The Development and Validation of the Short Form of the Foreign Language Enjoyment Scale (S-FLES)

Authors:

1. Elouise Botes (contact person)
University of Luxembourg
2 Avenue de l'Université
4365
Esch-sur-Alzette
Luxembourg

2. Prof. Jean-Marc Dewaele (Birkbeck College, University of London)
3. Prof. Samuel Greiff (University of Luxembourg)

Abstract

The Foreign Language Enjoyment Scale (FLES) was originally introduced by Dewaele and MacIntyre (2014). We used a data set with n = 1,603 learners of foreign languages (FL) to develop and validate the Short Form of the Foreign Language Enjoyment Scale (S-FLES). The data was split into two groups and we used the first sample to develop the short-form measure. A three-factor hierarchical model of FLE was uncovered, with FLE as a higher order factor and Teacher Appreciation, Personal Enjoyment, and Social Enjoyment as three lower order factors. We selected three items for each of the three lower order factors of the S-FLES. The proposed nine-item short form of FLE was validated in the second sample, and the fit statistics for the factor structure indicated close fit. Further evidence was found to support the internal consistency, convergent validity, and discriminant validity of the S-FLES. The S-FLES provides a valid and reliable short-form measure of FLE, which can easily be included in any battery of assessments examining individual differences in language learning.

The Development and Validation of the Short Form of the Foreign Language Enjoyment Scale (S-FLES)

In the field of applied linguistics in recent years, Foreign Language Enjoyment (FLE) has been established as one of the cornerstones of positive psychology research (Dewaele, Chen, Padilla & Lake, 2019; Khajavy et al., 2018; Li, 2020). FLE is defined as a broad positive emotion that is experienced by language learners when their psychological needs are met during challenging language-learning activities (Dewaele & MacIntyre, 2014). Evidence for the positive impact of FLE in the foreign language (FL) classroom has emerged in recent years, with FLE linked to better academic achievement (Li, 2020), quicker development of comprehensibility (Saito, Dewaele, Abe & In'nami, 2018), greater self-perceived competence (Dewaele & Dewaele, 2017), stronger motivation (Pavelescu, 2019; Saito et al., 2018), a greater willingness to communicate in the target language (Dewaele, 2019; Khajavy et al., 2018), and a lower Foreign Language Classroom Anxiety (FLCA; Dewaele & MacIntyre, 2014, 2019).

Initially, FLE was introduced by an accompanying 21-item Foreign Language Enjoyment Scale (FLES) as the measurement tool, which has since been increasingly used, with several adaptations and translations available (see Jin & Zhang, 2018; Li, Jiang, & Dewaele, 2018). For instance, FLE is frequently measured as part of a battery of instruments, each examining a separate variable, in studies that have hypothesised various relationships between applied psychology variables in a language context (see Khajavy et al., 2018; Saito et al., 2018; Wei et al., 2019). Of note, such batteries of instruments can result in rather lengthy questionnaires, which, in turn, have been found to have a negative effect on completion rates (Rolstad et al., 2011). A balance is therefore required between the effectiveness of the measure and its efficiency. To this end, the 21-item FLES can be seen as somewhat long when compared with other commonly used measures of

related constructs that have been used in similar studies, such as the eight-item Foreign Language Anxiety Scale (Dewaele & MacIntyre, 2014), which is a short version of Horwitz, Horwitz, and Cope's (1986) original 33-item scale, the 12-item Self-Perceived Competence Scale (McCroskey & McCroskey, 1988), and the 12-item Willingness to Communicate Scale (McCroskey & Baer, 1985).

A need has therefore arisen to develop a sound short version of the FLES that offers an optimal trade-off between psychometric strength and time saved without compromising the reliability and validity of the measure. In this paper, we therefore used the 21-item FLES to develop the Short Form of the Foreign Language Enjoyment Scale (S-FLES).

Measurement of FLE

The original 21-item FLES can be viewed as a global measure of the enjoyment of language learning and as such is broad in scope (Dewaele & MacIntyre, 2014). Its items were developed on the basis of Ryan et al.'s (1990) Interest/Enjoyment scale and refer to the enjoyment of learning, the atmosphere in the classroom, peers, and the teacher. Use of the scale has increased since its inception in 2014, and it is currently the most popular measure for examining positive emotion in language learning. It is most often used in its full 21-item form (see Mierzwa, 2018; Shirvan & Taherian, 2018). However, a shortened 10-item version based on Dewaele and MacIntyre's (2016) finding of separate dimensions for private and social FLE in the 21-item scale has been used by Dewaele, Witney, Saito, and Dewaele (2018) and Dewaele, Magdalena-Franco, et al., (2019). In addition to the 10- and 21-item versions, two Chinese versions for FL learners are available (see Jin & Zhang, 2018; Li et al., 2018). A summary of all FLE scales that are currently available is provided in Table 1. We discuss the development of these alternative scales at length later in this section.

Irrespective of the specific version that is used, the psychometric evidence for the validity and reliability of the scales employed to measure FLE is still emerging. Dewaele and MacIntyre (2014) reported an internal consistency reliability measured with Cronbach's alpha of $\alpha = .86$. Additional high levels of internal consistency reliability ($\alpha > .90$) have also been reported in other studies (Dewaele, Özdemir, et al., 2019; Shirvan & Taherian, 2018). The discriminant validity of the scale is often examined through a comparison with FLCA because FLE was developed as a positive emotion that should exhibit negative relations with FLCA. Indeed, the two constructs share a moderate negative correlation (Dewaele & MacIntyre, 2014, 2016). In their analysis of the SFLES instrument, Shirvan and Taherian (2018) conducted a confirmatory factor analysis of a unidimensional factor solution with the 21 items of the FLES loading on a single FLE factor and found satisfactory fit (RMSEA = .05; GFI = .91). However, several other studies have reported a multidimensional solution for the FLE construct as can be seen in more detail in Table 1 (Dewaele & MacIntyre, 2016; Li et al., 2018).

