



Serbian Version of Foreign Language Classroom Anxiety Scale: Psychometric Analysis

Jelisaveta Safranĳ¹, Marija Volarov², Milan Oljaca³

Abstract

The validity and psychometric properties of Serbian translation of the Foreign Language Learning Anxiety Scale (Horwitz, Horwitz, & Cope, 1986) were investigated. The scale was administered to 296 undergraduate engineering students. The aim was the comparison of several competing models related to factor structure of FLCAS. The following models were tested: (1) four-factor model, (2) bifactor model with four specific factors, (3) three-factor model, (4) bifactor model with three specific factors, (5) two-factor model and (6) bifactor model with two specific factors. Bifactor model seemed to be the best solution according to all fit indices that were used. This is the first study with the focus on Serbian version of FLACS, and the first study in general with sophisticated statistical procedures (e.g. IRT, testing of bifactor models) that were used in order to address conceptual problems regarding the FLCA. The results of this study provide information regarding factor structure of the scale and functioning of every single item from the IRT perspective and open some new questions for further investigation.

Keywords

Foreign Language Classroom
Anxiety Scale (FLCAS)
Foreign Language Classroom
Anxiety (FLCA)
Bifactor model
IRT
Validation

Article Info

Received: 03.22.2018
Accepted: 08.28.2019
Online Published: 10.08.2020

DOI: 10.15390/EB.2020.7826

Introduction

Foreign Language Learning Anxiety As A Distinct Construct

Foreign language classroom anxiety (FLCA) has been recognized as different from general anxiety identified in psychology; it is a situation-specific anxiety endured only in formal language learning environment (MacIntyre, 1999). It is described as feelings of apprehension, tension, and worry experienced in foreign language classes and it is supposed to have the substantial impact towards foreign language learning (Horwitz et al., 1986). It is defined by Horwitz et al. (1986) as "a distinct complex of self-perceptions, beliefs, feelings, and behaviors related to classroom language learning which arise from the uniqueness of the language learning process" (p. 128). In addition, psychology also recognizes three performance related anxieties: (1) communication apprehension, (2) test anxiety, and (3) fear of negative evaluation. Since MacIntyre and Gardner (1989) tested this original model of foreign language anxiety they concluded that the above mentioned three dimensions contribute to language anxiety. Therefore, it is important to study, comprehend and tackle this phenomenon as it contributes to an affective filter inhibiting the learner's ability to learn the target language and thus, foreign language acquisition is unlikely to advance.

¹ University of Novi Sad, Faculty of Technical Sciences, Department of Fundamental Sciences in Engineering, Serbia, jsafranĳ@eunet.rs

² University of Novi Sad, Faculty of Philosophy, Department of Psychology, Serbia, m.volarov@yahoo.com

³ University of Novi Sad, Faculty of Philosophy, Department of Psychology, Serbia, milanoljaca@ff.uns.ac.rs

The previous research showed that language anxious students often display avoidance behavior such as missing class and postponing homework (Horwitz et al., 1986). Some stated difficulties were the too fast pace of the class and that students feel left behind (MacIntyre & Gardner, 1991). In addition, anxiety may hinder retrieval of previously learned knowledge (Tobias, 1986). Unfortunately, although language-anxious students study more than their low-anxious peers, their level of language learning achievement often fails to show that endeavor (Horwitz et al., 1986; Price, 1991; Tsai & Li, 2012). This negative correlation between anxiety and language performance has been reported by many researchers (e.g., Chen & Chang, 2004; Elkahafi, 2005; Horwitz et al., 1986; Liu & Jackson, 2008; MacIntyre & Gardner, 1991). However, it is unlikely a unidimensional construct and has not been experienced equally in all modalities of communication (Baghaei, Hohensinn, & Kubinger, 2014). It is an omnipresent phenomenon, and thus, factors such as classroom context, social and cultural contexts should be considered due to understanding it properly.

Anxiety consistently corresponds with the settings and therefore, it should be expressed within a context (Oxford, 1999). It is also certain that cultural issues are a significant part of the settings. Various studies have been carried out (Daly, 1991; Horwitz et al., 1986; Price, 1991; Young, 1990) and they have shown that FLCA is present primarily in speaking and listening in the foreign language. However, these results are unlikely to be common across different cultures. Thus, the circumstances that initiate anxiety, socially acceptable behavior, and some other factors are empirical proof that can be fluctuated depending on wide cultural standards and classroom progress (Al-Saraj, 2014). Namely, when learners are silent during the class it could mean showing respect for the teacher or might be because of a certain level of anxiety. Thus, broader contextual factors, such as cross-cultural issues, need to be considered due to recognize and measure anxiety level.

