presented in the far-right column of Table 1, the psychometric properties of the FLCAS were analysed in terms of an exploratory factor analysis (EFA), confirmatory factor analysis (CFA), reliability and Rasch analyses. The EFA and reliability tests were conducted using SPSS Inc. (*IBM SPSS Statistics Version 23, Release 23.0.0, 2016*), the CFA in Amos Development Company [*Amos 23.0.0 (Build 817), 2016*] and the Rasch analysis in Winsteps [*Winsteps*© *Rasch Measurement (Version 3.63.2), 2006*)]. The analyses were prepared in two cycles. The first cycle consisted of the following:

- a) Determining frequencies and descriptive statistics these give an overview of the data obtained from questionnaires;
- b) Exploratory factor analysis (EFA) to identify factors representative of the underlying relationships within the data;
- c) Confirmatory factor analysis (CFA) to confirm the best model for this data set (i.e. one, two or three factor models);
- d) Reliability (Cronbach's Alpha) to test the reliability of the factors resulting from the previous analyses in terms of the number of factors extracted;
- e) Rasch analysis (RA) includes various elements such as unidimensionality of each factor, the functionality of the response categories to determine to what extent the data fits the model (i.e. one, two of three factor model) that emerged from the CFA.

The second cycle included:

- a) Rasch analysis testing to what extent the data fits the model resulting from the RA in the first cycle
- b) Reliability (Cronbach's Alpha) confirming the reliability of the new proposed model
- c) Confirmatory factor analysis to re-evaluate the goodness of fit statistics

Details of the investigation are discussed in the next sections.

4. Results: first cycle of investigation

The application of the methods outlined above and the subsequent results will be discussed in the order in which the analyses were conducted.

Frequencies and descriptive statistics

Frequencies and descriptive statistics (means and standard deviations) reported per item provides an overview of the data. Most of the items in this study have an equal

distribution of respondents choosing the different options of the rating scale (Strongly Agree = 1, Strongly Disagree = 5) (cf. Table 2). However, about 80% of the respondents chose Disagree and Strongly Disagree for items 5 and 17 and Strongly Agree and Agree for item 10. The item mostly agreed upon is statement 10 (Mean = 1.94, Standard Deviation (SD) = 1.07) and the most disagreed upon item is statement 5 (Mean = 4.10, SD = 1.14). Based on these distributions, it was decided to exclude items 5, 10 and 17 (closely related to item 5) from the subsequent analysis.

	Frequencies							Descriptives	
Item (Cf. Horwitz et al. 1986, p.129 for a full description of the items)	% Strongly agree (SA)	% Agree (A)	% Neither agree nor disagree (N)	% Disagree (D)	% Strongly disagree (SD)	# Missing	Mean	SD	
1.	8,06	41,94	24,19	20,16	5,65	0	2,73	1,05	
2.	12,90	31,45	7,26	30,65	17,74	0	3,09	1,36	
3.	5,69	30,08	10,57	37,40	16,26	1	3,28	1,22	
4.	9,68	27,42	22,58	33,06	7,26	0	3,01	1,14	
5.	6,45	5,65	7,26	32,26	48,39	0	4,10	1,17	
6.	2,44	10,57	18,70	46,34	21,95	1	3,75	1,00	
7.	8,87	28,23	29,84	21,77	11,29	0	2,98	1,15	
8.	6,50	20,33	35,77	29,27	8,13	1	3,12	1,04	
9.	17,74	27,42	29,03	20,97	4,84	0	2,68	1,14	
10.	40,32	39,52	10,48	4,84	4,84	0	1,94	1,07	
11.	4,03	11,29	45,16	24,19	15,32	0	3,35	1,01	
12.	11,48	26,23	20,49	30,33	11,48	2	3,04	1,22	
13.	4,88	26,02	20,33	35,77	13,01	1	3,26	1,13	
14.	20,97	34,68	14,52	22,58	7,26	0	2,60	1,25	
15.	5,69	27,64	26,02	26,83	13,82	1	3,15	1,15	
16.	8,87	29,03	19,35	28,23	14,52	0	3,10	1,23	
17.	1,64	5,74	10,66	34,43	47,54	2	4,20	0,96	
18.	8,87	30,65	34,68	19,35	6,45	0	2,84	1,05	
19.	2,42	7,26	14,52	45,97	29,84	0	3,94	0,98	
20.	10,48	31,45	22,58	25,00	10,48	0	2,94	1,19	
21.	3,28	5,74	27,05	37,70	26,23	2	3,78	1,01	
22.	8,87	38,71	26,61	17,74	8,06	0	2,77	1,10	
23.	8,06	27,42	39,52	21,77	3,23	0	2,85	0,96	

