

Validating the Short-Form Foreign Language Classroom Anxiety Scale (S-FLCAS)

This is a pre-print of the article:

Botes, E., van der Westhuizen, L., Dewaele, J. M., MacIntyre, P., & Greiff, S. (2022). Validating the Short-form Foreign Language Classroom Anxiety Scale. Applied Linguistics. <https://doi.org/10.1093/applin/amac018>

Please cite the published article.

Authors:

1. Dr. Elouise Botes (contact person)
University of Luxembourg
2 Avenue de l'Université
4365
Esch-sur-Alzette
Luxembourg
2. Lindie van der Westhuizen (University of Luxembourg)
3. Prof. Jean-Marc Dewaele (Birkbeck College, University of London)
4. Prof. Peter MacIntyre (Cape Breton University)
5. Prof. Samuel Greiff (University of Luxembourg)

Abstract

Foreign language classroom anxiety (FLCA) is a popular construct in applied linguistics research, traditionally measured with the 33-item Foreign Language Classroom Anxiety Scale (FLCAS). However, recent studies have started utilising the 8-item Short-Form Foreign Language Anxiety Scale (S-FLCAS). There is therefore a need, which this study addressed in five sequential steps, to validate the S-FLCAS in order to ensure the validity and reliability of the scale. A sample of $n = 370$ foreign language learners was utilised in the validation efforts, which included exploratory and confirmatory factor analyses, the establishment of convergent and discriminant validity, and invariance testing. The S-FLCAS was found to have a unidimensional structure with the 8 items loading on a single latent variable. Evidence was provided of the internal consistency and the convergent and discriminate validity of the S-FLCAS. In addition, the measure was found to be fully invariant across age, gender, educational levels, and L1 groups. It is therefore with some considerable confidence that we can recommend the future use of the S-FLCAS in peer-reviewed research.

Validating the Short-Form Foreign Language Classroom Anxiety Scale (S-FLCAS)

Anxiety has been found to negatively impact learning across numerous contexts, including mathematical learning (Hembree, 1990), science learning (Mallow, 2006), and foreign language (FL) learning (Dewaele & MacIntyre, 2014). In the case of FL learning, a domain-specific form of anxiety has been defined, namely, Foreign Language Classroom Anxiety (FLCA). In their now seminal paper, Horwitz et al. (1986) summarised FLCA as ‘a distinct complex of self-perceptions, beliefs, feelings, and behaviours related to classroom language learning arising from the uniqueness of the language learning process’ (p. 128). The introduction of the construct of FLCA also contained an accompanying 33-item measure, appropriately titled the Foreign Language Classroom Anxiety Scale (FLCAS). The measure has been extensively investigated with regard to its validity and reliability over the last three decades (see Horwitz, 1986; Caikang, 2003; Park, 2014) and has widely been accepted as a valid measure of the construct of FL learning anxiety.

The construct and its 33-item measure have remained staples in FL learning research on individual differences since its initial publication. As a consequence of its ubiquity among learners, FLCA has gained popularity as a variable in numerous studies and broad nomological network of variables (see MacIntyre, 2017). For example, FLCA has been linked to proficiency in FL learning (Arnaiz & Guillén, 2012), the willingness to communicate in the target language (Rastegar & Karami, 2015), motivation to learn the target language (Alrabai & Moskovsky, 2016), as well as interpersonal variables, such as the personality traits of neuroticism and perfectionism (Ghorbandordinejad & Nasab, 2013; Liu & Zhang, 2008), emotional intelligence (Shao, Ji & Yu, 2013), and self-esteem (Liu & Zhang, 2008). In addition, research examining FLCA with its accompanying 33-item FLCAS has taken place in diverse contexts, with studies examining gender differences in FL learning (e.g., Al-Saraj, 2014; Ezzi, 2012), studies examining adolescent FL

learners (e.g., Dordinejad & Ahmadabad, 2014) and adult FL learners (e.g., Dewaele, 2013), as well as studies examining a diverse set of target languages, including French (Gardner, Tremblay & Masgoret, 1997), Arabic (Elkhafaifi, 2005), and Chinese (Zhao & Whitchurch, 2011). The construct of FLCA and the 33-item FLCAS have been found to function across a variety of research contexts and are firmly entrenched in the research lexicon of FL learning studies.

Recent developments in the field language learning (Dewaele, 2012; Prior, 2019) have led to even greater research interest in the complex role of emotions, including FLCA and its measurement. FLCA has been studied alongside other emotions, such as Foreign Language Enjoyment (FLE; see Botes, Dewaele & Greiff, 2020a; Dewaele, 2019) and Foreign Language Boredom (FLB; see Kruk, 2021; Li, Dewaele & Hu, 2020). The paper which introduced FLE to the research lexicon measured FLCA with an eight-item shortened version of the FLCAS (see Dewaele & MacIntyre, 2014), which we refer to as the Short-Form Foreign Language Classroom Anxiety Scale (S-FLCAS) in the present paper.¹ The S-FLCAS was first developed in an appendix to MacIntyre's (1992) doctoral dissertation. However, the eight-item measure has rarely been used in lieu of the 33-item FLCAS in peer-reviewed articles, that is, until the article by Dewaele and MacIntyre (2014) which has increased research attention for the S-FLCAS. Indeed, numerous recent research publications have favoured the eight-item S-FLCAS over the longer, original 33-item measure (e.g., Botes et al., 2020; Dewaele et al., 2019, Moskowitz & Dewaele, 2020).

Although the use of short-form measures such as the S-FLCAS has advantages, the lack of a full validation study of the scale poses a significant research risk. Short-form measures have the crucial advantage of reducing administration time and thus allowing researchers to include a larger

¹ It should be noted that the eight-item scale was not given a specific name to differentiate it from the original 33-item FLCAS in either MacIntyre (1992) or Dewaele and MacIntyre (2014). For the sake of clarity, we decided to call the eight-item measure the Short-Form Foreign Language Classroom Anxiety Scale in this study.

number of measures in an assessment battery (Heene et al., 2014). This is a substantial advantage given the rise in the complexity of hypotheses and research questions in the social sciences, which creates the need for researchers to measure an ever-increasing number of variables as efficiently as possible in a single study (Ziegler et al., 2014). As such, it is unsurprising that the S-FLCAS has become increasingly popular since its 2014 revival, as the eight items can easily be incorporated into an assessment battery. However, unlike the 33-item FLCAS, no independent validation of the S-FLCAS has been carried out. Researchers therefore run the risk of utilising a measure that has not been consistently demonstrated to be valid and reliable or as usable across various diverse contexts with different gender, age, education, and target language groups.

The aim of this study is therefore to fill the gap in research and provide an independent validation of the S-FLCAS in order to determine the validity and reliability of the measure. Validation efforts will include invariance testing in order to ensure that items do not function differently across different age, gender, education level, and target language groups. In addition, an overview of the literature on the use, development, and current evidence of the validity and reliability of the FLCAS and S-FLCAS will be provided.

Foreign Language Classroom Anxiety Scale

The 33-item FLCAS was designed with the aim of providing a measure that captures the unique, situation-specific anxiety that arises in the FL classroom (Horwitz et al., 1986). Indeed, before the FLCAS, researchers in FL learning used a variety of general anxiety measures (see Chastain, 1975; Kleinmann, 1977) which led to a perplexing pattern of results (Scovel, 1978). Scovel (1978) attributed the contradictory results of his review – a mix of positive, negative, and near-zero correlations with language achievement measure – to the variety of anxiety measures that had been used, as anxiety was not conceptualised in a consistent manner across all the studies.

It is from this confounding phase in the literature (MacIntyre, 2017) that Horwitz et al. (1986) set out to design a measure that targeted anxiety as it is experienced in the FL classroom.

The introduction and popular use of the FLCAS has validated Scovel's (1978) opinion that the use of a consistent measure of language anxiety would lead to more consistent patterns of results. Recent meta-analyses have established clear, negative relationships between FLCA and language learning achievement (Botes et al., 2020b; Teimouri et al., 2019; Zhang, 2019). The success of the FLCAS in conceptualising language anxiety may be attributed to the strong design of its items. Items were based on the theoretical building blocks of test anxiety, communication apprehension, and fear of negative evaluation in the context of FL learning. However, it should be noted that Horwitz (1986, 2017) has specified that although FLCA is conceptually related to these three constructs, it is a separate and unique variable. FLCA can therefore not be reduced to the sum of test anxiety, communication apprehension, and fear of negative evaluation in the context of FL learning. Instead, FLCA took its theoretical inspiration from these three constructs. Items therefore refer to feelings of anxiety, nervousness, and unease in the FL classroom, with items such as 'In language class, I get so nervous I forget things I know' and 'I often feel like not going to language class' (Horwitz et al., 1986, p. 129).

Another theoretical foundation underlying the items of the FLCAS is the focus on measuring FLCA as a form of situation-specific anxiety. Situation-specific anxiety can be differentiated from trait and state forms of anxiety. Trait anxiety can be compared to a personality trait, with an individual being consistently predisposed to anxious behaviours or thoughts across a wide variety of contexts (Egloff & Hock, 2001). In turn, state anxiety is a temporary experience of anxiety at a particular moment (MacIntyre & Gardner, 1991). Situation-specific anxiety, such as FLCA, is seen as a delimited kind of trait anxiety in that it exhibits individual differences in the

levels of anxiety but only in response to the specific stimulus of language learning. Whereas a person with high trait anxiety might be afraid of snakes, flying in a airplane, and speaking FL in the classroom, it is not necessary to assume that someone who usually is nervous about FL speaking also fears snakes and flying (see Horwitz, 2017). Therefore, FLCA it is an anxiety that re-occurs in a particular type of situation, the language classroom.

The psychometric attributes of the FLCAS have been investigated and reinvestigated numerous times since the scale was introduced. Horwitz (1986) published a validation study of the scale with evidence of its test-retest reliability ($r_{tt} = .83$) as well as its convergent and discriminant validity. In order to establish validity, Horwitz (1986) correlated the FLCAS with trait anxiety ($r = .29, p < .01$), test anxiety ($r = .53, p < .01$), fear of negative evaluation ($r = .36, p < .01$), and communication apprehension ($r = .28, p > .05$). In addition, FLCA was found to be related to—but independent from—trait anxiety, test anxiety, and fear of negative evaluation. Additional evidence of the validity of the FLCAS has included high internal consistencies ($\alpha > .90$; Aida, 1994; Elkhafaifi, 2005; Gocer, 2014) and response validity (Tóth, 2008). However, the construct validity of the FLCAS has been less consistent. FLCA was therefore established as a unique construct, a form of anxiety conceptualized in a domain-specific way and worth investigating for its effects on learners, its role in language use, and its relationship to language learning.