In a first attempt to shorten the scale, Dewaele et al. (2018) created a 10-item short scale that was based on a follow-up study of the introduction of FLE and the 21-item scale in 2014 (Dewaele & MacIntyre, 2016). The initial intention behind conducting the study was to further establish that FLCA and FLE were two separate constructs and could not be considered two opposite ends of the same continuum. As such, one exploratory factor analysis was conducted on all items on the 21-item FLES as well as the eight-item FLCA Scale. The analysis yielded three factors, and the FLCA items loaded on one of them. This pattern of results was used to argue for the independence of anxiety and enjoyment in language learning. The second and third factors, both related to the FLE items, were labelled Private Enjoyment and Social Enjoyment of language learning, with the majority of the 21 items of the FLES loading on the two factors. On the basis of

the factor loadings from the exploratory factor study, 10 items were chosen to create a shorter two-factor FLES, which has been used in further research (Dewaele, 2019; Dewaele, Magdalena-Franco, et al., 2019). The 10-item FLES yielded an acceptable internal reliability of $\alpha = .89$ in the study by Dewaele and MacIntyre (2019). However, to our knowledge, the 10-item FLE measure has not been subjected to any further validation studies, and the selection of the items was largely based on expert knowledge and not on the psychometric properties of the subscales or the items.

Despite its shortcomings, the 10-item measure provided the first multidimensionally constructed FLE scale. This measure is at odds with the idea that the structure of FLE is unidimensional and instead corresponds to the original theory behind FLE, which is reflected in the broad scope of the content of the items in the 21-item FLES (Dewaele & MacIntyre, 2014). Although a few of the previous studies that have examined the measurement of FLE have been implicit rather than explicit about the factor structure of FLE, an examination of previous FLE measures seems to indicate a certain pattern in the results, with studies identifying similar factors. More specifically, a private/individual enjoyment and a social/group-based enjoyment of language learning have been found by all multifactor studies (see Table 1). In addition, two papers included an additional third factor that referred to the role of the teacher in engendering enjoyment in the foreign language classroom (Jin & Zhang, 2018; Li, 2019). As two- and three-factor structures have been found in the majority of previous studies, we therefore expected that the S-FLES that is to be developed in this study would also turn out to be a multidimensional measure.

Furthermore, to our knowledge, the presence of a higher order factor underlying the data on FLE has not yet been considered. Within the theoretical framework, FLE has been posited to be a broad overarching positive variable in the FL learning classroom. Throughout the literature on individual differences, the use of higher order factors is supported, such that complex higher

order models have been found in research on personality (DeYoung, 2006), intelligence (McGrew, 2005), and anxiety (Muris et al. 2001). Indeed, there is no reason to suppose that the enjoyment of foreign language learning would be any more simplistic in its inherent factor structure than other individual difference variables are. Therefore, should a multidimensional factor structure emerge from the data, due consideration will be given to the inclusion of a higher order (i.e., second order) factor that is at the apex of the underlying first-order factors.

Another noteworthy point is that through the diversity of existing versions (see Table 1), several different item sets are available to applied linguistics researchers. We argue that there is a strong need for a psychometrically validated short form that is based on the 21-item FLES. The existing reduced form of the FLES, namely, the 10-item scale, was based on expert knowledge and was not developed with the aim of producing a reliable and valid short version. We argue that, as FLE has been accepted by the research community as an important psychological construct in its own right—separate from the oft-studied FLCA—it is an appropriate time to re-examine the existing data and to develop a short form on the basis of the best-practice guidelines for scale development utilizing two independent samples to develop and validate the scale (see Marsh, Ellis, Parada, Richards, & Heubeck, 2005). The aim of this paper was therefore to develop a short form that has satisfactory psychometric strength, is strongly rooted in theory, and reflects the underlying factor structure of FLE, with a minimum number of items.

Table 1

Available FLE Scales

Measure	Language	# Items	Validation evidence	Factor structure
Dewaele & MacIntyre (2014)	English	21	Cronbach's alpha reported (α = .86).	NA
Dewaele & MacIntyre (2016)	English	21	Exploratory factor analysis conducted with FLCA and FLE, resulting in two FLE factors emerging from the data.	1. Private-FLE 2. Social-FLE
Dewaele et al. (2018, 2019), Dewaele & MacIntyre (2019)	English	10	Based on the exploratory factor analysis conducted in Dewaele & MacIntyre (2016). Cronbach's alpha reported (α = .88, and α = .87), respectively.	1. Private-FLE 2. Social-FLE
Li et al. (2018)	Chinese	11	Exploratory and confirmatory factor analyses resulted in a three-factor model that fit the data well.	1. FLE-Private 2. FLE-Teacher 3. FLE-Atmosphere

Jin & Zhang (2018, 2019)	Chinese	16	Exploratory and confirmatory factor analyses resulted in a three-factor model that fit the data well.	1. Enjoyment of Teacher Support 2. Enjoyment of English Learning 3. Enjoyment of Student Support

On Short-Form Development

A number of considerations should be made in the construction of a short form, and there are several best-practice guidelines on how to select items for a short-form measure (see Marsh et al., 2005; Smith, McCarthy, & Anderson, 2000). When selecting items for a short form, consideration should be given to the reliability, validity, and theoretical underpinnings of each item and the scale as a whole (Putnam & Rothbart, 2006). However, the research on the development of short-form scales of existing measures has warned that the shorter administration time is not always worth the inevitable loss of validity (Smith et al., 2000) and that the psychometric properties of short-form measures are sometimes overstated (Marsh et al., 2005).