	Table 2. Des	scriptive statistics	of individual	items of the	questionnaire
--	--------------	----------------------	---------------	--------------	---------------

Journal for Language Teaching | Ijenali Yekufundzisa Lulwimi | Tydskrif vir Taalonderrig ISSN: 0259-9570 | eISSN: 2958-9320

24.	10,57	36,59	21,95	26,83	4,07	1	2,77	1,09
25.	12,40	38,02	14,88	25,62	9,09	3	2,81	1,21
26.	6,45	16,13	16,94	37,90	22,58	0	3,54	1,19
27.	4,10	26,23	22,95	38,52	8,20	2	3,20	1,05
28.	2,44	13,82	30,08	34,96	18,70	1	3,54	1,03
29.	8,94	33,33	21,14	29,27	7,32	1	2,93	1,13
30.	15,32	29,84	27,42	20,97	6,45	0	2,73	1,15
31.	6,45	20,16	21,77	30,65	20,97	0	3,40	1,21
32.	10,48	35,48	34,68	13,71	5,65	0	2,69	1,02
33.	12,10	42,74	20,97	19,35	4,84	0	2,62	1,08

Exploratory factor analysis

An exploratory factor analysis (EFA) was conducted to investigate the underlying factors of the FLCAS in the South African context. It was decided to perform an EFA first because of the contradictory results from previous studies – different factors with varying items (See Table 1).

An assessment of the suitability of the data for a factor analysis (FA) was conducted using the Kaiser-Meier-Olkin (KMO) measure, which determines sampling adequacy and Bartlett's test, which measures whether the correlation between the items is high enough (Field, 2014, p. 685; Pallant, 2005, p. 174). According to Field (2014, p. 685), a KMO above 0.8 is deemed as very good. Thus, the resulting KMO of 0.874 of this study indicates that the sample size is sufficient to form compact factors. Bartlett's test measures the p-value that must be less than 0.05 to reflect adequate correlation. The reported p-value of p < 0.0001 shows high enough correlation between the items to continue with the FA.

If all the items in a questionnaire contribute to a general factor, all must have a factor loading or a communality of at least 0.3 on the first factor of the unrotated matrix (Field, 2014, pp. 681 - 682). The communalities reflect the proportion of the variances of the items explained by the factors. Item 22 does not contribute to the first factor of the unrotated matrix. In other words, if the FLCAS is viewed as a general factor, item 22 does not form part of this since the communality is 0.136 and only 13.6% of the variance of item 22 is explained by the extracted factors.

The pattern matrix contains the loadings of the items on the factors after the Oblimin with Kaiser Normalization method was applied. These loadings reflect the relative contribution of an item to its specific factor (or subscale) and should be 0.3 and higher

(Field, 2014, pp. 672 and 681 – 682). The pattern matrix revealed that the optimal number of factors extracted from the data is three. This makes theoretical sense considering the findings of Cao (2011) that a three-factor structure has the best fit. The rest of this investigation also focused on three factors. The factor loadings reported in the pattern matrix indicated that item 26 loaded onto all factors (loadings ranging between 0.311 - 0.345) when three factors have been extracted.

Considering the results discussed above, items 22 and 26 were to be omitted subsequently.

The three extracted factors relate to 'Speech anxiety in the classroom', 'Communication with native speakers' and 'Anxiety about foreign language learning processes and situations'. They relate to the items listed in brackets:

- 'Speech anxiety in the classroom' (items 1, 3, 9, 13, 18, 20, 23, 24, 27, 28, 31 and 33),
- 'Communication with native speakers' (items 11, 14 and 32) and
- 'Anxiety about foreign language learning processes and situations' (items 2, 4, 6, 7, 8, 12, 15, 16, 19, 21, 25, 29 and 30).

The total variance explained by the extracted factors is 45.71%. It was thus required at this stage to confirm whether one, two or three factors fit(s) the data best. A confirmatory factor analysis was thus conducted.

Confirmatory factor analysis

A Confirmatory Factor Analysis (CFA) by means of a Structural Equation Model (SEM) was conducted to test the fitness of the different possible models (i.e., one, two or three factors). Goodness-of-fit statistics comprise different measures that indicate how well the covariance structure predicted by the model corresponds to the covariance structure in the data (Cao, 2011, p. 79). These fit measures can be grouped into five broad categories, namely, absolute fit measures, relative fit measures, parsimony-based fit measures, fit indices based on the non-central chi-squared-distribution and information-theoretic fit measures (Blunch, 2008, pp. 113 – 116). It is advisable to report indices from at least three of the broad categories. For the purpose of this study five indices from four different categories were reported. The reported indices are chi-squared test statistic divided by the Degrees of Freedom (CMIN/DF), Comparative Fit Index (CFI), Root Mean Square Error Approximation (RMSEA) with a 90% confidence interval (LO 90 and HO 90), Akaike Information Criterion (AIC) and Browne-Cudeck

Criterion (BCC). The guideline values provided by Blunch (2008, pp. 113 - 116) and the relevant values are indicated in Table 3.