Assessment of Foreign Language Learning Anxiety

In 1986 Horwitz et al. (1986) developed the Foreign Language Classroom Anxiety Scale (FLCAS) to measure anxiety specific to a foreign language classroom setting. According to the author of the scale in designing the instrument Horwitz et al. (1986) drew on measures of test anxiety (Sarason, 1984), speaking anxiety, and communication apprehension (McCroskey, 1970) as well as included five items from the French Class Anxiety Scale (Gardner, Clement, Smythe, & Smythe, 1979). Thus, the FLCAS consists of 33 items and uses a five-point Likert scale from "strongly disagree" to "strongly agree" with a neutral category "neither agree nor disagree" in the middle.

There is some research on gathering psychometric evidence of the FLCAS, with a focus on score reliability and validity. Many studies have proved the high internal consistency of the FLCAS measured by Cronbach's alpha, but the underlying factor structure that comprises the FLCAS as determined by exploratory factor analysis has been inconsistent across studies ranging from two to four-factor solutions (Aida, 1994; Horwitz et al., 1986; Khodadady & Khajavy, 2013; Liu & Jackson, 2008). These papers suggest that the scores are reliable but the number of factors contained in the FLCAS may be hesitant, depending on students on different competence levels and learning situation. Thus, the score validity of FLCAS may be different across participants of different learning settings such as cross-cultural issues. In previous studies, the reliability of the FLACS score in terms of internal consistency ranged from .89 to .93 (Horwitz et al., 1986; Paredes & Muller-Alouf, 2000; Tóth, 2008).

Many researchers have attempted to find the underlying components of the FLCAS by performing factor analysis since the first attempt by Aida (1994). Park (2014) explains three main reasons for different components of the FLCAS regardless of the continuous attempts across countries in the last three decades. First, Horwitz et al. (1986) did not clarify the components of the FLCAS, leading subsequent researchers to misinterpret the aforementioned three components of the measure (Aida, 1994; Cao, 2011; Liu & Jackson, 2008; Tóth, 2008). Second, to find the underlying components of the FLCAS most researchers have used only exploratory factor analysis, which could be criticized for subjective judgment in factor rotation and labeling. Third, previous studies have used different versions of the FLCAS by translating the original version into the native language of the participants (Cheng, Horwitz, & Schallert, 1999; Matsuda & Gobel, 2004; Park, 2012; Tóth, 2008).

The research of Cao (2011) confirmed the three-factor model of foreign language classroom anxiety scale (FLCAS) composed of three domains: Communication Apprehension, Test Anxiety, Fear of Negative Evaluation. These three domains were empirically derived through factor analysis and further confirmed having the best fit for the observation. Previous studies (i.e. Zhao, 2007) investigated FLCA by using the four-factor model of FLCAS. However, Cao (2011) showed that the three-factor model has better fit, and thus, it is considered as better model for investigating foreign language classroom anxiety.

The main purpose of the study of Park (2014) was to find the underlying components of the FLCAS by utilizing both exploratory factor analysis (EFA), which is used to generate theoretical components, and confirmatory factor analysis (CFA), which is used to examine whether *a priori* components generated by exploratory factor analysis adequately fit the data. Park's study suggests that the two-factor model in general adequately fit the data as the most optimal solution.

In addition, the study of Panayides and Walker (2013) deals with the psychometric properties of the FLCAS for Cypriot senior high school EFL students which were investigated through Rasch measurement (Bond & Fox, 2007). The Greek version of the FLCAS was administered to a sample of 304 senior high school EFL students. The researchers clarified two discrepancies found in the literature: first, the factor structure of the scale and second, whether test anxiety is a component of FLCA. The results showed that after removing five items which fitted the Rasch Rating Scale model poorly, the remaining 28 items formed a unidimensional scale, one component of which were test anxiety indicators. The degree of reliability was a high and semantic analysis of the items revealed that one of the reasons was the inclusion of many parallel items. The Rasch person-item map showed that a second reason was the narrow coverage of the construct by the items.

Aims of the Present Study

Main research problem addressed in this study was the identification of optimal latent factor structure of FLCAS. Taking into account inconsistent results of previous studies regarding factor structure of FLCAS this study was comprised in order to provide important information about the adequacy of using scores in the same manner as it was suggested by the authors of the scale (Horwitz et al., 1986). The main aim of this study was a comparison of several competing models related to factor structure of FLCAS using CFA in order to detect model which is most potent to fit the data gathered on a sample of students by using Serbian adaptation of FLCAS. Considering Parks's (2014) notions described above about potential causes of inconsistent results regarding latent factor structure, the great contribution of this study can be found in its tendency to solve one of the three problems described by Park (2014).