While keeping these precautions in mind, it has been argued that short-form development through psychometrically sound methods will provide a valid and reliable alternative to the already existing 10-item measure in the case of FLE. The rising popularity of FLE as a research topic (Dewaele, Chen, et al., 2019; Dewaele & Li, 2020; Dewaele et al., 2019; Shao, Nicholson, Kutuk & Lei, 2020) has created a need for a validated measure that can be used as part of a battery of instruments where the trade-off between the number of items and the validity of the scale has been optimised. An additional benefit provided by the process of developing a short form of the 21-item FLES is that the dimensionality of the construct could also be examined further. There is still no consensus regarding the number of factors underlying FLE, with statistically significant fit found for a single factor model (Shirvan & Taherian, 2018), a two-factor model (Dewaele & MacIntyre, 2016), and a three-factor model (Li et al., 2019). The development of the short scale will therefore contribute to the ongoing research dialogue regarding the structure of FLE.

Method

Participants

The sample consisted of *n* = 1,603 adult FL learners from around the world. The average age of the sample was 24.81 years (*SD* = 8.34), and 72.56% of the sample was female. The sample self-reported a high level of multilingualism with the average participant listing 3.47 languages (*SD* = 1.28) in their repertoire. A total of 43 different languages were being learned by the sample, with the majority learning English (*n* = 738), followed by French (*n* = 208) and Spanish (*n* = 170).

These data were made available from a previous study examining FLCA and FLE (Dewaele & MacIntyre, 2014). The data were collected in 2012 on an online platform through the use of snowball sampling. The sample was previously used to examine the relationship between FLCA and FLE (Dewaele & MacIntyre, 2014) as well as the social and private enjoyment factors underlying the construct of FLE (Dewaele & MacIntyre, 2016). However, the data have not yet been used with regard to developing a short form of the FLE.

Instruments

FLE and FLCA were measured with previously established self-report questionnaires.

Foreign Language Enjoyment Scale. The original 21-item scale introduced by Dewaele and MacIntyre (2014) is aimed at measuring positive emotions in language learning, with items such as ‘I enjoy my FL class’ and ‘There is a positive environment in my FL class’. Items are rated on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*; α = .86).

Foreign Language Classroom Anxiety Scale. The eight-item measure is a shortened version of the 33-item scale developed by Horwitz, Horwitz, and Cope (1986), with items such as

'Even if I am well-prepared for my FL class, I feel anxious about it'. Items are related on a 5-point Likert scale ranging from 1 (*strongly disagree*) to 5 (*strongly agree*). The FLCAS was included in the study as a measure of discriminant validity because the moderate, negative relationship between FLE and FLCA was confirmed in a recent meta-analysis (α = .86; Botes et al., 2021).

Data Analysis

In the development and validation of the S-FLES, the analysis and interpretation of results followed five major sequential steps, following the criteria specified in Marsh et al. (2005). A flowchart of the steps can be found in Figure 1.

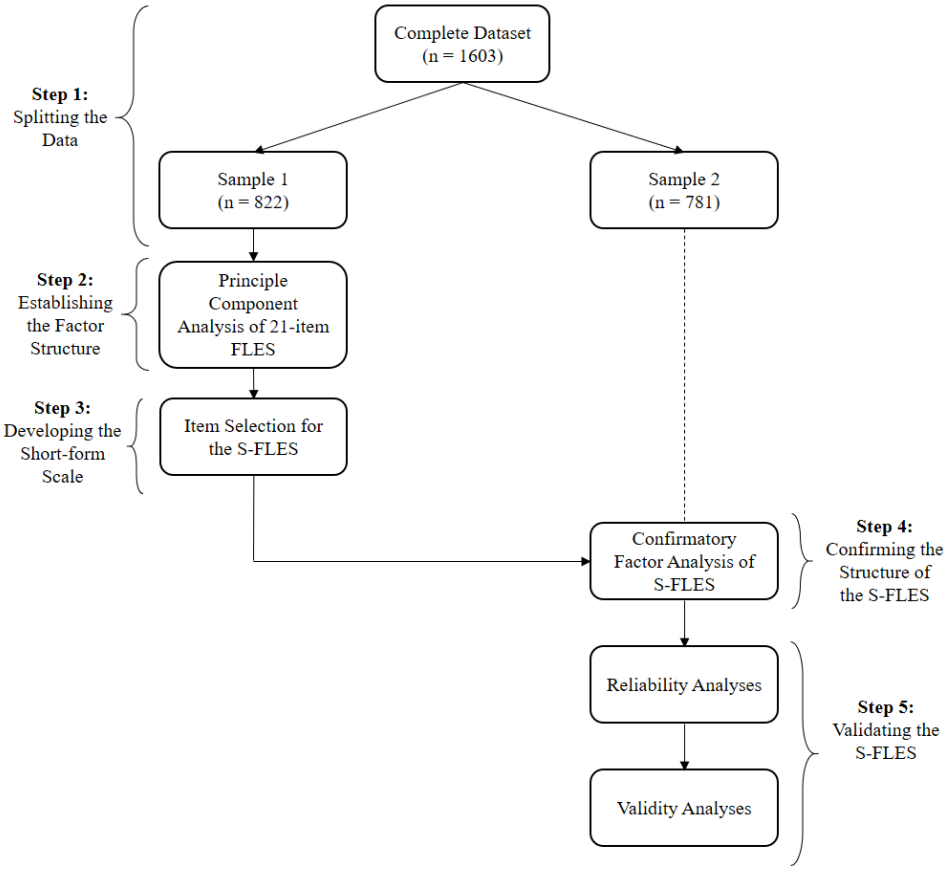

Figure 1. Research methods flowchart.