Model Guideline:	CMIN/DF Near 1	CFI Larger than 0.95	RMSEA Less than 0.10	AIC As small as possible	BCC As small as possible
1 Factor	1.845	0.783	0.082 (0.072; 0.092)	813.642	864.392
2 Factors	1.785	0.799	0.079 (0.069; 0.089)	793.054	844.408
3 Factors	1.681	0.827	0.074 (0.063; 0.084)	757.265	809.828

Table 3. Goodness-of-fit indices for the three possible models

It is evident from Table 3 that the three-factor model fits the data best. This model was investigated further in terms of reliability, as discussed below.

Reliability

Cronbach's Alpha is used to test the reliability of the factors resulting from the exploratory factor analysis (cf. above, *Exploratory factor analysis*). A Cronbach's Alpha coefficient of 0.7 and above is generally accepted as an indication of reliability, but in the early stages of research – as in this case – values of 0.5 or above will also be sufficient (Field, 2014, pp. 708 – 709). Table 4 indicates the Cronbach Alphas for each of the extracted factors, which are therefore deemed to be reliable, the lowest Cronbach's Alpha measure being 0.534.

Table 4. Cronbach's Alpha values per factor

Factor	Cronbach's Alpha
Guideline	>= 0.7
Speech anxiety in the classroom	0.898
Communication with native speakers	0.875
Anxiety about foreign language learning processes and situations	0.534

Rasch analysis

Although validity and reliability traditionally were tested using factor analysis and Cronbach Alphas, the investigation regarding the psychometric properties of a measuring scale could be supplemented with a Rasch analysis (Tennant & Conaghan, 2007, p. 1358). Rasch analysis transforms ordinal data into continuous, equal-interval units (logits). This allows for the summation of the raw scores of individual items, measuring a latent trait, into a continuum. This enables the researcher to conduct person-to-person, item-to-item and person-to-item analysis when evaluating the functioning of a rating scale (Boone et al., 2014, pp. 7 and 70). The current research focused on item functionality.

Rasch analysis is a unified approach to evaluate several measurement issues related to validity and reliability, for example: the unidimensionality of the scale, differential item functioning and appropriate category ordering. The various tests conducted are interrelated and should be considered simultaneously (Schutte et al., 2016, p. 4).

Unidimensionality of the FLCAS. Rasch analysis, and more specifically, a Rasch principal component analysis of the residuals were used to confirm undeniably whether or not more than one factor should be extracted from this specific data set in terms of the FLCAS.

The reported eigenvalue when performing a Rasch analysis on the FLCAS as a unidimensional factor is 2.46. Additional factors may exist if the eigenvalue of the first contrast is larger than 2.0 (Linacre, 2014, p. 521; Schutte et al., 2016, p. 4). Thus, the Rasch analysis once again confirms the decision to extract three factors (as indicated by the EFA) instead of only one factor, since this proves that the latent trait is indeed not unidimensional. Therefore, Rasch analysis was conducted for each of the three factors to investigate the psychometric properties (i.e. validity and reliability) of each of the factors.

Analysis per factor. These analyses identified items that might be problematic in the adaptation of the FLCAS for the South African context. The relevant Rasch analysis measures are reported simultaneously for all three factors, after which the findings will be discussed per factor. This will explain the reasons for the elimination of problematic items.

Unidimensionality. The first measure under review when conducting a Rasch analysis for individual factors is also the unidimensional. All the eigenvalues were lower than the recommended value of 2.0 (cf. Table 5). Therefore, each of the individual factors is unidimensional.

	1
Factor	Eigenvalue
Guideline	<= 2
Speech anxiety in the classroom	1.71
Communication with native speakers	1.69
Anxiety about foreign language learning processes and situations	1.79

Table 5. Eigenvalues when conducting a Rasch principal component analysis of the residuals for each of the factors

Differential Item Functioning (DIF) contrast. Rasch analysis could also be used to determine if item biases exist between groups (e.g., participants who have different first languages). Differential DIF implies that an item measures a specific characteristic in different ways for different groups (Boone et al., 2014, pp. 273 – 275). The guideline value provided is at least 0.5 logits for a difference to be noticeable, more specifically a value ≥ 0.43 indicates a slight to moderate contrast and > 0.64 a moderate to large contrast (Boone et al., 2014, pp. 273 – 275). The accompanied Mantel-Haenszel p-values should be less than 0.05, which indicates the probability of observing the same amount of DIF when systematic item bias does not exist (Boone et al., 2014, pp. 273 – 275).