Table 1*Previously Conducted FLCAS Factor Analyses*

Publication	L1	Target Language	Proposed factor structure	Methods
Aida (1994)	English	Japanese	<ol style="list-style-type: none"> 1. Speech anxiety and fear of negative evaluation 2. Fear of failing the class 3. Comfortableness in speaking with Japanese people 4. Negative attitudes towards Japanese class 	Principal component analysis with varimax rotation
Cheng et al. (1999)	Chinese	English	<ol style="list-style-type: none"> 1. Low self-confidence in speaking English 2. General English classroom performance anxiety 	Principal component analysis with varimax rotation
Liu and Jackson (2008)	Chinese	English	<ol style="list-style-type: none"> 1. Fear of negative evaluation 2. Communication apprehension 3. Test anxiety 	Factor analysis with varimax rotation
Park (2014)	Korean	English	<ol style="list-style-type: none"> 1. Communication apprehension and understanding 2. Communication apprehension and confidence 	Maximum likelihood exploratory factor analysis with direct oblimin rotation

Tóth (2008)	Hungarian	English	<ol style="list-style-type: none">1. Communication apprehension2. Fear of inadequate performance in English classes	Principal component analysis with direct oblimin rotation
-------------	-----------	---------	--	---

The original publication of the measure did not delve into the underlying factor structure, nor was the factor structure broached in Horwitz's (1986) validation study. It is therefore not surprising that examinations of the factor structure of the FLCAS became a frequently studied and debated topic. Different factor structures have been found across four studies, summarised in Table 1. Even though Horwitz (2001, 2017) has repeatedly stated that the conceptual building blocks of FLCA do not necessarily translate into the underlying factors, Table 1 demonstrates that numerous authors have labelled their factors in accordance with these conceptual building blocks. The majority of studies examining the factor structure underlying FLCA (see Table 1) disregarded a number of items that did not load on any of the selected factors. Indeed, Cheng et al. (1999) disregarded 13 items, and Park (2014) disregarded 10 items that did not load on any selected factors. The variability found in the underlying construct of FLCA may be attributed to the varying contexts in which the sample data were collected, such as different target languages or the proficiency levels of the sample groups (Park & French, 2013). In addition, different statistical analyses of the sample data sets may have resulted in different factor structures, as estimation and rotation methods can impact the results of exploratory factor analyses (Field, 2013). In effect, the discarded items become conceptual orphans with uncertain relationships to the underlying factors. In the interests of parsimony, the possibility of discarding some of the items in the FLCAS also ought to be considered.

Confirming the construct validity of the FLCAS is further complicated by a lack of a clear theoretical foundation with respect to the number of factors underlying FLCA. Horwitz (2017) advised against attempts to uncover the theoretical building blocks of test anxiety, communication apprehension, and fear of negative evaluation as factors underlying FLCA. Instead, Horwitz (2017) advocated for more practically oriented research aimed at assisting and alleviating FLCA in FL

learners, stating ‘my point is that we don’t need to thoroughly identify the components of Language Anxiety or understand the interactions among them to help anxious learners’ (p. 38). We might take Horwitz’ argument to suggest that a one-factor solution is optimal for conceptualizing FLCA as a construct for measurement purposes. The issue of dimensionality is important for understanding what one is getting when a specific test is employed, and accurate interpretation requires accurate conceptualization and measurement (Flake & Fried, 2020). Indeed, practical experimental or intervention-based research cannot be demonstrated to be effective if the targeted variables are not measured in a valid and reliable manner—including the construct validity of the scales (Flake & Fried, 2020).

The literature on the FLCAS has therefore established a measure with clear response reliability and internal consistency, with consistent pattern of validity correlates (Botes et al., 2020b). However, problematic items and an indistinct factor structure pose validity issues for the future use of the scale. The lack of clear construct validity is especially problematic should researchers aim to utilise multivariate statistical analysis techniques, such as structural equation modelling, where the results of hypotheses depend on specifying clear factor structures (Barret, 2007).

One solution to the measurement issues of the FLCAS may therefore be the use of a unidimensional scale, the S-FLCAS, if clear evidence of its validity and reliability can be obtained. Therefore, in this study, our aim was to validate the S-FLCAS to confirm that it offers a valid and reliable measurement of FLCA for use in future cross-sectional and intervention-based research.

Short-Form Foreign Language Anxiety Scale

The aim of developing the S-FLCAS was to create a short-form FLCA scale that more closely resembles the 10-item French Class Anxiety Scale and the 10-item French Use Anxiety Scale that helped inspire Horwitz et al. (1986) to create the FLCAS (MacIntyre, 1992). Items were selected through item analysis techniques, with items that did not adversely affect the internal consistency of the scale removed one-by-one (MacIntyre, 1992). The number of items were whittled down from 33 to eight, one item at a time with continuous item analyses conducted every time an item was removed, finally resulting in the eight-item S-FLCAS that showed minimal loss of internal consistency.

MacIntyre (1992) investigated the validity of the S-FLCAS by examining the internal consistencies, predictive validity, and convergent validity of the short form. The internal consistencies of the full 33-item FLCAS and the eight-item S-FLCAS were markedly similar ($\alpha = .94$; $\alpha = .93$; MacIntyre, 1992). Dropping 25 items had a negligible effect on the Cronbach alpha coefficient. In addition, the correlations between the FLCAS and the S-FLCAS provided evidence of convergent validity ($r = .98, p < .01$). The FLCAS and S-FLCAS were also found to have similar correlational patterns with third variables, such as grades in language courses ($r = -.38, p < .01$; $r = -.33, p < .01$), achievement test scores ($r = -.48, p < .01$; $r = -.44, p < .01$), and self-proficiency ratings ($r = -.61, p < .01$; $r = -.57, p < .01$; MacIntyre, 1992), respectively. In each case, some reduction in correlation was observed, which would be expected with fewer items, a smaller range of possible scores, and less variability available for analysis. However, the observed reductions were small and did not change the substantive interpretation of the correlations. As such, the preliminary validation of the S-FLCAS yielded promising results and provided preliminary evidence of the validity and reliability of the S-FLCAS, but the analysis was incomplete.

Several additional validity and reliability considerations remain for the eight-item S-FLCAS, specifically regarding its construct validity, convergent validity, and divergent validity. Given the contention regarding the factor structure of the 33-item FLCAS (see Table 1), a factor analytical study of the S-FLCAS is needed. As far as the authors are aware, no study has conducted a factor analysis of only the eight items of the S-FLCAS. The S-FLCAS was included in a previously published factor analysis alongside the 21 items from the Foreign Language Enjoyment Scale in order to establish the independence of FLCA and FLE (see Dewaele & MacIntyre, 2016). The eight items loaded on a single factor alongside two enjoyment factors, indicating that a single factor might underlie the S-FLCAS (Dewaele & MacIntyre, 2016). However, exploratory and confirmatory analyses are needed to fully determine the construct validity of the S-FLCAS. In addition, the relation of the S-FLCAS to variables associated with the construct of FLCAS—such as communication apprehension, trait anxiety, and fear of negative evaluation—should be examined to determine convergent validity. In addition, as the original validation attempt occurred three decades ago and was unpublished (MacIntyre, 1992); it would be a prudent time to re-examine the validity of the S-FLCAS, especially given its recent rise in popularity.

The S-FLCAS has been regularly utilised since it was included in the Dewaele and MacIntyre (2014) study. Recent studies featuring the S-FLCAS have examined relationships between FLCA and FLE (Uzun, 2017), FLCA and need for cognition (Razezadeh & Zarrinabadi, 2020), FLCA and language proficiency (Dewaele & Alfawzan, 2018), FLCA and emotional intelligence (Resnik & Dewaele, 2020), and FLCA and teacher related variables (Hung, 2020), whereas another study explored gender differences in FLCA (Alazeni, 2020). The rapid rise in the popularity of the measure is most likely due to the convenience provided by short-form measures

(Heene et al., 2014). The eight-item scale can therefore be easily incorporated alongside numerous other measures in an assessment battery.

Thus, MacIntyre (1992) developed the S-FLCAS through the use of sound psychometric methods and investigated the preliminary validity of the measure, though the analysis stopped short of a complete validation of the short form and assessment use a new sample is required. Although the measure was found to possess predictive validity, there was no attempt to address the factor structure underlying the S-FLCAS, and evidence of its confirmatory and discriminatory validity in a new sample is needed to avoid capitalizing on chance correlation in the previous sample. In addition, considerable time has passed since the initial validation efforts in 1992. Given increasing interest in emotion in SLA, the use of the S-FLCAS might be expected to increase, especially if a full examination of the reliability and validity of the S-FLCAS supports its use as a valid and reliable measure.

Method

Participants

The sample consisted of $n = 370$ international adult foreign language learners. The mean age of the sample was 27.56 years ($SD = 10.25$), with 168 male and 202 female participants. The average number of self-reported languages known by participants was 3.91 ($SD = 1.79$). There were a total of 45 different languages being learnt by the sample, with the majority learning Dutch ($n = 50$), followed by French ($n = 40$), and German ($n = 36$). More information regarding the descriptive statistics and distributions of data for age, education level, and the L1 of the sample can be found in the Supplementary Materials.

The data were collected in 2019 via the online platform SoSci Survey, utilising snowball sampling. All ethical requirements for data collection and storage were met as stipulated by the University of [redacted for peer-review].

Instruments

The following instruments were used in the validation of the S-FLCAS.

Short-Form Foreign Language Classroom Anxiety Scale (S-FLCAS)

The S-FLCAS consists of eight items identified by MacIntyre (1992) and as stipulated by Dewaele and MacIntyre (2014). The scale is aimed at measuring the broad construct of anxiety specific to foreign language learning, with items such as ‘It embarrasses me to volunteer answers in my FL class’. Two of the eight items are reverse coded (‘I don’t worry about making mistakes in FL class’ and ‘I feel confident when I speak in FL class’). Items are rated on a 5-point Likert scale, ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The sample Cronbach alpha was $\alpha = .89$.

Penn State Worry Questionnaire – Abbreviated (PSWQ-A)

This is an abbreviated eight-item measure that was designed to provide a general indication of worry and anxiety in adults, with items such as ‘My worries overwhelm me’. The eight-item version of the PSWQ was developed to address goodness-of-fit concerns in the original 16-item full-length version of the questionnaire (Hopko et al., 2003). Items are measured on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The Cronbach alpha for the PSWQ was $\alpha = .94$ in this sample.

Brief Fear of Negative Evaluation Scale (BFNES)

This is an eight-item measure for examining a general social anxiety and fear of being negatively judged by others, with items such as ‘I am afraid that people find fault with me’ (Rodebaugh et al., 2004). Items are measured on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The sample Cronbach alpha was $\alpha = .96$.