Step 1: Splitting the data set. In order to ensure that the validity and reliability of the final product—the Short Form of the Foreign Language Enjoyment Scale (S-FLES)—could be examined in line with best practice recommendations (see Hagtvet & Sipos, 2016; Marsh et al., 2005), the data set was randomly split into two samples using SPSS 25. The exploratory phase of the study (examining the underlying structure and developing a short-form scale) was conducted with Sample 1, whereas the confirmatory phase of the study utilised Sample 2 (examining the validity and reliability of the resultant S-FLES).

Step 2: Establishing the factor structure. On the basis of an initial Principle Component Analysis (PCA), we employed three criteria to uncover the number of factors and, thus, the factor structure of the 21-item FLES, namely: the eigenvalue greater than 1 criterion, the scree plot, and parallel analysis.

PCA with an oblique (promax) rotation was performed in JASP (JASP Team, 2020). PCA was chosen as it creates a simplified description of the data that is easy to interpret (Field, 2005). Furthermore, as PCA has the goal of reducing the number of items and variables, it has been found to be particularly useful in the creation of short-form measures (McGuire et al., 2010; Stevanovic, 2014). Oblique (promax) rotation was chosen as it is theoretically assumed that the factors underlying FLE are correlated (Field, 2005; Li et al., 2018). Factor loadings were classified as low (< .4), acceptable (.4 to .6), or high (> .6; Kline, 2014).

The first criterion used to determine the number of factors was the oft-used Kaiser criterion by which the number of components with an eigenvalue > 1 (Kanyongo, 2005) are retained. The second criterion involved the examination of the scree plot with the eigenvalues depicted against their ordinal numbers (Tabachnick & Fidell, 2001). Finally, the third criterion we used was that of parallel analysis, which has been described as a useful tool for examining the number of factors to

retain in the development of a measure, as it is less subject to sensitivity and variability as compared with, for example, the scree plot (Ledesma & Valero-Mora, 2007; Zwick & Velicer, 1986). Parallel analysis is a Monte Carlo-esque simulation that is 'based on the generation of random variables' (Ledesma & Valero-Mora, 2007, p. 3). Informed decisions about the number of factors that would be suitable for the S-FLES and which items to select were made on the basis of these three criteria plus the theoretical underpinnings of FLE as a construct (see Hagtvet & Sipos, 2016) as outlined in more detail below.

Step 3: Developing the Short-Form Scale. Based on the results of the factor analysis in Step 2 in Sample 1, the individual factors and items were examined in SPSS 25. Items were selected for each of the factors when they exhibited the following characteristics (Marsh et al., 2005): (a) when they had large item-total correlations, (b) when they had large factor loadings on the focal factor in question and small factor loadings on other factors (i.e., small cross-loadings), and (c) when they encapsulated the theoretical rationale of the focal factor and FLE as a whole.

In an effort to validate the hand-chosen items on the basis of the criteria described above, in addition to the above-mentioned procedure the ant colony optimisation (ACO) algorithm was used to confirm the selection of items on a purely mathematical basis. The name ACO algorithm was borrowed from the behavioural patterns of ants, which are capable of finding the shortest route to a food source by utilising pheromones (Olaru et al., 2015). Similarly, the ACO algorithm emulates the behaviour of ants in that the algorithm utilises probabilities to create a set of items that cannot be improved upon in terms of pre-specified criteria, model fit and reliability in our specific case—thus creating the shortest possible route to a well-fitting model (Dörendahl & Greiff, 2020). The ACO algorithm was implemented through the Shortform R package (Raborn & Leite, 2018). The algorithm requires the specification of the number of factors in the measure, the

pool of items from which the measure is to be constructed, and the number of items to include in each factor. Therefore, the ACO algorithm can identify the items that should be included in a measure in order to optimise the fit, but the numbers of factors and items to be selected need to be predetermined. Therefore, the ACO algorithm was used as a fail-safe check to determine which items should be selected for the S-FLES.

Furthermore, in the case of empirical support for a multidimensional solution to FLE (as theoretically expected), we explored whether a second-order factor should be included in the measurement model.

Step 4: Confirming the structure of the S-FLES. The resultant S-FLES from Sample 1 was independently tested via a confirmatory factor analysis in R with the lavaan package (Rosseel, 2012) using Sample 2 from the data set. The fit of the measurement model was examined via the comparative fit index (CFI), the root mean square error of approximation (RMSEA), and the standardised root mean square residual (SRMR) (Kline, 2014). Additionally, the absolute fit values of the chi-square (χ^2) and the chi-square relative to the degrees of freedom (χ^2/df) were taken into consideration. Finally, the factor loadings, cross-loadings, and error variables in the measurement model were examined to further determine the fit.

Step 5: Validity and reliability of the short form. Reliability and validity coefficients were analysed utilising Sample 2 of the data set. The reliability measures that we examined included internal consistency as measured with Cronbach's alpha and McDonald's omega as well as the split-half reliability.