Possible bias between groups is considered an important aspect in terms of this study since participants from two language groups (English and Afrikaans) completed the questionnaire in their own language. The noteworthy DIF contrasts and their associated p-values are shown in Table 6. Slight to moderate DIF contrast were reported for item 27 of the 'Speech anxiety in the classroom' and item 30 of the 'Anxiety about foreign language learning processes and situations' factor. The only moderate to large DIF contrast was reported for item 6 of the 'Anxiety about foreign language learning processes and situations' factor, indicating potential differentiation between the two language groups for the three items mentioned above.

Factor name and item number <i>Guideline</i>	DIF contrast (p-value) >= 0.43 slight to moderate > 0.5 noticeable > 0.64 moderate to large	p-value < 0.05
Speech anxiety in the classroom – item 27	0.42	0.0393
Anxiety about foreign language learning processes and situations – item 6	0.68	0.0044
Anxiety about foreign language learning processes and situations – item 30	0.46	0.0420

Table 6. Noteworthy DIF contrast and associated p-value per item

Journal for Language Teaching | Ijenali Yekufundzisa Lulwimi | Tydskrif vir Taalonderrig ISSN: 0259-9570 | eISSN: 2958-9320

Item fit statistics. Item fit statistics measure the association between the model and the data, i.e. how well the data fit the model (Boone et al., 2014, p. 164). The infit statistic focuses on person ability whereas the outfit statistic is sensitive to outliers. The infit and outfit mean square statistics (MNSQ) were preferred for reporting purposes for this study. Linacre (2014, p. 212) states that only values greater than 1.5 need to be reported as this indicates too much unexplained variance within the data. Therefore, only the infit and outfit statistics of items 2, 6 and 33 (per language group) are documented in Table 7. Only one of the four statistics of item 33 does not adhere to the guideline.

Factor name and item number	Infit M	NSQ	Outfit MNSQ			
Guideline	0.5 > measure < 1.5					
	Afrikaans	English	Afrikaans	English		
Speech anxiety in the classroom – item 33	1.49	0.83	1.58	0.80		
Anxiety about foreign language learning processes and situations – item 2	1.68	1.68	1.70	1.61		
Anxiety about foreign language learning processes and situations – item 6	0.99	1.51	0.85	1.65		

Rating category diagnostics. Since the majority of the studies thus far have used a 5-point scale and not an experimental 4-point scale as suggested by Panayides and Walker (2013, p. 13), we have opted to retain the 5-point scale to increase comparability of the results obtained between the current study and previous studies with 1 meaning 'Strongly Agree' and 5 meaning 'Strongly disagree'.

Rasch analysis also provides information to aid with the investigation regarding the optimal number of rating categories. Information resulting from the Rasch analysis includes the frequency distribution, the average measures, the threshold estimates as well as infit and outfit measures per category (Schutte et al., 2016). A uniform frequency distribution across rating categories, each category containing at least 10 observations is preferred (Schutte et al., 2016). The mean measure (ability mean) of the categories should increase monotonically. The infit and outfit MNSQ for each rating category should be less than 2.

Due to the relatively small sample size (n = 124), a number of items had rating categories with frequencies less than 10, therefore only the two items with disordered mean measures patterns and infit and outfit MNSQ less than two are reported in Table 8. Item 6 barely violates the guideline values; thus, no further investigation is required. However, when reviewing the original instrument for item 11 as suggested by Boone et

al. (2014, p. 204), it was discovered that the item was negatively phrased. This might explain the unwanted pattern in the rating scale data.

Factor name and item number <i>Guideline</i>	Rating category*	Frequency (#) > 10	Mean measure Increase mono- tonically	Infit MNSQ < 2	Outfit MNSQ < 2
Communication with	SA	5	-1.64	2.6	3.0
native speakers – item 11	A	14	-1.66	0.7	0.7
	N	54	-0.40	1.2	1.2
	D	30	0.23	1.3	1.3
	SD	18	1.31	1.2	1.2
Anxiety about foreign	SA	3	-0.38	1.7	2.0
language learning	А	11	-0.39	1.2	1.2
processes and situations –	N	19	-0.07	1.3	1.3
item 6	D	46	0.20	1.5	1.7
	SD	22	1.23	1.3	1.2
	Missing	1	-0.06		

 Table 8. Rating category diagnostics

* SA = Strongly Agree, A = Agree, N= Neutral, D = Disagree and SD = Strongly Disagree

An additional facet when investigating the functioning of a rating scale is the probability that a specific rating category will be chosen. Category probability curves can be used in this investigation as these curves indicate the probability of a specific rating category being chosen relative to item difficulty in a graphical form. In perfect circumstances these graphs should form a 'hill' for each of the numbered categories (Boone et al., 2014, p. 194; Linacre, 2014, p. 51).