Following models were tested: (1) four-factor model, (2) bifactor model with four specific factors, (3) three-factor model, (4) bifactor model with three specific factors, (5) two-factor model and (6) bifactor model with two specific factors. Bifactor models are specific because after extracting general factor from all items, from the rest of the variance specific factors are extracted, and all these factors are uncorrelated with each other. This is useful in a practical sense because if the best fit is provided for the bifactor model it would mean that sometimes for more sophisticated results scores from specific factor should be used, but also that total score is probably useful and enough in some occasions. The important conceptual implication is that bifactor structure actually implies that the construct of interest is unidimensional.

Additionally, the convergent and divergent validity of factors from the best fitting model was tested in relation to IPIP Big Five personality dimensions. According to the previous findings, negative correlation with Emotional stability, Imagination and Extraversion were expected (i.e. Asmali, 2017).

This is important because it can provide additional information about the validity of obtained latent factor structure of the scale. In other words, if we can replicate relations between FLACS scores on the one hand, and Big Five personality dimensions on the other hand, that were suggested in the previous studies, it would provide additional support to the identified latent factor structure.

The greatest contribution of this study is reflected in the fact that there is no any study of this kind conducted using Serbian adaptation of FLCAS. Therefore, its purpose was twofold: to examine whether the theoretical construct of FLCA is applicable in Serbian cultural context since there are no tools for evaluating FLCA among populations who speak Serbian as a first language, and to construct a psychometrically sound adaptation of a scale that can be used with Serbian-speaking populations. The Serbian translation was based on the English version of the instrument. Besides, only a few studies have been acknowledged that is related to the comparison of different structural models but bifactor models have never been tested so far. Using bifactor model provides an opportunity for further development of FLCA concept. Finally, as far as it is known, this is the first study that evaluates psychometric properties of FLCAS items using Item response theory.

Method

Sample and Procedure

The study was carried out using a sample comprised of 296 engineering students at University of Novi Sad, in fall semester 2016. The sample comprised of 81 power engineering student (27.4%), 14 civil engineering student (4.7%), 41 mechanical engineering student (13.9) and 156 industrial management student (52.7%). Genders were represented equally with 50.3% female age ranged from 18 – 28 ($M = 20.16$, $SD = 1.40$). Minimal duration of language learning was 2 years and maximum 8 years ($M = 2$, $SD = 2.37$). All of the students were entirely informed about the study and freely accepted to participate without any compensation. They filled in the questionnaires in paper-and-pencil at the beginning of their ESP class. The whole procedure lasted for 20 minutes.

The data collecting tool consisted of a background questionnaire interrogating participants' gender, age and study-year, and the Serbian version of the Foreign Language Classroom Anxiety Scale (FLCAS) which was administered to the examinees. First, the English version of FLAS was translated into Serbian. The Serbian version consisted of 33 items with the same content and range as the original English version. The process of translation included two stages. In the first stage, three translators were asked to translate the English version of the FLAS into Serbian. The translators held the Ph.D. degree in English language teaching and were experienced in research on foreign language anxiety. They work at University of Novi Sad. After they were fully informed about their roles in the procedure, they translated the English version of the FLAS scale into Serbian in a blind session. In the second stage, the other three translators compared translated versions in a panel and aimed to unify those three into one. Finally, all of the translators reached a satisfactory similarity while focusing on semantic and conceptual equivalence. Eventually, they agreed that the Serbian version ensured the equivalence of meaning and certain ideas and notions in English language.

Instruments

Foreign Language Classroom Anxiety Scale (FLCAS; Horwitz et al., 1986) consists of 33 items followed by 5-points Likert scale (1- totally disagree, 5 – totally agree). The scale was constructed in order to measure anxiety which occurs while learning English as a foreign language and demonstrating that knowledge. It operationalizes 3 aspects of foreign language anxiety - Communication Apprehension (11 items, $\alpha = .86$), Test Anxiety (15 items, $\alpha = .86$) i Fear of negative feedback by peers and teachers (7 items, $\alpha = .81$). The total score can be also used ($\alpha = .94$). The higher the score, the more anxiety the students experience.

IPIP-50 (IPIP Big Five Broad Domains: Goldberg, 1992). This instrument represents an operationalization of psycholexical Big Five model which is available on IPIP repositorium. The instrument is used for measuring five broad personality traits: Agreeableness ($\alpha = .77$), Conscientiousness ($\alpha = .73$), Extraversion ($\alpha = .75$), Emotional stability ($\alpha = .86$) and Intellect ($\alpha = .71$). It consists of 50 items, 10 items per subscale, followed by 5-points Likert scale (1 – totally disagree, 5 – totally agree).

Data Analysis

Data preparation and preliminary analysis were conducted using SPSS v21 (IBM Corp. Released, 2012). In addition, using the same software correlations and gender differences were calculated.