Short-Form Foreign Language Enjoyment Scale (S-FLES)

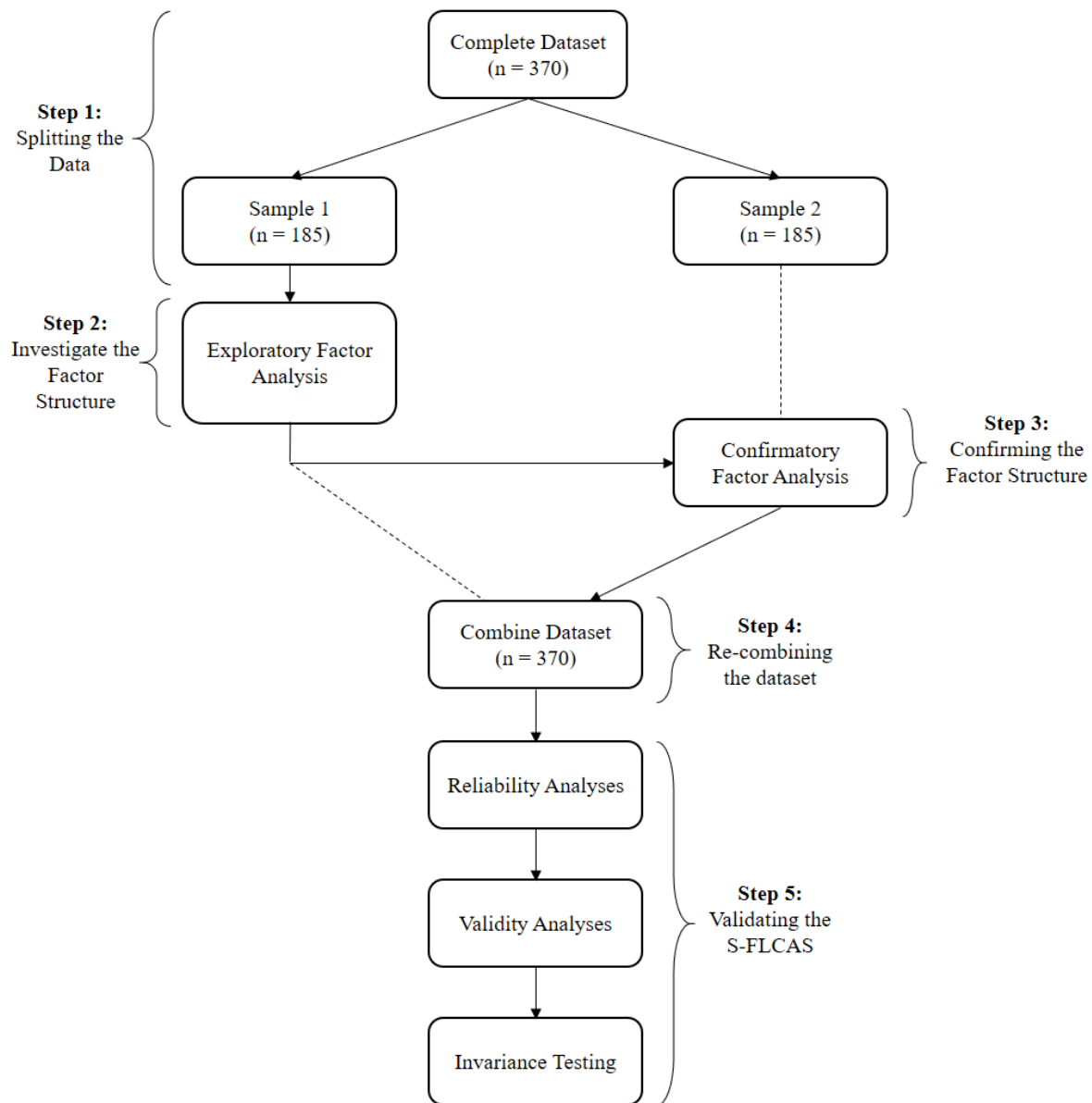
This is a nine-item broad measure of positive emotion experienced in the foreign language classroom (Botes et al., 2021a). This short-form version of the original 21-item scale developed by Dewaele and MacIntyre (2014) contains items such as ‘I enjoy FL learning’. Items are measured on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The sample Cronbach alpha was $\alpha = .89$.

Data Analysis

The validation of the S-FLCAS occurred in five sequential steps, detailed in Figure 1. The specific analyses and procedures involved in each step are described below.

Figure 1.

Method Flowchart.



Step 1: Splitting the Data

In accordance with the best practice guidelines (see Hagtvet & Sipos, 2016; Marsh et al., 2005), SPSS 25 was used to randomly split the data set in two. Subsequently, the two data sets were compared via t-tests to ensure that no statistically significant differences could be found between the two newly created samples. The examination of the factor structure in Step 2 utilised

the first sample that was created, and the confirmation of the measurement model in Step 3 utilised the second sample.

Step 2: Investigating the Factor Structure

The factor structure underlying the eight-item S-FLCAS was examined via maximum likelihood estimation with oblique (promax) rotation in JASP (JASP Team, 2020). Promax rotation was utilised as it would be theoretically expected that factors underlying FLCA would correlate, should a multidimensional factor structure emerge from the data (Toth, 2008). The first sample was utilised in this analysis. The factor structure was determined via the eigenvalue greater than 1 criterion and the scree plot.

Exploratory factor analysis with maximum likelihood estimation was chosen as the factor extraction method as it has been found to be less biased for uncovering the underlying factor structure and empirical fit than its rival principal axis factoring and generalised least square methods (de Winter & Dodou, 2012; Olssen et al., 2000). Furthermore, promax rotation was chosen as it is theoretically assumed that, should a multidimensional factor structure underlie FLCA, the factors in such a model will most likely be correlated (Field, 2013). Factor loadings were categorized as low ($< .4$), acceptable ($.4$ to $.6$), or high ($> .6$; Kline, 2014).

Step 3: Confirming the Factor Structure

The factor structure identified in Step 2 of the analysis was confirmed via a confirmatory factor analysis in R utilising the Lavaan package (Rosseel, 2012). The second sample generated in Step 1 was used for this analysis. Maximum likelihood estimation with robust standard errors were used to estimate all confirmatory factor analysis (CFA) models (including the invariance models). There were no missing values. The fit of the measurement model was evaluated via the following

fit indices: the Root Mean Square Error of Approximation (RMSEA), the Standardised Root Mean Square Residual (SRMR), the Comparative Fit Index (CFI), and the Tucker-Lewis Index (TLI). A model was deemed to have a close fit when CFI and TLI $> .95$, RMSEA $< .08$, and SRMR $< .08$ (Hu & Bentler, 1999). Factor loadings and cross-loadings were further investigated to determine the fit of the model.

It should be noted that the size of the sample utilised in this analysis can be considered somewhat small ($n = 185$). It would therefore be appropriate to examine statistical power before conducting the confirmatory factor analysis. As such, an extensive power analysis was conducted alongside the confirmatory factor analysis in Step 3.

Step 4: Recombining the Data Set

The two halves of the data set were recombined into a single data set of $n = 370$. The data set was recombined to provide sufficient statistical power to conduct invariance analyses. The full data set was utilised in the reliability, validity, and invariance analyses conducted in Step 5.

Step 5: Validity, Reliability, and Invariance Testing

Reliability was examined via internal consistency coefficients as measured by Cronbach's alpha and McDonald's omega.

Both the convergent and discriminant validity of the S-FLCAS were examined. Convergent and discriminant validity were established by contrasting the S-FLCAS with the PSWQ-A and the BFNES. The PSWQ-A provided a general measure of anxiety, and the BFNES provided a general indication of social anxiety. It is theoretically assumed that FLCA is related to, yet conceptually different from, general anxiety and social anxiety (Horwitz et al., 1986). Low to moderate correlations ($.15 \leq r \leq .35$) would therefore be expected between the S-FLCAS,

the PSWQ-A, and the BFNES and would provide evidence of convergent validity. In addition, the S-FLCAS should present as a distinct construct in an exploratory factor analysis alongside the PSWQ-A and the BFNES, thus providing evidence of discriminant validity. In addition, the moderate negative association between FLCA and FLE has been established in the literature ($r = -.30$; see Botes et al., 2021b), and as such, the S-FLES was further utilised to establish discriminant validity.

Furthermore, invariance testing was conducted on the S-FLCAS to demonstrate its generalisability across different subgroups in the population. Invariance was examined across gender, age groups, education level, and L1 groups. Invariance testing was conducted via JASP (JASP Team, 2020). A measure is said to be invariant when members of different groups (e.g., males and females) who have the same standing on the construct of interest receive the same observed score on the measure (Meredith, 1993). Thus, if measurement invariance is established, it can be assumed that the construct of interest is measured consistently across groups and that the properties of the scale are not affected by group differences. Using multigroup CFA, we tested measurement invariance by employing a series of increasingly restrictive invariance models across each of the aforementioned groups (Meredith, 1993; Millsap, 2011). First, we tested a configural invariance model in which all parameters were freely estimated across groups (Meredith, 1993). Configural invariance is used to test whether the factor structure is the same across groups. Next, a metric invariance model was tested in which the factor loadings were set to be invariant across groups (Meredith, 1993). Metric invariance implies that the different groups respond to the items in the same way. Lastly, scalar invariance was assessed by specifying a model in which both factor loadings and item intercepts were constrained to be equal across groups. Evidence of scalar invariance allows for comparisons of latent means across

groups (Meredith, 1993). In each consecutive step, the more restrictive model was compared with the previous, less restrictive model. The guidelines proposed by Cheung and Rensvold (2002) and Chen (2007) were used to evaluate invariance. Accordingly, invariance is supported when the $\Delta\text{CFI} \leq -.010$, $\Delta\text{RMSEA} \leq .015$, and $\Delta\text{SRMR} \leq .030$ (for metric invariance) or $\Delta\text{SRMR} \leq .015$ (for scalar invariance) when comparing the less restrictive with the more restrictive model. The change in CFI was used as the main criterion as RMSEA and SRMR tend to over-reject invariant models when the sample size is small (Chen, 2007). Invariance (and CFA model fit) cut-offs should, nevertheless, be considered rough guidelines instead of golden rules. Thus, in addition to the model fit indices, all available information, including the parameter estimates, statistical conformity, and theoretical adequacy of the models was considered to ultimately evaluate the CFA models (Marsh et al., 2004).