	+				-+-			-+				_				_	+-			_	-+					+-	
1.0.+																										.+	
																										- 1	
																							-			- 1	
																										51	
·.8 -+-11 ···														*										- 5	55	·+	
11																							- !	5 -		- 1	
	1													-								- 5	5			- 1	
	-11						1						4.4	-							1	5 -				- 1	
. 6 . +	1 -			222			• •				• •				• •		• •			- 5	5					+	
]	1	1	222		222	2 -							44	41	44	44			œ,	5						- 1	
		-122				2	• •		• •		• •)	44			• •		4	4 -	5			• •			• •	+	
]		221 -				- 2	22		• •		4	• •						5+	4							- 1	
4 -+	2	1			* *		1	2 -		- 4					• •		5			44						+	
	22		11 -				3	3+	3+	+3	3 -				•	- 5	5 -				4			•		-1	
]	22		1		* *	33	33				3	3 -			1	5		n . 4				44	•		* *	- 1	
22				11 -	- 33	3	• •	- 4	- 2	2	• •	- 3	3 -	- 1	55		• •						44	4 -	• •	1	
.2 + 22 -				+	3 -			4 -		- 2	2 -		- 3	+	•	-								- 4	4	+	
· · · · 2 · · · ·			* * *	33 -	11	4	44				22	2 -	55		33		• •	• •		• •				• •	• 4	41	
· · · ·] · · · · ·			333		* * *	++1	+	• •	• •	•	5	5+	2 -		• • •	33	53			• •				• •	•	- 1	
	3	3333		- 4 4	44		1	11	+ 5	55	5 -		- 2	22	22	• •	• •	33	53	33	5 -			* *	-	- 1	
0 -++++++	+++++	++++	+++	+ 55	55	555	55	55	- 1	11	11	11	11	11	11	+ +	++	+ +	+	+ +	+	+ +	+ -	++	++	+++	
	+		+		-+-								+			_	+-					+-		-	_	+-	
	-3 -			• • •	-1		• •	- 0) •	•	• •	• 1		*	• •	• •	2			• •	13	5 -		• •		4	



16 of 24

Figure 1. Category probability curves for 'Speech anxiety in the classroom' (a), 'Communication with native speakers' (b) and 'Anxiety about foreign language learning processes and situations' (c)

It is evident from Figure 1 that each rating category (1 - 5) peaks at some stage along the continuum for the first two factors (a & b), thus the 5-point Likert scale is considered to be optimal. The category probability curve of the 'Anxiety about foreign language learning processes and situations' factor did not illustrate the required trend since rating category 3 (Neither agree nor disagree) does not peak.

Item Separation and Reliability. The last aspects that need to be reviewed are Item Separation and Reliability. Separation is used to verify the item hierarchy, simply put, to verify the measurement effectiveness of the continuum. Item Reliability relates to the discrimination between person performance levels. Low item Separation (values smaller than three) and low item Reliability (values smaller than 0.9) imply that the data sample is not large enough to confirm the measurement effectiveness or the item difficulty hierarchy of the instrument (Linacre, 2014, pp. 79 and 620). Real values provide the lower limits, and model values the upper limits of these measures (Boone et al., 2014, p. 222).

Factor	Separatio	on	Reliability					
Guideline	>= 3		>= 0.9					
	Real	Model	Real	Model				
Speech anxiety in the classroom	3.75	3.86	0.93	0.94				
Communication with native speakers	4.46	4.65	0.95	0.96				
Anxiety about foreign language learning processes and situations	4.34	4.56	0.95	0.95				

Table 9. Item Separation and Reliability measurement values per factor

The Separation and Reliability measures of all three factors are in line with the guideline values (Table 9). Thus, the data sample is deemed large enough to confirm measurement effectiveness. Based on the adequate Item Separation and Reliability measures reported above, it will be assumed that the data sample was large enough to draw conclusions regarding the psychometric properties of the FLCAS based on the Rasch analysis conducted.

Discussion of cycle 1

As noted previously, the information provided by Rasch analysis should be reviewed simultaneously in order to determine what the next step should be, i.e. which items should be investigated further and how.

Considering that only items 27 and 33 of the 'Speech anxiety in the classroom' factor were flagged, in terms of DIF contrast and item fit statistics respectively, but no serious issues were reported, thus the 'Speech anxiety in the classroom' factor does not require any further investigation.

Item 11 of the 'Communication with native speakers' factor displayed a disordered mean measures pattern. The infit and outfit MNSQ measures of the Strongly Agree rating category were above the guideline value of 2 (2.6 and 3.0 respectively). This factor will be investigated further by considering the removal of item 11 from the factor.