Internal validity of the FLCAS was tested using confirmatory factor analysis (CFA) using the „lavaan“ package (Rosseel, 2012) within R software environment (R Core Team, 2016). Fit indices were estimated using maximum likelihood estimator (ML). Evaluation of model fit was provided using: comparative fit index (CFI), Tucker-Lewis index (TLI), the Root Mean Square Error of Approximation (RMSEA) and Standardized Root Mean Square Residual (SRMR). CFI – comparative fit index; optimal value $>.90$. TLI - Tucker-Lewis index; optimal value $>.90$. RMSEA - Root mean square error of approximation; optimal value $<.06$. SRMR - Standardized root mean square residual; optimal value $<.08$ (Hu & Bentler, 1999).

Psychometric properties of FLCAS items were tested via the analysis based on Item response theory (IRT) within „ltm“(Rizopoulos, 2006) package in R. Two-parameter model (2PL) for items with graded responses (GRM; Samejima, 1969) was used. GRM model is applied on scales which contain items with ordered response categories, such as the FLCAS (5-point Likert scale). Discrimination or the slope parameter (α) corresponds to factor loadings, represents the property of an item to discriminate between participants with different levels of the latent trait (DeMars, 2010; Morizot, Ainsworth, & Reise, 2007). The threshold parameter (beta), refers to the value of the latent trait (on a continuum with the mean of 0 and standard deviation of 1), where the participant has 50% likelihood of choosing a higher or lower response category (DeMars, 2010).

Results

Confirmatory Factor Analysis (CFA) and Item Response Theory (IRT)

CFA was used in order to test how models suggested in previous studies fit our data. The following models were tested: four-factor model (M1; according to Zhao, 2007, as cited in Cao, 2011), three-factor model (M3; Huang, 2008) and two-factor model (M5; Park, 2014). Very high correlations among all dimensions that were obtained in previous studies were taken into account, so for each of all proposed models conceptually equivalent bifactor model was tested, as well. According to results of CFA shown in Table 1, only model 6, with two specific (named Communication Apprehension and Understanding – CAU and Communication Apprehension and Confidence - CAC) and one general factor, had optimal fit indices within all applied criteria. Model is presented in Figure 1.

Table 1. Fit indices for all tested models¹

Model	χ^2	DF	CFI	TLI	RMSEA	SRMR
M1	1313.83	458	.813	.797	.079	.067
M2	1089.35	463	.870	.851	.068	.059
M3	1354.92	492	.820	.807	.077	.066
M4 *						
M5	616.29	229	.885	.873	.076	.060
M6	421.52	207	.936	.922	.059	.041

Note. M1 – four-factor model. M2 – Bifactor model with four specific factors. M3 – three-factor model. M4 – bifactor model with three specific factors. M5 – two-factor model. M6 – bifactor model with two specific factors. CFI – comparative fit index; optimal value $>.90$. TLI - Tucker-Lewis index; optimal value $>.90$. RMSEA - Root mean square error of approximation; optimal value $<.06$. SRMR - Standardized root mean square residual; optimal value $<.08$ (Hu & Bentler, 1999). * - The model does not converge.

¹ Models 5 and 6 were tested using 23 out of 33 FLACS items according to the results provided by Park (2014).

M6 - Bifactor model with two specific factors (CAC and CAU)

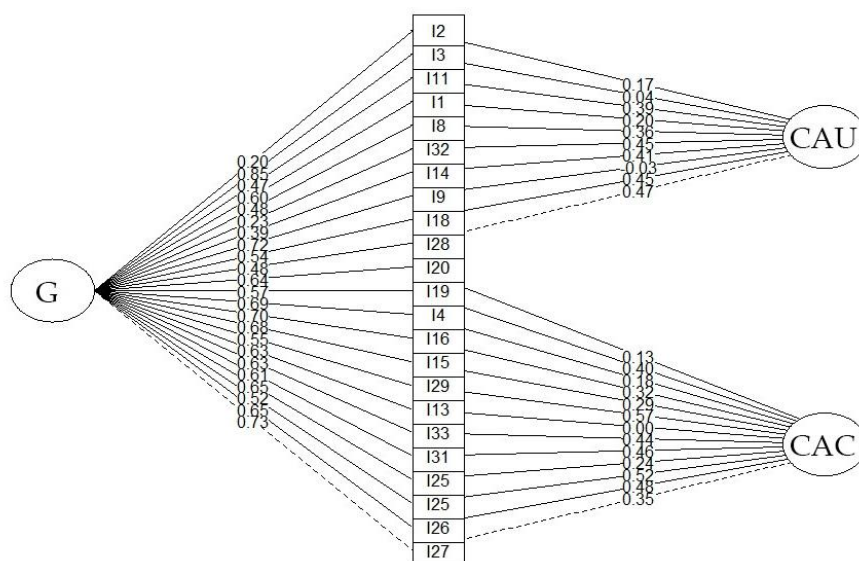


Figure 1. Structural representation of Model 6

Standardized factor loadings (for general and specific factors), as well as difficulty, discrimination, and information indices for FLCAS items, are shown in table 2. Most of the items from both dimensions demonstrated good functioning according to CFA and IRT indices. Items 13, 4 and 20 from CAU dimension were identified as problematic. Therefore, it is suggested for these items to be removed in future revisions of Serbian version of FLCAS. Considering CAU, all of these three items, especially item 13, had very low factor loadings. Items 4 and 20 have adequate IRT parameters but, since there are items with similar IRT parameters and higher factor loadings, they could also be removed due to shortening of the instrument.