In order to establish the construct validity of the S-FLES, we analysed both convergent and discriminant validity. Convergent validity was examined by computing the correlation between

the long form (i.e., the FLES) and the newly developed S-FLES in Sample 2. Discriminant validity was examined by conducting an average variance extracted (AVE) analysis. The analysis examined discriminant validity by comparing the square root of the AVE value of each factor with the correlation between any pair of factors (Farrell, 2010), where discriminant validity is established if the estimates of the variance extracted by the factors underlying FLE are greater than their squared correlation estimates (Hair et al., 2006). Thus, the AVE method establishes discriminant validity on a construct level. In addition, discriminant validity was further examined by correlating the S-FLES total values in Sample 2 with the respective scores on the FLCA scale—a variable that is known to be negatively related to enjoyment in language learning (Dewaele & MacIntyre, 2014).

Results

Step 1: Splitting the Datafile

Following the research steps outlined previously, the datafile was randomly split into two halves. The demographics of the two halves can be found in Table 2, which demonstrates that there were no statistically significant differences on these variables between Samples 1 and 2.

Table 2

Demographic Information for Samples 1 and 2

	Sample 1	Sample 2
Sample size	822	781
Gender	71.53% female	72.86% female
Age	*M* = 24.78 (*SD* = 8.11)	*M* = 24.84 (*SD* = 8.59)
Level of multilingualism	*M* = 3.54 (*SD* = 1.325)	*M* = 3.39 (*SD* = 1.24)

The descriptive statistics for Samples 1 and 2 regarding the full 21-item FLES and FLCAS can be found in Table 3. There were no statistically significant differences between the two samples with regards to demographics or item scores on the FLES or FLCAS.

Table 3

Descriptive Statistics

	M	*SD*	Min	Max
Sample 1				
FLES	80.68	9.27	40	103
FLCAS	21.70	6.72	8	40
Sample 2				
FLES	80.72	9.16	47	104
FLCAS	22.34	6.56	8	39

Step 2: Establishing the Factor Structure

The 21-item FLES administered to Sample 1 was analysed via PCA with an oblique (promax) rotation. The initial analysis produced a solution with four factors according to the eigenvalue criterion (i.e., eigenvalue greater than 1). At this point, we did not inspect the scree plot or compute a parallel analysis because four items exhibited weak or negative loadings (< .4) on all four factors that were extracted (Items 1, 2, 5, and 6). As these four items are not paramount to the design and theory of FLE, we decided not to proceed with the full set of items and to exclude these four items in all future analyses in order to ensure that items with unclear loading patterns and generally weak loadings would not unduly influence the factor structure.

A subsequent second PCA was conducted on the 17 remaining items from Sample 1. The factor loadings for each item can be found in Table 4 (please note that only the loadings on the first three factors are depicted).

Table 4

PCA of the 17-item FLES with a Promax Rotation

Item	Factor 1	Factor 2	Factor 3
3		.548[a]	
4		.799[b]	
7		.580[a]	
8		.769[b]	
9		.722[b]	
10	.425[a]		
11		.561[a]	
12		.711[b]	
13	.453[a]		
14	.520[a]		
15	.940[b]		
16	.985[b]		
17	.937[b]		
18	.656[b]		
19			.770[b]
20			.908[b]
21			.720[a]

Note. Only loadings > .4 are displayed.
[1]acceptable loading (.4 to .6). [2]high loading (> .6).

In order to decide on the most appropriate number of factors in the PCA, we applied the three criteria outlined in the methods section: eigenvalues > 1, scree plot and parallel analysis.

Three factors had an eigenvalue > 1, indicating that three components may underlie the FLES. In turn, the scree plot can be interpreted to support either a three-factor solution or a unidimensional solution (see the line with the circles in Figure 2). The scree plot shows an inflection point after the first factor, indicating that a single factor solution might underlie the 17-item measure of FLE. However, items loaded with acceptable loadings of > .4 on three separate factors, and a second inflection point is visible after the third factor in the scree plot, suggesting that a three-factor solution might be feasible as well. Finally, the simulated data from the parallel analysis further supported the finding that a three-factor solution underlies FLE (see the line with the triangles in Figure 2 and the point where the two lines cross).

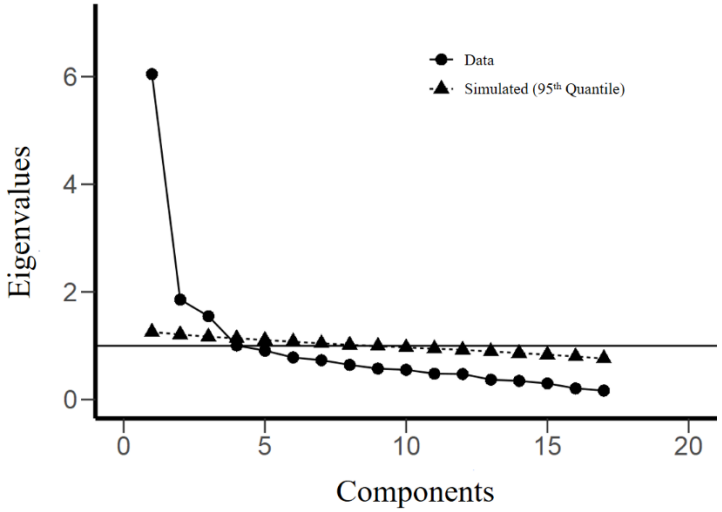

Figure 2. Scree plot of FLE.

Overall, the results of the eigenvalue analysis, scree plot, and parallel analysis tended to indicate that a three-factor solution underlies the FLES. This multidimensional interpretation of FLE is also in line with previous research (see Table 1) and the theoretical considerations of the variable itself (Dewaele & MacIntyre, 2014). However, because of the first inflection point in the scree plot, a one-factor solution might also reflect the number of factors according to these criteria.