Items 2, 6 and 30 of the 'Anxiety about foreign language learning processes and situations' factor reported some violation of the guideline values provided. Items 6 and 30 were flagged in terms of DIF contrast, item 2 and 6 in terms of item fit statistics, and item 6 (again) in terms of rating category diagnostics. All four of the item fit statistics of item 2 were larger than 1.5, indicating that the data does not fit the model for this particular item. Item 6 displayed moderate to large bias between the two language groups (based on the reported DIF contrast of 0.68), item fit statistics of 1.51 and 1.65 for the English group and an unfavourable mean measure pattern. Since the DIF contrast reported for item 30 is considered to be moderate this item will not be investigated any

further. Therefore, the next step in this particular investigation is to consider removing items 2 and 6 from the factor. This decision is reinforced by the undesired pattern of the rating category probability curve.

5. Results: second cycle of investigation

The Rasch analysis, reliability tests and confirmatory factor analysis were re-conducted for the factors 'Communication with native speakers' and 'Anxiety about foreign language learning processes and situations' without the problematic items identified in the previous section (i.e., items 2, 6, 11, 30).

Rasch analysis

The newly calculated Eigenvalues of the 'Communication with native speakers' and 'Anxiety about foreign language learning processes and situations' factors were still below the guideline value of 2, indicating that the factors are still unidimensional after the removal of the above-mentioned items.

The largest DIF contrast reported was 0.48 (p-value = 0.1488) for item 30 of the 'Anxiety about foreign language learning processes and situations' factor, thus no moderate to large DIF contrasts. This indicates that there was no differentiation between the two language groups. No exceptional item fit statistics were indicated. Thus, none of the items were flagged in terms of too much variance in the data that are unexplained by the model.

The mean measure (ability mean) of all items increases monotonically across the five rating categories. Also, each of the rating categories peaked for some range of the continuum when considering the category probability curves. This illustrates that the rating scale is optimal. Item 19 of the 'Anxiety about foreign language learning processes and situations' factor reflected an Outfit MNSQ of 2.4 which is slightly high, but no further investigation is required since no other issues were raised in terms of rating category diagnostics.

The measures for the factor 'Anxiety about foreign language learning processes and situations' are in line with the guideline values, with the Separation being 4.66 (Real) and 4.80 (Model) and Reliability being 0.96 (for both Real and Model). Thus, the data sample is deemed large enough to confirm measurement effectiveness in this case too. Separation and Reliability is not applicable in the 'Communication with native speakers' analysis since this factor consists of only two items.

No concern regarding the psychometric properties of the FLCAS based on the Rasch analysis conducted was raised during the further investigation. Thus, the reliability of the factors needs to be confirmed next.

Reliability

The Cronbach Alphas for the 'Communication with native speakers' and 'Anxiety about foreign language learning processes and situations' factors without items 2, 6 and 11 were recalculated. The reliability of both factors improved from 0.875 to 0.882 for the 'Communication with native speakers' factor and from 0.537 to 0.627 for the 'Anxiety about foreign language learning processes and situations' factor. Thus, the reliability of the factors improved when omitting the three items. The model fit of the three-factor model without items 2, 6 and 11 was tested again by means of a CFA.

Confirmatory factor analysis

The recalculated goodness-of-fit statistics still indicated a relatively good fit to the data. The recalculated measure reported is CMIN / DF = 1.779, CFI = 0.837 and RMSEA = 0.079 (0.067; 0.090).

All items included in the model under investigation contributed significantly to the relevant factors, the largest p-value being 0.001 and the smallest standard regression weight being 0.456.

Thus, the final model is:

- 'Speech anxiety in the classroom' (items 1, 3, 9, 13, 18, 20, 23, 24, 27, 28, 31 and 33),
- 'Communication with native speakers' (items 14 and 32) and
- 'Anxiety about foreign language learning processes and situations' (items 4, 7, 8, 12, 15, 16, 19, 21, 25, 29 and 30).

The correlations between the factors are as follows (cf. Figure 2):

- 'Speech anxiety in the classroom' (speech) and 'Communication with native speakers' (natives): r=-0.231
- 'Speech anxiety in the classroom' (speech) and 'Anxiety about foreign language learning processes and situations' (fll): r=0.391
- 'Communication with native speakers' (natives) and 'Anxiety about foreign language learning processes and situations' (fll): r=-0.262

Due to the way the questions were formulated, the correlations indicate that as the anxiety levels of respondents increased on one aspect, it increased on the other aspects as well.