On the other hand, considering CAC dimension, items 9, 3, 32 and 2 demonstrated poor psychometric properties. Items 9 and 3 have very low factor loadings, while items 32 and 2 have very poor information and discrimination. Considering these values, these four items should be removed in future revisions of the instrument.

Table 2. CFA and IRT Parameters for FLCAS Items

Item	G	CAU	CAC	β_1	β_2	β_3	β_4	a	I (-3/3)	%
									99.77	92.51%
27	.725	.346		-0.37	0.36	0.97	1.60	3.25	9.24	99.63%
26	.655	.480		-0.22	0.41	1.05	1.80	2.90	7.86	98.89%
25	.516	.516		-0.59	0.27	1.29	2.13	1.84	4.29	92.84%
12	.652	.243		-0.30	0.40	1.09	1.94	2.14	5.08	96.11%
31	.608	.462		-0.42	0.25	1.03	2.03	2.31	5.97	96.32%
33	.635	.444		-0.59	0.19	1.01	1.94	2.20	5.64	96.45%
13	.632	.000		-1.26	-0.31	0.80	1.75	1.40	2.92	90.09%
29	.552	.567		-0.62	0.18	1.20	2.18	1.96	4.82	93.33%
15	.679	.288		-0.61	0.20	1.20	2.24	2.08	5.33	93.53%
16	.698	.316		-0.31	0.40	1.17	2.06	2.69	7.43	97.37%
4	.690	.176		-0.19	0.52	1.22	2.20	2.10	4.96	93.70%
19	.571	.396		-0.68	0.24	1.24	2.11	1.90	4.60	93.53%
20	.636	.131		-0.83	-0.03	0.85	1.79	1.84	4.35	95.35%
28	.476		.466	-1.53	-0.08	1.26	2.54	1.24	2.50	80.04%

Table 2. Continued

Item	G	CAU	CAC	β_1	β_2	β_3	β_4	a	I (-3/3)	%
18	.540		.454	-1.95	-0.40	1.02	2.00	1.31	2.80	83.67%
9	.717		-.035	-0.89	0.00	0.85	1.67	1.74	3.96	95.22%
14	.393		.406	-1.53	-0.12	1.54	2.92	1.02	1.76	71.85%
32	.232		.455	-2.32	-0.19	2.29	4.47	0.69	0.86	51.30%
8	.484		.356	-1.09	0.11	1.31	2.39	1.37	2.86	84.83%
1	.600		.201	-1.46	-0.63	0.30	1.62	1.43	3.04	90.58%
11	.467		.395	-1.56	-0.26	1.31	2.30	1.03	1.79	77.12%
3	.846		.042	-0.37	0.24	0.91	1.64	2.78	7.36	99.14%
2	.195		.166	-4.89	-1.43	1.07	4.58	0.45	0.37	38.87%

Note. Item – number of item in the questionnaire. CAU - Communication Apprehension and Understanding. CAC - Communication Apprehension and Confidence. β_1 to β_4 – item difficulty indices. a – item discrimination. I – item information in the range from -3 to 3 logits. % - percent of the information in the range from -3 to 3 logits.

Descriptive Statistics, Convergent and Divergent Validity

Scores obtained on these two dimensions - CAU ($Sk = .19$, $Ku = -.66$) and CAC ($Sk = -.06$, $Ku = -.04$) are normally distributed (Tabachnick & Fidell, 2013), as well as score obtained from general factor ($Sk = .05$, $Ku = -.21$). Further, these two dimensions are highly and positively correlated ($r = .70$). Internal consistency reliabilities for dimensions are high ($\alpha_{CAU} = .93$; $\alpha_{CAC} = .83$). Gender differences were obtained when score on CAC dimension was used ($t(289) = -3.61$, $p < .001$) suggesting that female participants have tendency to score higher ($AS = 29.10$, $SD = 7.84$) than male participants ($AS = 25.88$, $SD = 7.26$). The same pattern of differences were obtained for general factor ($t(289) = -2.50$, $p < .05$), suggesting that female participants have tendency to score higher ($AS = 61.84$, $SD = 19.10$) than male participants ($AS = 56.54$, $SD = 16.98$). Gender differences for CAU score were not statistically significant.