The sample was overwhelmingly composed of college-educated, young adults with an English L1 (see the Supplementary Materials). We limited the number of subgroups for age, education, and L1 to two groups because additional categorisations would have resulted in severely unbalanced groups (due to small sample sizes), which may have affected the results (Yoon & Lai, 2018). Therefore, this redistribution of groups was aimed at establishing groups of more or less equal sizes and of sufficient size to make invariance testing viable (e.g., Finch et al., 2018; Meade & Bauer, 2007). As such, participants were grouped into two major age groups, namely, young adults (18 to 25 years; $n = 208$) and adults (26+ years; $n = 162$). Two major education groups were also formed, namely, those with a secondary school education ($n = 117$) and those with a post-secondary school education ($n = 253$). Furthermore, due to the overwhelming number of English L1 participants in the sample, two L1 groupings were made,

namely, English L1 ($n = 212$) and non-English L1 ($n = 158$). Lastly, gender was examined via self-identified gender categories of female ($n = 202$) and male ($n = 168$).

Results

Step 1: Splitting the Data Set

The data set was randomly split into two separate samples. The two samples were examined via t-tests to determine that no statistically significant differences were present. The results of the t-tests can be found in the Supplementary Materials. Descriptive statistics for each sample can be found in Table 2.

Table 2

Descriptive Statistics for Samples 1 and 2

	Sample 1	Sample 2
Sample size	185	185
Gender	56.2% female	53% female
Age	$M = 27.34$ ($SD = 10.11$)	$M = 27.78$ ($SD = 10.4$)
Level of multilingualism	$M = 3.95$ ($SD = 1.69$)	$M = 4.01$ ($SD = 1.88$)
Average total FLCA	$M = 2.82$ ($SD = .92$)	$M = 2.96$ ($SD = .93$)

Step 2: Investigating the Factor Structure

An EFA with ML estimation and oblique (promax) rotation of the eight-item S-FLCAS was applied to Sample 1. The eigenvalue criterion (eigenvalue greater than 1) indicated that FLCA has a unidimensional factor structure as the first factor had an eigenvalue of 4.670, with all following factors indicating eigenvalues $< .850$ (see Table 3). In addition, the scree plot further demonstrated that a single factor underlies the S-FLCAS (see Figure 2) because an inflection point

after the first factor is visible, and no other inflection points were shown. It is therefore with some confidence that we can state that FLCA as captured by the S-FLCAS is a unidimensional construct.

Table 3

EFA of the S-FLCAS

Item	Factor 1
1	.775 ²
2	.611 ²
3	.779 ²
4*	.476 ¹
5*	.754 ¹
6	.737 ¹
7	.840 ¹
8	.781 ¹

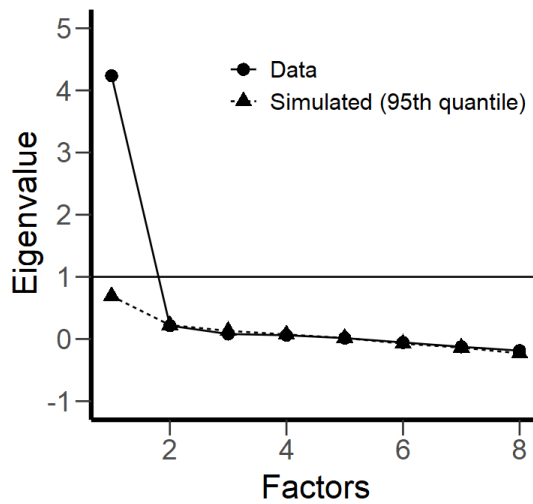
Note. *Reverse-scored items.

¹Acceptable loading (.4 to .6). ²High loading (> .6).

The factor loadings of the individual items were also acceptable (> .60; Kline, 2005), with the exception of Item 4: ‘I do not worry about making mistakes in FL class’. This item had a somewhat low factor loading, which could be attributed to using a negation (“not”) to create an item that is indicative of a *lack* of anxiety and therefore is reversely-scored (Conrad et al., 2004). However, on the whole, the EFA in Step 2 of the data analysis process yielded a clear unidimensional factor solution for FLCA.

Figure 2

Scree Plot of S-FLCAS

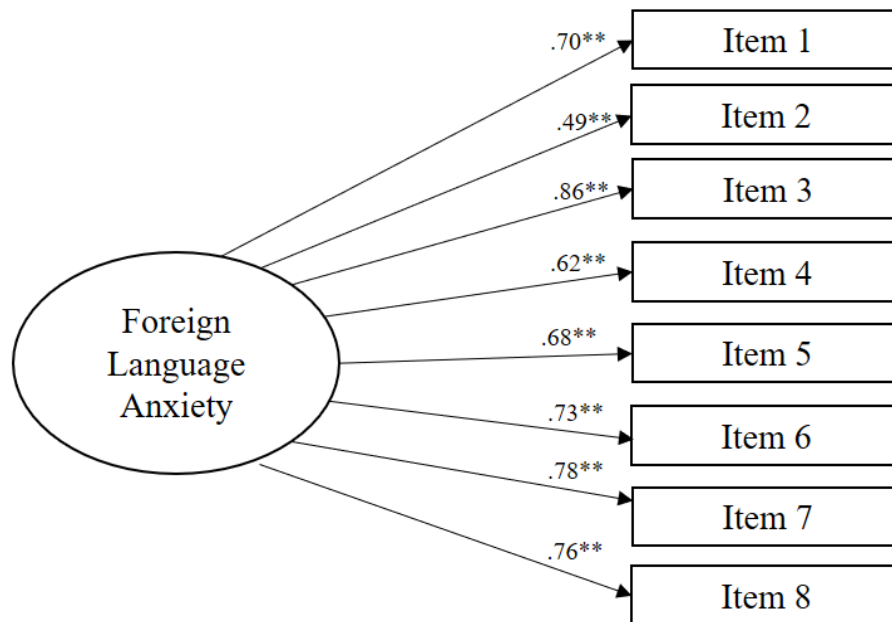


Step 3: Confirming the Factor Structure

As the factor structure of the S-FLCAS was found to be unidimensional in Step 2 of the data analysis, Step 3 proceeded with a confirmatory factor analysis of the eight-item one-factor S-FLCAS. The measurement model of the eight items of the S-FCLAS loading on a single latent variable was tested with ML estimation with robust standard errors in Sample 2 of the data (see Figure 3).

Figure 3

Model 1 of the S-FLCAS.



Note. ** $p < .001$.

Overall, the fit statistics indicated a good fit, $\chi^2(20) = 56.324, p < .001$. The CFI and TLI were both above the desired cut-off of $> .90$ and indicated a close fit (CFI = .984; TLI = .927; Kline, 2005). The SRMR further demonstrated good fit as it was below the cut-off of $< .05$ (SRMR = .048; Kenny, 2020). However, the RMSEA indicated a potential issue with fit (RMSEA = .099), as it was well above the desired cut-off of $< .08$ (Kenny, 2020).

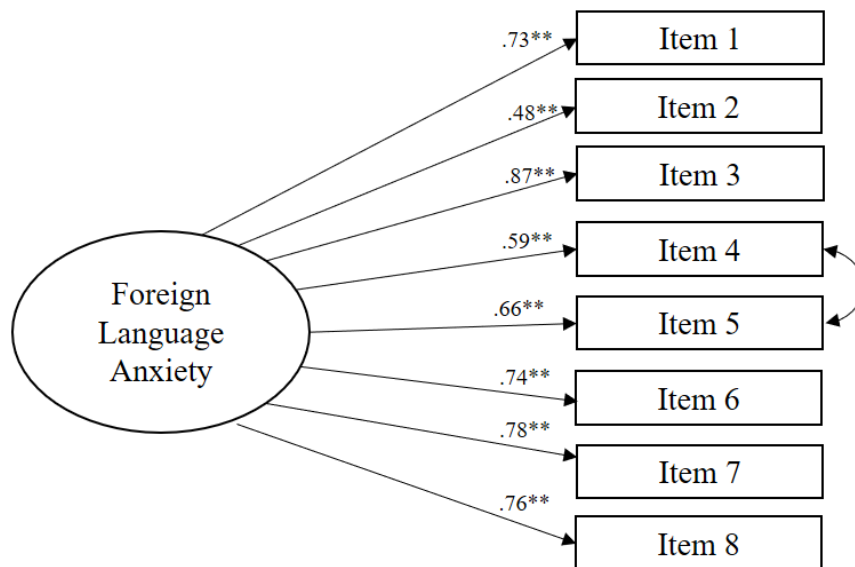
The factor loadings and modification indices provided some insight into the RMSEA results. Item 4 had a considerably higher standard error ($SE = .102$) in comparison with other items. Items 4 and 5 are reverse-scored, with item 4 in particular using the negative adverb of 'not' to create a negatively worded statement ('I don't worry about making mistakes in FL class'). Reverse-scored items are well-known to be associated with measurement difficulties, as reverse-scored items may lead to atypical responses (Carlson et al., 2011), which in turn impacts the models' fit statistics (Conrad et al., 2004) and may lead to Type II errors in model rejection

(Woods, 2006). The S-FLCAS contains two reverse-scored items (Items 4 and 5). The modification index in the confirmatory factor analysis suggested correlating Items 4 and 5 and thus ‘fixing’ the pathway between the two reverse-scored items ($MI = 15.921$).

Therefore, due to the known measurement difficulties caused by reverse-scored items and the results of the modification index, a second measurement model was tested with a correlation added between Items 4 and 5. The second model again utilised ML estimation with robust standard errors in Sample 2 (see Figure 4).

Figure 4.

Model 2 of the S-FLCAS.



Note. ** $p < .001$.

The second model demonstrated an improved, close-fitting model (see Table 4). In particular, the CFI and TLI values both increased to indicate a very close fit ($> .95$; Kline, 2005).

In addition, the SRMR further decreased to indicate a close fit ($< .05$; Kenny, 2020). However, the most significant improvement could be seen in the results of the RMSEA, which now indicated an adequate fit in the second measurement model ($< .08$; Kenny, 2020).

Table 4

Comparison of the Fit Statistics for Models 1 and 2

	Model 1	Model 2
χ^2 (df)	56.324 (20), $p < .001$	40.414 (19), $p < .001$
CFI	.948	.969
TLI	.927	.955
RMSEA	.099	.078
SRMR	.048	.040

Overall, the unidimensional model of the S-FLCAS provided a good fit to the data, with the caveat that if the individual items are used in SEM we would recommend correlating the reverse-scored Items 4 and 5 for the measurement model of the S-FLCAS in the future. As such, all validity and invariance analyses conducted in Step 5 of the data analysis included the symmetrical effect between Items 4 and 5.