Figure 2. Graphic representation of the final model of the FLCAS for the South African context

6. Discussion and conclusion

The objective of this study was to evaluate the psychometric properties of the FLCAS, specifically in a South African context, in order to identify the different constructs that are tested by the instrument, to verify whether and how speaking anxiety was tested and to ensure that the scores obtained for the different constructs are interpreted correctly. It was administered to a sample of 124 novice learners of French at the NWU.

From the frequencies per question, it was clear that the distributions of items 5, 10 and 17 necessitated their removal before conducting the next step in the analysis. An EFA was deemed necessary since the literature showed contradictory results in previous studies. The EFA indicated three factors, which were labelled 'Speech anxiety in the classroom', 'Communication with native speakers' and 'Anxiety about foreign language learning processes and situations'. Results from the EFA led to the removal from further investigation of items 22 and 26.

Elaboration on the fit of the three-factor model suggested by the EFA confirmed it via a CFA by means of an SEM model, after which Cronbach's Alpha was calculated for each of the factors to determine and confirm their reliability.

A Rasch analysis suggested that by omitting one item (11) from the 'Communication with native speakers' factor and two items (2 and 6) from the 'Anxiety about foreign language learning processes and situations' factor, it might be possible to improve the model. These items were therefore removed and the analyses (except for the EFA) were repeated. This resulted in a final model consisting of three factors with 25 items instead

of the original 33: 'Speech anxiety in the classroom' (12 items), 'Communication with native speakers' (2 items) and 'Anxiety about foreign language learning processes and situations' (11 items).

A validated and reliable version of the measuring instrument is thus ready for implementation in the context for which it was prepared. Because of our confidence in the value of the instrument in its new format, we used it with a group of beginner students in French for the purpose of measuring their levels of anxiety at different points during an academic semester: before each of the two written language tests and after the two individual orals that form part of their formal assessment. This allowed us to discover if there are any differences in anxiety levels in terms of the two types of assessment activities and whether the levels changed over time.

This, together with other relevant information on self-directed learning, computer-user self-efficacy, for example, would be used to build a profile of these learners in order to allow teachers to design a technology enhanced out-of-class speaking practice environment that is both suitable and beneficial in this context (cf. Grobler, 2020). The results could also be used in an investigation of the relation between speaking anxiety and speaking achievement for this specific group of learners in order to determine whether the expected negative correlations are confirmed.

Further investigation into the psychometric properties of the FLCAS in different contexts could contribute valuable information in the quest for a more standardised instrument for measuring the different aspects of anxiety in learners of foreign languages.

This study proposes guidelines for teams of researchers wanting to thoroughly investigate the psychometric properties of adapted or newly created questionnaires in any field of study, not only domains related to anxiety or language learning. This could be particularly valuable in the new phase of broader research into language anxiety – and more specifically language-skill-specific anxieties (Cheng, 2017, p. 2; Pan & Lou, 2023, p. 1762) – that is emerging. Language teachers are encouraged to work in multidisciplinary teams, such as that demonstrated by the authors of this paper, to increase the validity and strengthen the importance of research done in a language teaching and learning environment.

References

- Al-Saraj, T. M. (2014). Revisiting the Foreign Language Classroom Anxiety Scale (FLCAS): the anxiety of female English language learners in Saudi Arabia. *L2 Journal*, *6*(1), 50-76.
- Amos 23.0.0 (Build 817). (2016). IBM Corporation. http://amosdevelopment.com
- Awan, R.-U.-N., Azher, M., Anwar, M. N., & Naz, A. (2010). An investigation of foreign language classroom anxiety and its relationship with students' achievement. *Journal of College Teaching & Learning*, 7(11), 33-40.
- Blunch, N. J. (2008). *Introduction to structural equation modelling using SPSS and Amos*. Sage Publications.
- Boone, W. J., Staver, J. R., & Yale, M. S. (2014). Rasch analysis in the human sciences. Springer.
- Çağatay, S. (2015). Examining EFL students' foreign language speaking anxiety: The case at a Turkish state university. *Procedia Social and Behavioral Sciences*, 199, 648-656.
- Cao, Y. (2011). Comparison of two models of foreign language classroom anxiety scale. *Philippine ESL Journal*, 7, 73–93.
- Chen, I.-J. (2023). Effects of online international peer learning program on speaking ability and language anxiety. *The Educational Review, USA*, 7(8), 1052-1058. https://doi.org/10.26855/er.2023.08.002
- Cheng, Y.-s., Horwitz, E. K., & Schallert, D. L. (1999). Differentiating writing and speaking components. *Language Learning*, *49*(3), 3417-3446.
- Cheng, Y. (2017). Development and preliminary validation of four brief measures of L2 language-skill-specific anxiety. *System*, 1-11. https://doi.org/10.1016/j.system.2017.06.009
- Duong, L. T. T., & Nguyen, T. H. (2023). Freshman EFL learners' anxiety in speaking classes and coping strategies. *European Journal of Foreign Language Teaching*, 7(3), 32-44. https://doi.org/10.46827/ejfl.v7i3.5019
- Field, A. (2014). Discovering statistics using ISM SPSS Statistics. Sage Publications Ltd.
- Grobler, C. (2020). *Designing a model for a technology-enhanced environment developing the oral interactional competence of beginner language learners* (Doctoral thesis, University of Antwerp). https://repository.uantwerpen.be/docstore/d:irua:897
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, *70*(2), 125-132.
- Hung, R. W. C., & Swanto, S. (2023). Validating ESL speaking instrument for vocational college students: An exploratory factor analysis. In *Proceedings of Educational*