Relations between FLCAS and personality dimensions were tested using three separated regression analyses. Within each regression model, one of three FLCAS dimensions (CAU, CAC or G factor) was used as a criterion, while in all three models Big five personality dimensions were used as predictors. Results that were obtained are presented in table 3. All tested models reached statistical significance, explaining from 17.4% to 19.7% variance of the criterion. Within all three models, statistically significant predictors were Emotional stability and Intellect, in the negative direction. Conscientiousness was positively correlated with CAC and G factor, while this relation was insignificant in case of CAU. Extraversion and Agreeableness were not related to CAU, CAC or General factor of FLCAS.

Table 3. Results of the Regression Analyses for FLCAS Scores and Personality Dimensions

	CAU		CAC		General factor		VIF
	β	t-test	β	β	t-test	β	
Extraversion	.089	1.47	-.011	-.190	.054	.898	1.284
Conscientiousness	.073	1.27	.161	2.785*	.117	2.038*	1.181
Agreeableness	-.053	-0.88	.054	.892	-.012	-.203	1.273
Emotional stability	-.315	-5.62**	-.248	-4.432**	-.312	-5.647**	1.102
Intellect	-.278	-4.72**	-.326	-5.522**	-.321	-5.513**	1.223
R	.419		.417		.443		
R ²	.176		.174		.197		
F-test	12.34**		12.222**		14.187**		

Note. CAU - Communication Apprehension and Understanding. CAC - Communication Apprehension and Confidence. VIF – variance influence factor.

Discussion

Foreign language learning anxiety is defined by its authors as a concept related to self-perceptions, beliefs, feelings, and behaviors which are typical for anxious reactions, but in this case related to a specific context – classroom learning of foreign language and its unique features (Horwitz et al., 1986). The present study was conducted in order to validate an adaptation of Foreign Language Classroom Anxiety Scale (FLCAS) which represents a measure of foreign classroom language anxiety. To address this question, several factor models were tested and information regarding the internal validity of the scale was obtained. Among all tested models, the best fit was provided for the bifactor model which consists of one general factor (G) of foreign language anxiety and two specific factors (S) - Communication Apprehension and Understanding (CAU) and Communication Apprehension and Confidence (CAC). These S factors are named after two-factor structure provided by Park (2014). These two factors are comparable to those from Park's study (2014) but only in terms of items that are loading on them because of the different factor analytic strategy. Two factors in previous research are the result of exploratory factor analysis (EFA) with oblimin rotation method used. In the other hand, in this study, these two factors represent two uncorrelated factors extracted from the variance that left after G factor of language anxiety was extracted. Park (2014) suggested that FLCAS reflects the core component of Communication Apprehension to which other subordinary components were related. Interpretation proposed by Park in his study seems to be applicable in this case, too, with the advantage that the usage of the bifactor model is more appropriate for inferring such a statement. In our study, G factor can be interpreted as a core communication apprehension, while the existence of S factors suggest that specific individual differences may be also identified depending on the self-assessed ability to understand the foreign language and perceived level of confidence regarding language performance. Examining correlations between CAU and CAC on the one side, and Big Five personality traits on the other imply that CAU and CAC have satisfactory convergent and divergent validity.

Potential implications of the results from this study are related to research and practical usage of scores obtained from FLCAS. It seems that depending on research and practical questions that one is targeting, general test score and subscales test scores can be used offering different information (more general and more specific). This assumption needs to be further examined by testing the predictive and incremental validity of G and S factors identified in this study. Based on gender differences in Communication Apprehension and Understanding, future studies should always take into account gender as a potential moderator variable. Measurement invariance across gender should be tested, and if test norms are created, possibly they should be different across genders, too.

Analysis based on item response theory (IRT) provided additional support to the quality of Serbian translation of this scale. Most of the items seem to function adequately in terms of difficulty, information and discrimination, except seven items (4, 13 and 20 loaded on CAU; 2, 3, 9, 32 loaded on CAC) which should be removed in future revisions of this instrument. The potential explanation of poor psychometric properties of these items is further examined. The content of 4 mentioned items from CAC contain words such as "worry", "tremble", "panic" referring to more extreme levels of anxiety that might be less frequent in the sample of students from common population. This probably initiates poor performance of these items. The similar explanation can be offered for the rest three items from CAU. Benefits of the removal of these seven items is in the fact that psychometric properties of the scale will be improved, and the scale will be shorter and less time-consuming for the participants.

The limitation of the research refers to the sample which includes only engineering students, so it is not clear to what extent the results can be generalized to the entire student population. Therefore, the future research should include students of natural and social sciences. In addition, the number of the respondents is not particularly large, although there are enough respondents in this research.