Step 4: Recombining the Data Set

It should be noted that the two samples ($n = 185$; $n = 185$) utilised separately in Steps 2 and 3 of the data analyses were again recombined into a single data set ($n = 370$). This combined data set was used in the reliability, validity, and invariance testing in Step 5 of the data analysis.

Step 5: Validating the S-FLCAS

Reliability

The internal consistency of the eight-item S-FLCAS was acceptable, as measured by both Cronbach's alpha ($\alpha = .891$) and McDonald's omega ($\omega = .893$).

Validity

In order to establish validity, we compared the total score from the S-FLCAS with the scores from the PSWQ-A, the BFNES, and the S-FLES (see Table 5).

Table 5

Correlation Table

	1	2	3	4
1. S-FLCAS	-	.322**	.008	-.264**
2. PSWQ-A		-	.090	-.050
3. BFNES			-	-.014
4. S-FLES				-

** $p < .001$.

The S-FLCAS was found to be moderately positively correlated with general anxiety as measured by the PSWQ-A ($r = .322$, $p < .001$). Language anxiety is theoretically expected to be associated with, yet a distinct construct from, general anxiety (Horwitz et al., 1986). As such, the finding of a moderate positive correlation provided evidence of both convergent and discriminant validity. Furthermore, an EFA of the items in both the S-FLCAS and the PSWQ-A yielded two distinct factors, one language anxiety factor and one general anxiety factor (see the Supplementary Materials), further providing evidence of discriminant validity.

Contrary to our expectations, no statistically significant correlation was found between the S-FLCAS and the BFNES. There was therefore no discernible relationship between language anxiety and fear of negative evaluation in this data set. To this end, this result was somewhat

unexpected as the fear of negative evaluation has been theorised to be one of the building blocks of FLCA (Horwitz et al., 1986). Indeed, several studies using the 33-item FLCAS have found moderate positive correlations between FLCA and fear of negative evaluation (Aydin, 2008; Shabani, 2012; Tzoannopoulou, 2016), with Šafranĳ and Zivlak (2018) finding that fear of negative evaluation positively predicted FLCA ($\beta = 0.13, p < .05$). The statistically insignificant correlation found between the S-FLCAS and BFNES reported here suggests that this component of the original FLCAS is not well reflected in the short form of the scale. Researchers with a particular interest in anxiety related to social comparison processes and/or teacher evaluation might prefer to use the original long form of the FLCAS scale.

Lastly, the moderate negative correlation found between the S-FLCAS and the S-FLES was as expected. A recent review of the literature found an overall moderate correlation between FL learning anxiety and FL learning enjoyment ($r = -.30, k = 25, N = 13,421$; Botes et al., 2021b). The S-FLCAS therefore followed the trend in the literature and provided further confirmation that FLA and FLE are two distinct constructs.

Invariance Testing

We tested invariance by means of multigroup CFA models across gender, age, education level, and L1, respectively. Although the configural, metric, and scalar models were run consecutively for each of the groups separately (i.e., first running all three models for gender, then all three models for age, etc.), we jointly report the results for gender, age, education level, and L1 for each level of invariance for the sake of brevity. First, we (separately) estimated the configural invariance models for gender, age, education level, and L1. All the configural models provided a good fit to the data (Table 6), suggesting that the overarching factor structure was equivalent across all the gender, age, education level, and L1 groups that we tested. Next, we tested for metric

invariance (separately) across gender, age, education level, and L1. When comparing the configural invariance models with the metric invariance models, the ΔCFI , ΔRMSEA , and ΔSRMR values fell within the recommended cut-offs (i.e., $\Delta\text{CFI} \leq -.010$, $\Delta\text{RMSEA} \leq .015$, $\Delta\text{SRMR} \leq .030$), supporting invariant factor loadings across gender, age, education level, and L1 groups (see Table 6). Lastly, we tested for scalar invariance (separately) across gender, age, education level, and L1. When comparing the less restrictive metric models with the more restrictive scalar models, the results suggested that the item intercepts could be assumed to be invariant across gender, age, education level, and L1 groups because the ΔCFI , ΔRMSEA , and ΔSRMR values did not exceed the recommended cut-offs.²

Overall, our invariance results suggest that foreign language classroom anxiety was measured similarly by the S-FLCAS across the different groups. The scale properties (i.e., factor structure, factor loadings, and item intercepts) were fully invariant across the specified groups, suggesting that participants from different genders, ages, educational levels, and L1s understood the S-FLCAS items in comparable ways. Our results attest to the generalisability of the S-FLCAS across different subgroups of a population and indicate that latent mean comparisons on the S-FLCAS across gender, age, education level, and L1 would be permissible, meaningful, and valid.

Table 6

Measurement Invariance Across Gender, Age, Educational Level, and L1

² To account for the unbalanced group sizes in our invariance testing, we conducted a simulation study using a subsampling approach developed by Yoon and Lai (2018). The method and results of these analyses are reported in the supplementary material. In sum, the results of the simulation study confirmed our conclusion that the S-FLCAS could be considered scalar invariant across gender, age, education level, and L1 groups.

Invariance model	χ^2	<i>df</i>	<i>p</i> -value	CFI	RMSEA	SRMR	Δ CFI	Δ RMSEA	Δ SRMR
Invariance across gender									
Configural	74.997	38	.001	.974	.073	.038			
Metric	80.560	45	.001	.975	.065	.048	.001	-.008	.010
Scalar	86.882	52	.002	.975	.060	.047	.000	-.005	-.001
Invariance across age									
Configural	63.605	38	.006	.982	.060	.031			
Metric	71.682	45	.007	.981	.057	.049	-.001	-.003	.018
Scalar	78.503	52	.010	.981	.052	.047	.000	-.005	-.002
Invariance across educational level									
Configural	76.585	38	.001	.973	.074	.034			
Metric	88.499	45	.001	.970	.072	.056	-.003	-.002	.022
Scalar	101.810	52	.001	.965	.072	.054	-.005	.000	-.002
Invariance across L1									
Configural	67.552	38	.002	.979	.065	.033			
Metric	71.781	45	.007	.981	.057	.042	.002	-.008	.009
Scalar	79.155	52	.009	.981	.053	.041	.000	-.004	-.001

Discussion

The S-FLCAS was found to be a valid and reliable measure after the five sequential validation steps were followed in this study.

The exploratory factor analysis uncovered a unidimensional structure underlying FLCA, with all eight items loading on a single latent variable. The unidimensional structure was further confirmed by the confirmatory factor analysis, although a minor measurement concern regarding the reverse-scored items (Items 4 and 5) were raised. The two reverse-scored items were therefore correlated as suggested by modification indices, and the second confirmatory factor analysis indicated a close fit (RMSEA = .078; CFI = .969). The confirmation of a

unidimensional FLCA construct as measured by the S-FLCAS is not an unexpected finding, as Dewaele and MacIntyre's (2016) results already indicated this possibility.

The unidimensional structure has the advantage of simplicity, in that the scale's total score can easily be used in linear regression and correlational studies without compromising the underlying construct. In addition, when more advanced statistical techniques such as structural equation modelling are used with individual item scores, the specification of the proposed measurement model is straightforward with the recommended addition of fixing a path between the two reverse-scored items (specifically items 4 and 5, see Figure 4). The inclusion of two reverse-scored items might reduce response set bias (Borgers et al., 2004) but it does pose some measurement limitations, as reverse-scoring has been found to adversely affect the model fit (Conrad et al., 2004) and item responses (Carlson et al., 2012). As such, we do recommend that the correlational path be included between the reverse-scored Items 4 and 5 when the S-FLCAS measurement model is used in the future. In addition, fruitful future research may be carried out by examining the value provided by the inclusion of the two negatively worded items and exploring the option of creating an only-positively worded FLCAS. Nevertheless, the clear unidimensional solution found to underlie the S-FLCAS can be considered a boon.

However, the acceptance of a unidimensional S-FLCAS does create some measurement contention. The S-FLCAS is meant to capture the construct of FLCA in the same manner as the full 33-item FLCAS, yet—as far as we are aware—no unidimensional solution of the FLCAS has ever been proposed. However, given that no factor structure has consistently and repeatedly been identified as underlying the FLCAS (see Table 1), and available solutions discard a substantial number of items, we suggest that the S-FLCAS and the FLCAS both measure FLCA, especially because MacIntyre (1992) found that the total scores from the two scales were nearly

identical ($r = .98, p < .01$). Indeed, the numerous factor analyses of the full 33-item FLCAS may have hinted at a unidimensional structure, with previous factor analytic studies often finding a first factor that explained the majority of the variance in the data. For example, Aida (1994) and Tóth (2008) both found multidimensional solutions, but both authors' first factors explained the considerable majority of the variance. It is likely the case that variance accounted for by the first factor is largely responsible for the pattern of validity correlations observed over the 30+ years of using the FLCAS in research. In addition, should a multidimensional structure of the full 33-item FLCAS be the preferred solution, the possibility of a hierarchical structure with a global FLCA factor cannot be discounted. As far as we are aware, a hierarchical solution of the full 33-item FLCAS has not been tested or considered in previous studies. As such, the S-FLCAS was found to have a clear unidimensional structure, which we argue does not hinder the S-FLCAS from being considered a valid short-form of the FLCAS and does indeed measure anxiety experiences the same conceptualisation of the construct of FLCA.

Beyond the factor structure, additional validation results were promising. The S-FLCAS was found to have an acceptable internal consistency ($\alpha = .891$; $\omega = .893$), although as expected it was somewhat lower than the internal consistency found by MacIntyre (1992; $\alpha = .93$). The statistically significant positive correlation between the S-FLCAS and the PSWQ ($r = .322, p < .001$) indicated both convergent and divergent validity. The result confirmed the theoretically expected relationship between FLCA and trait anxiety (Horwitz et al., 1986), as measured by the S-FLCAS and the PSWQ, but the moderate correlation further indicated that although the two constructs are related, they are independent. Furthermore, the statistically significant negative correlation found between the S-FLCAS and the S-FLES ($r = -.264, p < .001$) further provides evidence of convergent and divergent validity as the two variables of FLCA and FLE have

consistently been found to be moderately negatively correlated (see Botes et al., 2020c). The only fly in the proverbial ointment of the validation attempt is the nonsignificant result found between FLCA and fear of negative evaluation as measured by the S-FLCAS and the BFNES, respectively. A relationship between FLCA and fear of negative evaluation has been established in the literature (see Aydin, 2008; Shabani, 2012; Tzoannopoulou, 2016). The null result we found might be explained by the type of items that were removed in the original reduction of 33 items in the FLCAS to the eight items in the S-FLCAS. In fact, given the substantial reduction in items, many of the items referring to the social evaluation of the FL classroom were cut. The 33-items FLCAS included four items that mention social comparison to other students in the class ('I keep thinking that the other students are better at languages than I am,' 'I always feel that the other students speak the foreign language better than I do,' 'I am very self conscious about speaking the foreign language in front of other students,' and 'I am afraid that the other students will laugh at me when I speak the foreign language'). In addition, five of the original 33 items referred to teacher ('It frightens me when I don't understand what the teacher is saying in the foreign language', 'I get upset when I don't understand what the teacher is correcting,' 'I'm afraid that my language teacher is ready to correct every mistake I make,' 'I get nervous when I don't understand every word the language teacher says,' and 'I get nervous when the language teacher asks questions which I haven't prepared in advance'). Reducing a scale by approximately 75%, from 33 to eight items, inevitably requires some sacrifice of detail in the measurement. However, given the results from MacIntyre (1992), the recent published studies using the S-FLCAS and the results in present study, the short form seems to capture efficiently individual difference in anxiety in the foreign language classroom.