Thus, further steps were taken to confirm or disconfirm whether a unidimensional solution best captures the variance in the FLES via the quick and efficient method of the ACO algorithm (see above for further details on the method). Using the ACO algorithm, the unidimensional solution did not fit the data. The results of the ACO algorithm with regard to the unidimensional solution can be found in the Supplementary Materials.

Thus, and in line with previous theoretical considerations, the three-factor solution was used in the further development of the S-FLES. The first factor that we extracted had three items with high loadings > .9 (Items 15 to 17; Kline, 2014). These items all referred to the role of the teacher in creating a positive environment in the FL classroom, with items such as 'The teacher is encouraging' (Item 15). The second factor seemed to capture a personal enjoyment of FL learning, as items such as 'I enjoy it' (Item 4) and 'I've learned interesting things' (Item 8) loaded most strongly on the factor. In turn, the third factor tapped into the social enjoyment of FL learning. The items on this factor all seemed to indicate a positive social environment linked to social cohesion and solidarity with peers, as items such as 'We form a tight group' (Item 19) loaded on this factor.

The 17-item FLES therefore yielded a three-factor solution to the data: Teacher Appreciation, Personal Enjoyment, and Social Enjoyment of FL learning.

Step 3: Development of the S-FLES

We investigated the factor structure of the S-FLES by using PCA. A multidimensional solution for the FLES was uncovered with three factors considered for inclusion in the S-FLES: Teacher Appreciation, Personal Enjoyment, and Social Enjoyment. Subsequently, the number of

items was selected from each of the three factors on the basis of their respective factor loadings as well as the design intent and theoretical reasoning underpinning FLE.

Teacher Appreciation Subscale. The first factor that emerged was that of Teacher Appreciation, which can be defined as the extent to which the learner perceived that their psychological needs were met by the FL teacher. The importance of the teacher for students' overall enjoyment of foreign language classes has been found in several studies (Dewaele et al., 2018; Dewaele & MacIntyre, 2019; Dewaele, Magdalena-Franco, et al., 2019; Jin & Zhang, 2018; Li et al., 2018). The three items that had the highest loadings on the first factor all specifically referred to the role of the teacher (Items 15, 16, and 17) and were therefore retained in the S-FLES.

The internal consistency of the subscale as measured by Cronbach's alpha and McDonald's omega was high ($\alpha = .92$, $\omega = .93$), with high inter-item correlations (see the Supplementary Materials).

Personal Enjoyment Subscale. The second factor sindicated a personal enjoyment of language learning. A personal or private enjoyment factor underlying FLE has been substantiated in the literature (see Dewaele & MacIntyre, 2016; Li et al., 2018). As such, the second factor included in the S-FLES was that of Personal Enjoyment. Items selected for this subscale were based on the inter-item correlations, the internal consistency of the subscale, and the factor loadings. Therefore, we selected Items 4, 8, and 9. Each item referred to the individual personal enjoyment of foreign language learning and had a satisfactory loading on the second factor that was extracted in the EFA.

The subscale yielded an acceptable internal consistency ($\alpha = .71$, $\omega = .72$), with acceptable inter-item correlations (see the Supplementary Materials).

Social Enjoyment Subscale. The third factor chosen for the S-FLES was that of Social Enjoyment, which refers to the fulfilment of social psychological needs in the FL classroom. Only three items (Items 19, 20, and 21) loaded on this factor, and as such, all three were retained in the S-FLES. This subscale encapsulates the enjoyment of the social interactions that take place and the social environment of the FL class as a whole, echoing previous validation studies (Jin & Zhang, 2018; Li et al., 2018). Not surprisingly, the items on this subscale all begin with the plural pronoun 'we', which refers to the participant and their peers.

The internal consistency of the subscale was acceptable ($\alpha = .77$, $\omega = .77$), with moderate inter-item correlations (see the Supplementary Materials).

ACO algorithm of the multidimensional model. The ACO algorithm was further used to provide a purely mathematical, a-theoretical confirmation of the items that had been selected. This additional analysis was implemented to ensure that the selection of the items as outlined above could be confirmed through ACO a data driven method.

An ACO algorithm was implemented to identify the three items from each of the three factors (Teacher Appreciation, Personal Enjoyment, Social Enjoyment) that would optimise the fit statistics of the measurement model. The algorithm could choose from all items belonging to this subscale. The ACO algorithm identified items that were nearly identical to the items selected above on theoretical grounds and on factor loadings. Teacher Appreciation was indicated by Items 15, 16 and 17; Personal Enjoyment by Items 3, 4, and 8; and Social Enjoyment by Items 19, 20, and 21 according to the ACO.

The difference between the ACO algorithm and the selection of items on the basis of underlying theory and the factor loadings can be seen in the selection of Item 3 by ACO versus

Item 9 on the Personal Enjoyment factor. Item 3, which reads ‘I don’t get bored’, was selected by the ACO algorithm, whereas we hand-selected Item 9 instead (‘In class, I feel proud of my accomplishments’). Boredom as represented by Item 3 in the FL classroom has recently been garnering increased research attention (see Pawlak et al., 2020a, 2020b), and has been demonstrated to be a separate, unique emotion in FL learning (Pawlak et al., 2020a). Therefore, due to theoretical considerations, Item 3, which was suggested by the ACO algorithm, was not retained. Instead, Item 9 was included in the S-FLES.

Thus, the S-FLES was constructed as a short form of the FLES, comprising three subscales (Teacher Appreciation, Personal Enjoyment, and Social Enjoyment) with three items on each subscale. The subscales and the subsequent items were developed on the basis of both statistical analyses and theoretical considerations as indicated by best-practice guidelines (Hagtvet & Sipos, 2016; Marsh et al., 2005).