The second limitation concerns anxiety, as a personality dimension or a sub-clinical line that sometimes occurs in students. In other words, the limitation of the study is that language learning anxiety is not directly related to other anxiety scales, but only indirectly in connection with neuroticism which includes anxiety indicators, but only partially.

Conclusion

As it was mentioned at the beginning, this study has several great contributions to the topic of foreign language anxiety. This is the first study with the focus on Serbian translation of FLCAS, and the first study, in general, focused on conceptual aspects of foreign language anxiety using sophisticated statistical procedures (e.g. IRT, testing of bifactor models) that were used in order to address research questions. Except that the results of this study provide important information regarding factor structure of the scale and functioning of every single item from the IRT perspective, they open some new questions which seem to be worth of further investigation. Besides proposal for gender differences to be necessarily considered in future studies, and testing of the predictive and incremental validity of factors from demonstrated bifactor model, cross-cultural validation should be tested. It seems that cultural context plays an important role in the expression of foreign language anxiety and here maybe lies the reason for inconsistent results of previous factor analytic studies related to FLCAS.

The practical implication of this paper is that the scale is adequate for application, and thus, it leaves the possibility of screening students who experience language learning anxiety, so this scale could be used to identify these students and make them aware of the fact that the foreign language is unlikely complicated and difficult to learn.

The theoretical implication of the study is that the latent space of anxiety of learning foreign language can be best described from two lower order factors and one G factor. This further means that researchers in the field should use this type of questioning, rather than four-factor solutions, and the like.

The guidelines for further research are to examine the incremental validity of the G factor in relation to the S factors, as well as to check the scale invariability in relation to age, and develop adaptations of the scale for primary and secondary schools.

References

- Aida, Y. (1994). Examination of Horwitz, Horwitz, and Cope's construct of foreign language anxiety: The case of students of Japanese. *The Modern Language Journal*, 78(2), 155-168. doi:10.1111/j.1540-4781.1994.tb02026.x
- 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. doi:10.5070/L26121650
- Asmali, M. (2017). Big five personality traits and test anxiety among English as a foreign language learners. *Igdir University Journal of Social Sciences*, 11, 1-21.
- Baghaei, P., Hohensinn, C., & Kubinger, K. D. (2014). Persian adaptation of foreign language reading anxiety scale: A psychometric analysis 1. *Psychological Reports*, 114(2), 315-325. doi:10.2466/08.11.PR0.114k16w1
- Bond, T. G., & Fox, C. M. (2007). *Applying the Rasch model: Fundamental measurement in the social sciences* (2nd ed.). Mahwah, NJ: Lawrence Erlbaum.
- Cao, Y. (2011). Comparison of two models of foreign language classroom anxiety scale. *Philippine ESL Journal*, 7(1), 73-93.
- Chen, T. Y., & Chang, G. (2004). The relationship between foreign language anxiety and learning difficulties. *Foreign Language Journal*, 37(2), 279-289. doi:10.1111/j.1944-9720.2004.tb02200.x
- Cheng, Y. S., Horwitz, E., & Schallert, L. (1999). Language anxiety: Differentiating writing and speaking components. *Language Learning*, 49(3), 417-446. doi:10.1111/0023-8333.00095
- Daly, J. (1991). Understanding communication apprehension: An introduction for language educators. In D. J. Young & E. K. Horwitz (Eds.), *Language anxiety: From theory and research to classroom implications* (pp. 3-13). Englewood Cliffs, NJ: Prentice-Hall.
- DeMars, C. (2010). *Item response theory*. New York: Oxford University Press.
- Elkahafi, H. (2005). Listening comprehension anxiety in the Arabic language classroom. *The Modern Language Journal*, 89(2), 206-220. doi:10.1111/j.1540-4781.2005.00275.x
- Gardner, R. C., Clément, R., Smythe, P. C., & Smythe, C. L. (1979). Attitudes and Motivation Test Battery (AMTB). *Revised Manual. Research Bulletin No. 15*. London, Ontario: University of Western Ontario.
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *The Modern Language Journal*, 70(2), 125-132. doi:10.2307/327317
- Hu, L., & Bentler, P. M. (1999). Cut-off criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55. doi:10.1080/10705519909540118
- Huang, H. (2008). *University EFL students' and their teachers' preferences for in-class activities and their relationships to the students' foreign language anxiety* (Unpublished master's thesis). Providence University, Taiwan.
- IBM Corp. Released. (2012). *IBM SPSS Statistics for Windows, Version 21.0*. Armonk, NY: IBM Corp.
- Khodadady, E., & Khajavy, G. H. (2013). Exploring the role of anxiety and motivation in foreign language achievement: A structural equation modeling approach. *Porta Linguarium*, 20, 269-286.
- Liu, M., & Jackson, J. (2008). An exploration of Chinese EFL learners' unwillingness to communicate and foreign language anxiety. *The Modern Language Journal*, 92(1), 71-86. doi:10.1111/j.1540-4781.2008.00687.x
- MacIntyre, P. D. (1999). Language anxiety: A review of the research for language teachers. In D. J. Young (Ed.), *Affect in foreign language and second language learning: A practical guide to creating a low-anxiety classroom atmosphere* (pp. 24-45). New York: McGraw-Hill.
- MacIntyre, P. D., & Gardner, R. C. (1989). Anxiety and second language learning: Toward a theoretical clarification. *Language Learning*, 39(2), 251-275. doi:10.1111/j.1467-1770.1989.tb00423.x