Strong results emerged from invariance testing which yielded overwhelmingly positive results supporting the use of the S-FLCAS. An equivalent factor structure, factor loadings, and item intercepts were found across age, gender, educational level, as well as L1 groups. Thus, FLCA was measured similarly across groups, and the use of the S-FLCAS in comparing age, gender, educational level, and L1 groups would be a valid endeavour. The importance of the invariance results should not be understated as it provided the first statistical evidence of the fairness of the S-FLCAS because the items did not function differently across groups (Kline, 2013). This is an encouraging finding as previous research examining FLCA has spanned numerous cultural, educational, and language settings (see Teimouri et al., 2019). In addition, the result of full invariance across English and non-English L1 learners is an especially advantageous finding for the future use of the S-FLCAS in research. Researchers may therefore administer the S-FLCAS in English to non-English L1 FL learners with confidence, provided that the FL learners possess at least an intermediate proficiency in English.

The overall promising results on the validity and reliability of the S-FLCAS notwithstanding, the study and measure are not without limitations. Firstly, the sample size ($n = 370$) placed some constraints on the data analysis, as larger sample sizes are often recommended for both structural equation modelling and invariance testing (Kenny, 2020). The sample size especially affected the groupings in the invariance testing, with groups limited to two per category due to statistical power constraints. That said, the results of our simulation study using the Yoon and Lai (2018) subsampling method suggest that we can confidently conclude that our unbalanced sample sizes did not impact our conclusions concerning the invariance of the S-FLCAs across age, gender, educational level, and L1 groups. Secondly, the sample itself was also composed of a majority of highly educated young adults and L1 English speakers. These

skewed demographics further limited the groupings that were possible in the invariance analyses. Lastly, the S-FLCAS itself is also limited to a given context, namely, FL anxiety experienced in the FL classroom by adult or adolescent FL learners. The S-FLCAS might not be suitable for use with non-traditional or self-taught FL learners who did not learn in a classroom and is also not advised to be administered to young children.

The limitations posed by this study will hopefully be addressed in future research. Additional validation efforts of the S-FLCAS are needed, specifically confirming the unidimensional structure and the fixing of the pathway between the two negatively worded items. Additional studies need to confirm the proposed pathway between items 4 and 5 of the measurement model of the S-FLCAS in order to ensure that the proposed residual correlation is not merely the result of sampling error. In addition, future validation efforts ought to be conducted in research contexts that differ from the one used in this study. For example, the sample contained a majority of English L1 speakers. Thus, future research administering the S-FLCAS to a uniform, non-English L1 sample may provide additional insights into the construct validity of the measure. Invariance testing with more varied and larger samples is also recommended. In addition, validated translations of the S-FLCAS would be especially valuable.

It is our hope that future FLCA research will utilise the S-FLCAS and address some of the questions raised in this study. Indeed, the 33-item FLCAS has had three decades of psychometric and anecdotal evidence to recommend its use. Considerable future research is therefore needed to establish the S-FLCAS to the same extent. However, the ease of use, interpretation, validity, and reliability of the measure confirmed in this study as well as the general benefits that short-form measures provide (Heene et al., 2014) undoubtedly recommend the S-FLCAS for future use in peer-reviewed research.

Conclusion

The study aimed to validate the S-FLCAS. Although the scale was introduced in 1992, it was only due to recent developments in the field that there has been an exponential rise in the use of the S-FLCAS. As such, a need to ensure the validity and reliability of the S-FLCAS has emerged. The validation efforts uncovered a unidimensional factor structure, with all eight items loading on a single latent variable. Evidence of the internal consistency as well as the convergent and discriminant validity of the S-FLCAS was found. In addition, invariance testing confirmed that the scale properties of the S-FLCAS are fully invariant across age, gender, educational level, and L1 groups. On the whole, the psychometric evidence behind the S-FLCAS is overwhelmingly positive, and we recommend its future use in applied linguistics research.

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. <https://doi.org/10.1111/j.1540-4781.1994.tb02026.x>
- Alenezi, S. M. (2020). Foreign language enjoyment and anxiety among the Northern Borders University EFL students: Links to gender and majors. *Journal for Educational, Psychological, and Social Research*, 39(185), 1203-1233.
<https://dx.doi.org/10.21608/jsrep.2020.86081>
- Arabai, F., & Moskovsky, C. (2016). The Relationship between Learners' Affective Variables and Second Language Achievement. *Arab World English Journal*, 7(2), 77–103.
- Al-Saraj, T. M. (2014). Foreign language anxiety in female Arabs learning English: Case studies. *Innovation in Language Learning and Teaching*, 8(3), 257-278.
- Arnaiz, P., & Guillén, F. (2012). Foreign Language Anxiety in a Spanish University Setting: Interpersonal Differences. *Revista de Psicodidáctica*, 17(1), 5–26.
- Aydin, S. (2008). An Investigation on the Language Anxiety and Fear of Negative Evaluation among Turkish EFL Learners. *Asian EFL Journal*, 31, 421-444.
- Barrett, P. (2007). Structural equation modelling: Adjudging model fit. *Personality and Individual differences*, 42(5), 815-824.
- Borgers, N., Sikkels, D., & Hox, J. (2004). Response effects in surveys on children and adolescents: The effect of number of response options, negative wording, and neutral

- mid-point. *Quality & Quantity*, 38, 17–33. <https://doi-org/10.1023/B:QUQU.0000013236.29205.a6>
- Botes, E., Dewaele, J.-M., & Greiff, S. (2020a). The development and validation of the Short-form Foreign Language Enjoyment Scale (S-FLES). *Unpublished Manuscript*. <https://psyarxiv.com/984hb>
- Botes, E., Dewaele, J.-M., & Greiff, S. (2020b). The foreign language classroom anxiety scale and academic achievement: An overview of the prevailing literature and a meta-analysis. *Journal for the Psychology of Language Learning*, 2(1), 26-56.
- Botes, E., Dewaele, J.-M., & Greiff, S. (2020c). Taking stock: An overview of the literature and a preliminary meta-analysis of Foreign Language Enjoyment and other individual differences variables. *Unpublished Manuscript*. <https://psyarxiv.com/guaj5>
- Caikang, W. (2003). The adaptation and validation of the Foreign Language Classroom Anxiety Scale when applied to Chinese college students. *Psychological Science Shanghai*, 26(2), 281-284.
- Carlson, M., Wilcox, R., Chou, C. P., Chang, M., Yang, F., Blanchard, J., Marterella, A., Kuo, A., & Clark, F. (2011). Psychometric properties of reverse-scored items on the CES-D in a sample of ethnically diverse older adults. *Psychological Assessment*, 23(2), 558.
- Chastain, K. (1975). Affective and ability factors in second-language acquisition. *Language learning*, 25(1), 153-161.

- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14, 464–504.
<https://doi.org/10.1080/10705510701301834>
- Cheng, Y., Horwitz, E. K., & Schallert, D. L. (1999). Language Anxiety: Differentiating Writing and Speaking Components. *Language Learning*, 49(3), 417–446.
<https://doi.org/10.1111/0023-8333.00095>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 9, 233–255. https://doi.org/10.1207/S15328007SEM0902_5
- Conrad, K. J., Wright, B. D., McKnight, P., McFall, M., Fontana, A., & Rosenheck, R. (2004). Comparing traditional and Rasch analyses of the Mississippi PTSD Scale: revealing limitations of reverse-scored items. *Journal of Applied Measurement*, 5(1), 15-30.
- De Winter, J. C., & Dodou, D. (2012). Factor recovery by principal axis factoring and maximum likelihood factor analysis as a function of factor pattern and sample size. *Journal of Applied Statistics*, 39(4), 695-710.
- Dewaele, J.-M. (2013). The link between foreign language classroom anxiety and psychoticism, extraversion, and neuroticism among adult bi- and multilinguals. *The Modern Language Journal*, 97(3), 670–684. <https://doi.org/10.1111/j.1540-4781.2013.12036.x>
- Dewaele, J.-M. (2019). The effect of classroom emotions, attitudes toward English, and teacher behavior on willingness to communicate among English foreign language

learners. *Journal of Language and Social Psychology*, 38(4), 523-535.

<https://doi.org/10.1177/0261927X19864996>

Dewaele, J.-M., & Alfawzan, M. (2018). Does the effect of enjoyment outweigh that of anxiety in foreign language performance? *Studies in Second Language Learning and Teaching*, 8(1), 21–45.

Dewaele, J.-M., & MacIntyre, P. D. (2014). The two faces of Janus? Anxiety and enjoyment in the foreign language classroom. *Studies in Second Language Learning and Teaching*, IV(2), 237–274.

Dewaele, J.-M., & MacIntyre, P. D. (2016). Foreign language enjoyment and foreign language classroom anxiety. The right and left feet of FL learning? In P. MacIntyre, T. Gregersen, & S. Mercer (Eds.), *Positive Psychology in SLA* (pp. 215–236). Bristol, UK: Multilingual Matters. <http://www.multilingual-matters.com/display.asp?K=9781783095353>

Dewaele, J.-M., Chen, X., Padilla, A.M., & Lake, J. (2019). The flowering of positive psychology in foreign language teaching and acquisition research. *Frontiers in Psychology. Language Sciences*, 10, 2128. Doi: 10.3389/fpsyg.2019.02128.

Dordinejad, G. F., & Ahmadabad, M. R. (2014). Examination of the Relationship between Foreign Language Classroom Anxiety and English Achievement among male and female Iranian High School students. *International Journal of Language Learning and Applied Linguistics World*, 6(4), 446–460.

Egloff, B., & Hock, M. (2001). Interactive effects of state anxiety and trait anxiety on emotional Stroop interference. *Personality and Individual Differences*, 31(6), 875-882.

Elkhafaifi, H. (2005). Listening Comprehension and Anxiety in the Arabic Language Classroom.