Sciences: I-RoLE 2023: International Conference of Research on Language Education (pp. 243-249). European Proceedings. https://doi.org/10.15405/epes.23097.22

- *IBM SPSS Statistics Version 23, Release 23.0.0.* (2016). IBM Corporation. http://www-01.ibm.com/software/analytics/spss/
- Lee, M.-L. (2011). Differences in the learning anxieties affecting college freshman students of EFL. *Global perspectives, local initiatives, 1*, 169-182.
- Linacre, J. M. (2014). A user's guide to Winsteps Ministep Rasch-model computer programs: Program manual 3.81.0. Winsteps.com.
- Matsuda, S., & Gobel, P. (2004). Anxiety and predictors of performance in the foreign language classroom. *System*, 32(1), 121-136.
- Pallant, J. (2005). SPSS Survival Manual (2nd ed.). Allen & Unwin.
- Pan, J., & Lou, L. (2023). Research on the phenomenon of "Foreign Language Anxiety" based on individual differences of learners. *Creative Education*, 14, 1759-1772. https://doi.org/10.4236/ce.2023.149113
- Panayides, P., & Walker, M. J. (2013). Evaluating the psychometric properties of the Foreign Language Classroom Anxiety Scale for Cypriot senior high school EFL students: The Rasch measurement approach. *Europe's Journal of Psychology*, 9(3), 493-516.
- Philips, E. M. (1992). The effects of language anxiety on students' oral test performance and attitudes. *The Modern Language Journal*, 76(1), 14-26.
- Schutte, L., Wissing, M.P., Ellis, S.M. et al. Rasch analysis of the Meaning in Life Questionnaire among adults from South Africa, Australia, and New Zealand. *Health and Quality of Life Outcomes* 14,12. https://doi.org/10.1186/s12955-016-0414-x
- Simons, M., & Decoo, W. (2007). Cette peur d'ouvrir la bouche: Comment vaincre l'anxiété dans l'apprentissage des langues ? *Le français dans le monde*, *352*, 40-42.
- Sperber, A. D. (2004). Translation and validation of study instruments for cross-cultural research. *Gastroenterology*, *126 Supplement 1*, 124-128.
- Tennant, A., & Conaghan, P. G. (2007). The Rasch measurement model in rheumatology: What is it and why use it? When should it be applied, and what should one look for in a Rasch paper? *Arthritis & Rheumatism (Arthritis Care & Research)*, 57(8), 1358–1362.
- Tóth, Z. (2008). A foreign language anxiety scale for Hungarian learners of English. *Working Papers in Language Pedagogy*, 2, 55-77.
- Weiwei, L. (2023). An exploration of foreign language reading anxiety of young EFL students. Journal of English Language Teaching and Applied Linguistics, 5(3), 101-107. https://doi.org/10.32996/jeltal.2023.5.3.12
- Winsteps© Rasch Measurement (Version 3.63.2). (2006). John M. Linacre. www.winsteps.com.

ABOUT THE AUTHORS

Carina Grobler

North-West University, South Africa

Email: carina.grobler@nwu.ac.za ORCID: https://orcid.org/0000-0001-7451-7115

Carina Grobler is a senior lecturer in French at the NWU. She started her teaching career in 2006 at the University of the Free State and moved to the NWU in 2009. She obtained her master's degree from the Université Stendhal (France) and her PhD at the University of Antwerp (Belgium). Her research interests include instructional design, artificial intelligence, open educational resources (OER) and multilingualism.

Erika Fourie

North-West University, South Africa

Email: erika.fourie@nwu.ac.za ORCID: https://orcid.org/0000-0003-3321-0576

Erika Fourie joined the NWU's Statistical Consulting Services in 2010 where she has since been involved in various research projects and postgraduate studies. She completed her PhD in Risk Analysis in 2015 and became a lecturer in the School of Mathematical and Statistical Sciences in 2017. She remains active at the NWU as a senior subject specialist and lecturer.