Higher order FLE factor. As stated in the research design section, we also explored whether a second-order factor was appropriate in the development of the S-FLES. The nine-item, three-factor design of the S-FLES allows for a fine-grained assessment of FLE but also lends itself to the possibility of a higher order (i.e., second-order) FLE factor (see Figure 3).

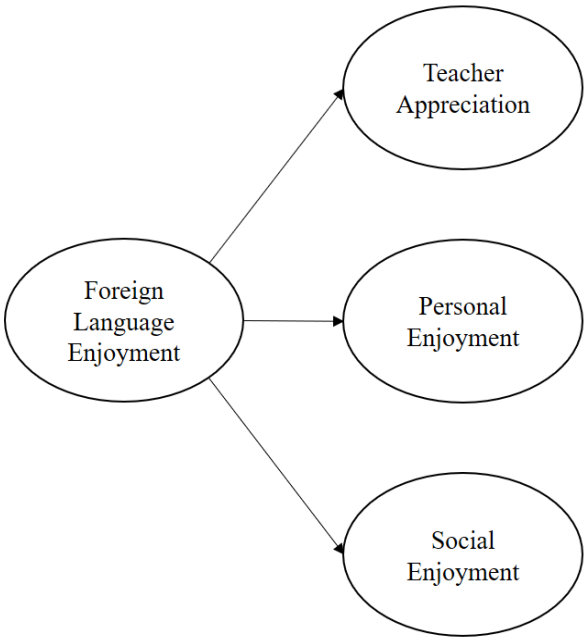

Figure 3. Higher order factor model of FLE.

Several arguments can be made in support of such a second-order factor. The majority of research in the field of individual differences utilises sum scores of variables to determine correlations and linear trends, for example, the relationship between FLE and FLCA (Dewaele & MacIntyre, 2014), FLE and academic achievement (Li, 2020), and FLE and willingness to communicate (Dewaele, 2019). A practical argument can therefore be made to include a higher order factor in the measurement model in order to provide a clear consensus that future use of a sum score utilising the S-FLES is permissible in research studies as an alternative to the more fine-grained use along the three sub-dimensions. Furthermore, a theoretical argument can also be made for the inclusion of a higher order factor, as the variable was developed as an overarching enjoyment factor in FL learning that encapsulates positive emotions in the FL classroom (Dewaele & MacIntyre, 2014).

It should be noted that the second-order factor structure as proposed in Figure 3 cannot be empirically compared with a correlational model without a second-order factor (i.e., a first-order

correlated factor model). The two models are mathematically identical as the parameters to be estimated do not differ. Therefore, the decision to include a second-order factor (instead of a correlated first-order factor model) was largely a practically and theoretically driven decision.

Step 3 of the development and validation of the S-FLES therefore concluded with a nine-item, hierarchical model consisting of three first-order factors (Teacher Appreciation, Personal Enjoyment, and Social Enjoyment) and a single second-order factor (FLE) as depicted in Figure 3.

Step 4: Confirming the Structure of the S-FLES

Subsequently, a confirmatory factor analysis of the proposed nine-item higher order S-FLES with the three dimensions Teacher Appreciation, Personal Enjoyment, and Social Enjoyment loading on a second-order FLE factor was tested in Sample 2 (see Figure 4).

Overall, the fit statistics suggested good fit, with the Root Mean Square Error of Approximation (RMSEA =.059) falling below the recommended cut-off of .08 (Kline, 2005). In turn, the Comparative Fit Index (CFI = .978) and the Tucker-Lewis Index (TLI = .967) both indicated close fit (Kline, 2005). The Satorra-Bentler Chi-Square was significant, $\chi^2(24) = 89.25$; $p < .001$, with a chi-square to degrees of freedom ratio of $\chi^2/df = 3.72$, which was above the recommended ratio of 2 (Byrne, 1989). However, it is important to acknowledge that the sample in question ($n = 781$) can be considered large, and the correlations between the factors can be considered moderate to large ($.39 \leq r \leq .84$). The chi-square and the chi-square/*df* ratio have been found to be sensitive to the sample size and prone to Type I errors (Kenny, 2020).

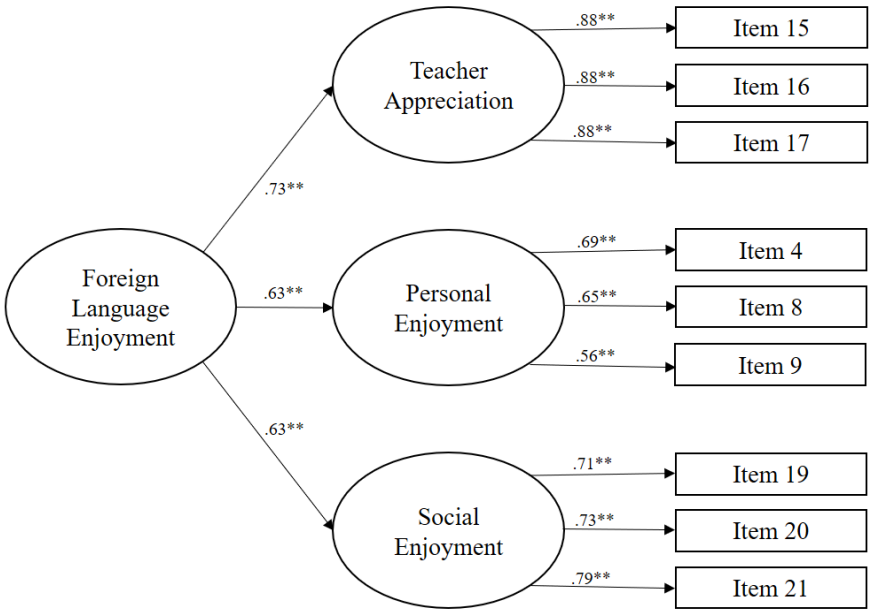

Figure 4. Measurement model of the S-FLES. ** $p < .001$.