- MacIntyre, P. D., & Gardner, R. C. (1991). Language anxiety: Its relationship to other anxieties and to processing in native and second languages. *Language Learning*, 41(4), 513-534. doi:10.1111/j.1467-1770.1991.tb00677.x
- Matsuda, S., & Gobel, P. (2004). Anxiety and predictors of performance in the foreign language classroom. *System*, 32, 21-36. doi:10.1016/j.system.2003.08.002
- McCroskey, J. C. (1970). Measures of communication-bound anxiety. *Speech Monographs*, 37(4), 269-277. doi:10.1080/03637757009375677
- Morizot, J., Ainsworth, A. T., & Reise, S. P. (2007). Toward modern psychometrics. In R. W. Robins, C. R. Fraley, & R. F. Krueger (Eds.), *Handbook of research methods in personality psychology* (pp. 407-423). New York: Guilford Press.
- Oxford, R. L. (1999). Anxiety and the language learner: New insights. In J. Arnold (Ed.), *Affect in language learning* (pp. 58-67). Cambridge: Cambridge University Press.
- 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. doi:10.5964/ejop.v9i3.611
- Paredes, P. F. P., & Muller-Alouf, H. (2000). A Spanish version of the foreign language classroom anxiety scale: Revisiting Aida's factor analysis. *Revista Española de Lingüística Aplicada*, 14(1), 337-352.
- Park, G. P. (2012). Investigation into the constructs of the FLCAS. *English Teaching*, 67(2), 207-220.
- Park, G. P. (2014). Factor analysis of the foreign language classroom anxiety scale in Korean learners of English as a foreign language. *Psychological Reports*, 115(1), 261-275. doi:10.2466/28.11.pr0.115c10z2
- Price, M. L. (1991). The subjective experience of foreign language anxiety: Interview with highly anxious students. In E. K. Horwitz & D. J. Young (Eds.), *Language anxiety: From theory and research to classroom implications* (pp. 101-108). Englewood Cliffs, NJ: Prentice Hall.
- R Core Team. (2016). *R: A language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Retrieved from <https://www.R-project.org/>
- Rizopoulos, D. (2006). ltm: An R package for latent variable modelling and item response theory analyses. *Journal of Statistical Software*, 17(5), 1-25. doi:10.18637/jss.v017.i05. <http://www.jstatsoft.org/v17/i05/> adresinden erişildi.
- Rossee, Y. (2012). lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, 48(2), 1-36. doi:10.18637/jss.v048.i02. Retrieved from <http://www.jstatsoft.org/v48/i02/>
- Samejima, F. (1969). *Estimation of latent trait ability using a response pattern of graded scores* (Psychometrika Monograph Suppl. No. 17). Richmond, VA: Psychometric Society. Retrieved from <http://www.psychometrika.org/journal/online/MN17.pdf>
- Sarason, I. G. (1984). Stress, anxiety, and cognitive interference: reactions to tests. *Journal of Personality and Social Psychology*, 46(4), 929-938. doi:10.1037/0022-3514.46.4.929
- Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson.
- Tobias, S. (1986). Anxiety and cognitive processing of instruction. In R. Schwarzer (Ed.), *Self-related cognition in anxiety and motivation*. Hillsdale, NJ: Lawrence Erlbaum.
- Tóth, Z. (2008). A foreign language anxiety scale for Hungarian learners of English. *Working Papers in Language Pedagogy*, 2, 55-77.
- Tsai, Y., & Li, Y. (2012). Test anxiety and foreign language reading anxiety in a reading proficiency test. *Journal of the Social Sciences*, 8(1), 95-103. doi:10.3844/jssp.2012.95.103
- Young, D. (1990). An investigation of students' perspectives on anxiety and speaking. *Foreign Language Annals*, 23(6), 539-553. doi:10.1111/j.1944-9720.1990.tb00424.x
- Zhao, N. (2007). Study of high school students' English learning anxiety. *Asian EFL Journal*, 9(3), 22-34.