The Modern Language Journal, 89(2), 206–220. <https://doi.org/10.1111/j.1540-4781.2005.00275.x>

Ezzi, N. A. A. (2012). The impact of gender on the foreign language anxiety of the Yemeni university students. *International Journal of Applied Linguistics & English Literature*, 1(2), 65-75.

Field, A. (2013). *Discovering statistics using SPSS* (5th edition). London, UK: Sage Publications.

Finch, H. W., French, B. F., & Hernández Finch, M. E. (2018). Comparison of Methods for Factor Invariance Testing of a 1-Factor Model With Small Samples and Skewed Latent Traits. *Frontiers in Psychology*, 9, 332. <https://doi.org/10.3389/fpsyg.2018.00332>

Flake JK, Fried EI. Measurement Schmeasurement: Questionable Measurement Practices and How to Avoid Them. *Advances in Methods and Practices in Psychological Science*. December 2020:456-465. doi:[10.1177/2515245920952393](https://doi.org/10.1177/2515245920952393)

Gardner, R. C., & MacIntyre, P. D. (1991). An Instrumental Motivation In Language Study: Who Says It Isn't Effective? *Studies in Second Language Acquisition*, 13(1), 57–72. <https://doi.org/10.1017/S0272263100009724>

Gardner, R. C., Tremblay, P. F., & Masgoret, A.-M. (1997). Towards a Full Model of Second Language Learning: An Empirical Investigation. *The Modern Language Journal*, 81(3), 344–362. <https://doi.org/10.1111/j.1540-4781.1997.tb05495.x>

GhorbanDordinejad, F., & Nasab, A. H. F. (2013). Examination of the relationship between perfectionism and English achievement as mediated by foreign language classroom

- anxiety. *Asia Pacific Education Review*, 14(4), 603–614. <https://doi.org/10.1007/s12564-013-9286-5>
- Gocer, A. (2014). The Relationship between Anxiety and Attitude of Students Learning Turkish as a Foreign Language and Their Achievement on Target Language. *Educational Research and Reviews*, 9(20), 879–884. <https://doi.org/10.5897/ERR2014.1784>.
- Hagtvet, K., & Sipos, K. (2016). Creating short forms for construct measures: The role of exchangeable forms. *Pedagogika*, 66(6), 689–713.
<https://doi.org/10.14712/23362189.2016.346>
- Heene, M., Bollmann, S., & Bühner, M. (2014). Much ado about nothing, or much to do about something? Effects of scale shortening on criterion validity and mean differences. *Journal of Individual Differences*, 35(4), 245.
- Hembree, R. (1990). The nature, effects, and relief of mathematics anxiety. *Journal for Research in Mathematics Education*, 21(1), 33-46.
- Hopko, D. R., Reas, D. L., Beck, J. G., Stanley, M. A., Wetherell, J. L., Novy, D. M., & Averill, P. M. (2003). Assessing worry in older adults: confirmatory factor analysis of the Penn State Worry Questionnaire and psychometric properties of an abbreviated model. *Psychological Assessment*, 15(2), 173.
- Horwitz, E. K. (1986). Preliminary Evidence for the Reliability and Validity of a Foreign Language Anxiety Scale. *TESOL Quarterly*, 20(3), 559–562.
<https://doi.org/10.2307/3586302>

- Horwitz, E. K. (2001). Language anxiety and achievement. *Annual Review of Applied Linguistics*, 21, 112–126. <https://doi.org/10.1017/S0267190501000071>
- Horwitz, E. K. (2017). On the Misreading of Horwitz, Horwitz and Cope (1986) and the Need to Balance Anxiety Research and the Experiences of Anxious Language Learners. In C. Gkonou, M. Daubney, & J.-M. Dewaele (Eds.), *New Insights into Language Anxiety* (pp. 31-48). Bristol, UK: Multilingual Matters.
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign Language Classroom Anxiety. *The Modern Language Journal*, 70(2), 125–132. <https://doi.org/10.1111/j.1540-4781.1986.tb05256.x>
- Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6, 1–55. <https://doi.org/10.1080/10705519909540118>
- Hung, M. T. (2020). *Teacher support and intrinsic motivation: The mediating roles of enjoyment, anxiety, and self-efficacy* (Doctoral dissertation). Available from <https://ir.library.msstate.edu/handle/11668/16660>
- JASP Team (2020). JASP (Version 0.13.1) [Computer software]. Amsterdam, The Netherlands: JASP.
- Kenny, D. A. (2020). Measuring Model Fit. Retrieved from <http://davidakenny.net/cm/fit.htm>
- Kleinmann, H. H. (1977). Avoidance behavior in adult second language acquisition 1. *Language learning*, 27(1), 93-107.

- Kline, P. (2014). *An easy guide to factor analysis*. Milton Park, UK: Taylor and Francis.
- Kline, R. B. (2013). Assessing statistical aspects of test fairness with structural equation modelling. *Educational Research and Evaluation*, 19(2-3), 204-222.
- Kline, T. (2005). *Psychological testing: A practical approach to design and evaluation*. Sage.
- Kruk, M. (2021). Empirical Investigations into WTC, Motivation, Language Anxiety and Boredom in Traditional and Digital Contexts. In *Investigating Dynamic Relationships Among Individual Difference Variables in Learning English as a Foreign Language in a Virtual World* (pp. 31-60). Springer, Cham.
- Li, C., Dewaele, J. M., & Hu, Y. (2020). Foreign language learning boredom: conceptualization and measurement. *Applied Linguistics Review*. <https://doi/10.1515/applirev-2020-0124/html>
- 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. <https://doi.org/10.1111/j.1540-4781.2008.00687.x>
- Liu, M., & Zhang, W. (2008). An exploration of Chinese EFL learners' foreign language anxiety, personality and self-esteem. *Journal of Applied Linguistics*, 5(2), 181–203. <http://dx.doi.org/10.1558/japl.v5i2.181>
- MacIntyre, P. D. (1992). *Anxiety And Language Learning From A Stages Of Processing Perspective* (Unpublished doctoral dissertation). The University of Western Ontario, London, ON, Canada. Retrieved from <https://ir.lib.uwo.ca/digitizedtheses/2155>

- MacIntyre, P. D. (2017). An overview of language anxiety research and trends in its development. *New Insights into Language Anxiety: Theory, Research and Educational Implications*, 11–30.
- Mallow, J. V. (2016). Science Anxiety: Research and action. In J. J. Mintzes (Ed.), *Handbook of College Science Teaching* (pp. 3–14). National Science Teachers Association.
<https://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.485.4166&rep=rep1&type=pdf>
- Marsh, H. W., Ellis, L. A., Parada, R. H., Richards, G., & Heubeck, B. G. (2005). A short version of the Self-Description Questionnaire II: Operationalizing criteria for short-form evaluation with new applications of confirmatory factor analyses. *Psychological Assessment*, 17(1), 81–102. <https://doi.org/10.1037/1040-3590.17.1.81>
- Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In search of golden rules: Comment on hypothesis-testing approaches to cutoff values for fit indexes and dangers in overgeneralizing Hu & Bentler's (1999). *Structural Equation Modeling: A Multidisciplinary Journal*, 11, 320 – 341. http://dx.doi.org/10.1207/s15328007sem1103_2
- Meade, A. W. & Bauer, D. J. (2007). Power and Precision in Confirmatory Factor Analytic Tests of Measurement Invariance, *Structural Equation Modeling: A Multidisciplinary Journal*, 14, 611-635, DOI: 10.1080/10705510701575461
- Meredith, W. (1993). Measurement invariance, factor analysis and factorial invariance. *Psychometrika*, 58, 525–543. <https://doi.org/10.1007/BF02294825>

- Millsap, R. E. (2011). *Statistical approaches to measurement invariance*. Routledge.
<https://doi.org/10.4324/9780203821961>
- Moskowitz, S., & Dewaele, J.-M. (2020). The role of intellectual humility in foreign language enjoyment and foreign language classroom anxiety. *Eurasian Journal of Applied Linguistics*, 6(3), 521-541.
- Olsson, U. H., Foss, T., Troye, S. V., & Howell, R. D. (2000). The performance of ML, GLS, and WLS estimation in structural equation modeling under conditions of misspecification and nonnormality. *Structural equation modeling*, 7(4), 557-595.
- 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, 261–275.
- Park, G.-P., & French, B. F. (2013). Gender differences in the Foreign Language Classroom Anxiety Scale. *System*, 41(2), 462–471. <https://doi.org/10.1016/j.system.2013.04.001>
- Rastegar, M., & Karami, M. (2015). On the Relationship between Foreign Language Classroom Anxiety, Willingness to Communicate and Scholastic Success among Iranian EFL Learners. *Theory and Practice in Language Studies*, 5(11), 2387–2394.
<https://doi.org/10.17507/tpls.0511.25>
- Resnik, P., & Dewaele, J.-M. (2020). Trait emotional intelligence, positive and negative emotions in first and foreign language classes: A mixed-methods approach. *System*, 94, 102324. <https://doi.org/10.1016/j.system.2020.102324>

- Rezazadeh, M., & Zarrinabadi, N. (2020). Examining need for closure and need for cognition as predictors of foreign language anxiety and enjoyment. *Journal of Multilingual and Multicultural Development*, 1-13. <https://doi.org/10.1080/01434632.2020.1798972>
- Rodebaugh, T. L., Woods, C. M., Thissen, D. M., Heimberg, R. G., Chambless, D. L., & Rapee, R. M. (2004). More information from fewer questions: the factor structure and item properties of the original and brief fear of negative evaluation scale. *Psychological assessment*, 16(2), 169.
- Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling and more. Version 0.5–12 (BETA). *Journal of statistical software*, 48(2), 1-36.
- Šafranĳ, J., & Zivlak, J. (2019). Effects of big five personality traits and fear of negative evaluation on foreign language anxiety. *Croatian Journal of Education*, 21(1), 275-306.
- Scovel, T. (1978). The Effect of Affect on Foreign Language Learning: A Review of the Anxiety Research. *Language Learning*, 28(1), 129–142. <https://doi.org/10.1111/j.1467-1770.1978.tb00309.x>
- Shabani, M. B. (2012). Levels and sources of Language Anxiety and Fear of Negative Evaluation among Iranian EFL learners. *Theory & Practice in Language Studies*, 2(11), 2378-2383.
- Shao, K., Yu, W., & Ji, Z. (2013). An exploration of Chinese EFL students' emotional intelligence and foreign language anxiety. *The Modern Language Journal*, 97(4), 917-929.

- Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition*, 1–25.
<https://doi.org/10.1017/S0272263118000311>
- Tóth, Z. (2008). A Foreign Language Anxiety Scale for Hungarian Learners of English. *Working Papers in Language Pedagogy*, 2, 55–78.
- Tzoannopoulou, M. (2016). Foreign Language Anxiety and Fear of Negative Evaluation in the Greek university classroom. *Selected Papers on Theoretical and Applied Linguistics*, 21, 823-838.
- Uzun, K. (2017). Compulsory English Courses in Higher Education: A Source of Angst or Thrill? *The Journal of Language Teaching and Learning*, 7(2), 1-20.
- Woods, C. M. (2006). Careless responding to reverse-worded items: Implications for confirmatory factor analysis. *Journal of Psychopathology and Behavioral Assessment*, 28(3), 186.
- Yoon, M., & Lai, M. H. C. (2018). Testing factorial invariance with unbalanced samples. *Structural Equation Modeling: A Multidisciplinary Journal*, 25, 201–213.
<https://doi.org/10.1080/10705511.2017.1387859>
- Zhang, X. (2019). Foreign language anxiety and foreign language performance: A meta-analysis. *The Modern Language Journal*, 103(4), 763-781. doi:10.1111/modl.12590

Zhao, A., & Whitchurch, A. (2011). Anxiety and its associated factors in college-level Chinese classrooms in the US. *Journal of the Chinese Language Teachers Association*, 46(3), 21-47.

Ziegler, M., Kemper, C. J., & Kruey, P. M. (2014). Short scales—five misunderstandings and ways to overcome them. *Journal of Individual Differences*, 35, 185-189.

Supplementary Materials

Short-form Foreign Language Classroom Anxiety Scale (S-FLCAS)

1. Even if I am well prepared for FL class, I feel anxious about it
2. I always feel that the other students speak the FL better than I do
3. I can feel my heart pounding when I'm going to be called on in FL class
4. I don't worry about making mistakes in FL class (reverse coded)
5. I feel confident when I speak in FL class (reverse coded)
6. I get nervous and confused when I am speaking in my FL class
7. I start to panic when I have to speak without preparation in FL class
8. It embarrasses me to volunteer answers in my FL class

Further Descriptive Statistics of the Data

Table 1

Age Distributions as Captured in the Data Collection.

Age	Frequency	Percentage
≤ 20 years	78	21.1%
21-25 years	130	35.1%
26-30 years	72	19.5%
31-35 years	38	10.3%
36-40 years	14	3.8%
41-45 years	14	3.8%
46-50 years	7	1.9%
51-55 years	5	1.4%
56-60 years	5	1.4%
≥ 60 years	7	1.9%
	370	100

Table 2

Education Level as Captured in the Data Collection

Highest Qualification	Frequency	Percentage
Secondary School	117	31.6%
Bachelor Degree	155	41.9%
Master's Degree	80	21.6%
PhD	18	4.9%
	370	100

Table 3*L1 Distributions as Captured in the Data Collection*

Language	Frequency	Percentage
Afrikaans	2	0.5%
Arabic	6	1.6%
Bulgarian	1	0.3%
Cantonese	1	0.3%
Catalan	3	0.8%
Chinese	2	0.5%
Czech	4	1.1%
Danish	1	0.3%
Dutch	15	4.1%
English	212	57.3%
Finnish	2	0.5%
French	13	3.5%
German	26	7.0%
Greek	2	0.5%
Hungarian	1	0.3%
Indonesian	3	0.8%
Italian	7	1.9%
Korean	1	0.3%
Lithuanian	2	0.5%
Luxembourgish	1	0.3%
Macedonian	1	0.3%
Malayalam	2	0.5%
Norwegian	3	0.8%
Polish	11	3.0%
Portuguese	9	2.4%
Romanian	1	0.3%

Russian	7	1.9%
Serbian	2	0.5%
Slovakian	1	0.3%
Slovenian	1	0.3%
Spanish	13	3.5%
Tamil	2	0.5%
Turkish	4	1.1%
Vietnamese	1	0.3%
370		100

T-test results of Sample 1 and 2

Statistically significant differences based on age, level of multilingualism, FLCA, and the PSWQ between Sample 1 and Sample 2 of the data were investigated via t-tests. No statistically significant differences between the two sample groups were found (see Table 4).

Table 4.

T-test results between Sample 1 and 2.

Variable	n	Mean	SD	t-value	p-value
Age				-.41	.682
Sample 1	185	27.34	10.11		
Sample 2	185	27.78	10.40		
Multilingualism				-.32	.750
Sample 1	185	3.94	1.69		
Sample 2	185	4.01	1.89		
FLCA				-1.42	.157
Sample 1	185	2.82	.92		
Sample 2	185	2.96	.93		
PSWQ				.154	.877
Sample 1	185	3.22	1.08		
Sample 2	185	3.20	1.01		

Pearson's Correlation Coefficient Matrix of All Items

Table 4

S-FLCAS Inter-Item Correlation Matrix

Age	1	2	3	4	5	6	7	8
Item 1	-	.403	.625	.384	.448	.538	.638	.537
Item 2		-	.395	.321	.441	.407	.485	.424
Item 3			-	.458	.552	.584	.672	.660
Item 4				-	.535	.364	.390	.396
Item 5					-	.528	.569	.570
Item 6						-	.630	.578
Item 7							-	.596
Item 8								-

Note. All correlations were significant at $p < .001$

EFA Results of the S-FLCAS and PSWQ-A Items

An EFA with ML estimation and oblique (promax) rotation of the eight-item S-FLCAS, the eight-item PSWQ-A, and the eight-item BFNES (see Table 5). The full dataset of $n = 370$ was used for the analysis. The items of each scale loaded onto unique, separate factors. Thus, the EFA clearly indicated that language anxiety as measured through the S-FLCAS, general anxiety as measured through the PSWQ-A, fear of negative evaluation as measured by the BFNES were three distinct constructs. The results further substantiated the discriminate validity of the S-FLCAS.

Table 5

EFA of S-FLCAS and PSWQ-A.

Item	Factor 1	Factor 2	Factor 3
FLCA Item 1	.704 ²		
FLCA Item 2	.580 ¹		
FLCA Item 3	.826 ²		
FLCA Item 4	.536 ¹		
FLCA Item 5	.738 ²		
FLCA Item 6	.733 ²		
FLCA Item 7	.808 ²		
FLCA Item 8	.765 ²		
PSWQ-A Item 1		.830 ²	
PSWQ-A Item 2		.849 ²	
PSWQ-A Item 3		.838 ²	
PSWQ-A Item 4		.768 ²	
PSWQ-A Item 5		.848 ²	
PSWQ-A Item 6		.759 ²	
PSWQ-A Item 7		.768 ²	
PSWQ-A Item 8		.906 ²	
BFNES Item 1			.873 ²
BFNES Item 2			.885 ²

BFNES Item 3	.881 ²
BFNES Item 4	.892 ²
BFNES Item 5	.848 ²
BFNES Item 6	.850 ²
BFNES Item 7	.897 ²
BFNES Item 8	.816 ²

Note. *Reverse-scored items.

¹Acceptable loading (.4 to .6). ²High loading (> .6).

Additional Invariance Testing using a *Monte Carlo Simulation*

The use of unequal groups sizes in invariance tests could result in invalid conclusions due to the larger group overdetermining the fit function in the multigroup confirmatory factor analyses (CFA) analysis (Yoon & Lai, 2018). To account for this, we complemented the invariance testing reported in the main manuscript with a Monte Carlo simulation technique proposed by Yoon and Lai (2018). This analyses was conducted in Mplus 8.3 (Muthén & Muthén, 1998-2019) with the R-script provided by Yoon and Lai (2018). For each multigroup CFA conducted (i.e., for age, gender, educational level, and L1), 100 multiple random samples were drawn from the largest of the two groups (i.e., the young adults group, the female group, the post-secondary school education group, the English L1 group), which corresponded to the size of the smallest of the two groups (i.e., the adults group; the male group, the secondary school education group, the non-English L1 group). Tests of configural, metric, and scalar measurement invariance were then performed 100 times, using the subsamples from the larger group together with the sample of smaller groups. The mean value of each fit statistic was thus calculated over 100 replications. The same testing sequence described in the main manuscript was used to test invariance across age, gender, educational level, and L1 with the Yoon and Lai (2018) simulation.

Similar to the results reported in the main manuscript, all the configural models provided a good fit to the data (Table 6), supporting the notion that the factor structure was equivalent across gender, age, education level, and L1. The results also seemed to support invariant factor loadings across all the groups (see Table 6), as the majority of the ΔCFI , $\Delta RMSEA$, and $\Delta SRMR$ values did not exceed the recommended cut-offs when comparing the configural invariance models with the metric invariance models. Likewise, the ΔCFI , $\Delta RMSEA$, and $\Delta SRMR$ values were within the acceptable range when comparing the metric models with the scalar models. Scalar invariance was

thus supported, indicating that the item intercepts can be assumed to be invariant across gender, age, education level, and L1 groups. Thus, based on the results of the simulation analyses, we can confidently conclude that our unbalanced sample sizes did not alter the conclusions on the measurement invariance of the S-FLCAS across gender, age, education level, and L1 groups.

Table 6

Measurement Invariance Across Gender, Age, Educational Level, and L1: Monte Carlo

Simulation

Invariance model	χ^2	df	p-value	Mean CFI	Mean RMSEA	Mean SRMR	Δ CFI	Δ RMSEA	Δ SRMR
Invariance across gender									
Configural	62.280	39	-	.976	.059	.043			
Metric	70.859	47	-	.975	.055	.057	-.001	-.004	.014
Scalar	75.423	52		.976	.051	.049	.000	-.004	-.008
Invariance across age									
Configural	49.524	39	-	.988	.039	.037			
Metric	59.111	47	-	.987	.038	.053	-.001	-.001	.016
Scalar	64.440	52	-	.986	.037	.053	-.001	-.001	.000
Invariance across educational level									
Configural	56.974	39	-	.974	.061	.042			
Metric	70.305	47	-	.967	.064	.076	-.007	.003	.034
Scalar	79.167	52	-	.961	.066	.069	-.006	.002	-.007
Invariance across L1									
Configural	54.001	39	-	.984	.049	.034			
Metric	62.199	47	-	.983	.044	.060	-.001	-.005	.026
Scalar	67.272	52	-	.983	.042	.045	.000	-.002	-.045

References

Yoon, M., & Lai, M. H. C. (2018). Testing factorial invariance with unbalanced samples.

Structural Equation Modeling: A Multidisciplinary Journal, 25, 201–213.

<https://doi.org/10.1080/10705511.2017.1387859>

Muthén, L. K., & Muthén, B. O. (1998-2019). *Mplus Version 8.3* [statistical software]. Muthén & Muthén.