Overall, the model provided a sufficient fit to the data, and the structure of the S-FLES as hypothesised with Sample 1 was confirmed in Sample 2.

Step 5: Validating the S-FLES

As the structure of the S-FLES was empirically validated in Step 4, we examined the reliability and validity of the proposed S-FLES in Sample 2.

Reliability

The internal consistency of the 9 items of the S-FLES in Sample 2 as measured with Cronbach's alpha and McDonald's omega was high (α = .81, ω = .82). In addition to this, each of the factors demonstrated high internal consistency: Teacher Appreciation (α = .91, ω = .91), Personal Enjoyment (α = .71, ω = .70), and Social Enjoyment (α = .79, ω = .79).

Convergent Validity

In order to establish convergent validity, we calculated the correlation between the full 21-item FLES and the nine-item S-FLES. The correlation between the FLES and S-FLES was very high ($r = .90$, $p < .001$), indicating nearly identical rank orders for the full and the short versions of the FLE.

Discriminant Validity

Discriminant validity was examined by comparing the average variance extracted (AVE) with the squared correlation coefficients (r^2) between the three subscales. The correlation matrix generated in the CFA was utilised in order to take measurement error into account (Farrell, 2010). Discriminant validity was indicated for each subscale as the AVE of each of the subscales was larger than the highest squared correlation (r^2) for all subscales (see Table 5). Therefore, the items on each of the subscales explained more variance in the specific subscale than the items on the other subscales, thus establishing discriminant validity between the subscales of the S-FLES (Zait & Bertea, 2011).

Table 5

Average Variance Extracted and Coefficients of Determination in Sample 2

		Coefficients of determination (r^2)			
Subscales	AVE	1.	2.	3.	AVE > r^2
Teacher Appreciation	.780	-	.185**	.212**	Yes
Personal Enjoyment	.467		-	.102**	Yes
Social Enjoyment	.554				Yes

**$p < .001$.

Furthermore, as the moderate negative relationship between FLCA and FLE is well-established in the literature (see Dewaele & MacIntyre, 2014, 2016), discriminant validity was indicated by the moderate negative correlations between the S-FLES and the FLCA ($r = -.241$, $p < .001$).

Discussion

The research aim of this study was to develop a psychometrically sound short-form measure of FLE. With a second-order factor and three first-order factors, the nine-item S-FLES was derived from the original 21-item scale by following five major steps. The newly developed S-FLES was found to be both valid and reliable.

In the development of the S-FLES, a crucial step involved establishing the factor structure underlying FLE. In the process of uncovering the factor structure of the S-FLES, both unidimensional and multidimensional solutions were considered. For theoretical and psychometric reasons, a unidimensional factor structure was rejected. Instead, a three-factor structure emerged

from the data. The three factors underlying FLE in the S-FLES were Teacher Appreciation, Personal Enjoyment, and Social Enjoyment.

The three-factor structure further replicated the findings of Jin and Zhang (2018) and Li et al. (2018), who found similar factor structures underlying two Chinese samples. Indeed, the Personal Enjoyment and Social Enjoyment aspects underlying FLE have been found in the majority of studies examining the factor structure of FLE (see Table 1). The role of the teacher in establishing and furthering students' enjoyment in FL class has also been found in quantitative (Dewaele, 2019; Dewaele et al., 2018; Jin & Zhang, 2018) and qualitative studies (Dewaele & Pavelescu, 2019; Pavelescu & Petrić, 2018; Shirvan & Talebzadeh, 2019). Therefore, we have considerable confidence in our choice of a three-factor structure of the S-FLES as captured by Teacher Appreciation, Personal Enjoyment, and Social Enjoyment in this study.

The proposed structure of the measurement model that we tested included a higher order FLE factor. We included the higher order factor on the basis of theoretical considerations. FLE was introduced to the FL learning literature as a broad overarching positive emotion that facilitates the outcome of FL learning. When using the scale in the future, it is therefore possible to utilise a total score on the S-FLES or, depending on the specific research questions at hand, a more fine-grained use of the three S-FLES subscales that were identified in this study.

The proposed measurement model of the S-FLES indicated good fit. In addition, the S-FLES demonstrated good reliability and validity. The internal consistency of the scale was acceptable on a subscale level (Teacher Appreciation: $\omega = .91$; Personal Enjoyment: $\omega = .70$; Social Enjoyment: $\omega = .79$) and a global level (S-FLES: $\omega = .82$). Evidence was found for the convergent and divergent validity of the scale.

The use of the S-FLES is therefore recommended for studies examining enjoyment in FL learning in a classroom FL learning context and with time constraints. Due to the limited number of items, the measure can easily be included in research studies examining multiple variables. It is our hope that introducing this measure to the field will increase the understanding of the nomological network that FLE is located in. Past research has established a clear nomological network of the negative emotion of FLCA, which can be linked to language proficiency (Botes, Dewaele, & Greiff, 2020), learning difficulties (Chen & Chang, 2004), language attitudes (Phillips, 1992), language beliefs (Oh, 1997), willingness to communicate in the target language (MacIntyre et al., 2002), personality traits (Dewaele, 2017), age (Onwuegbuzie, Bailey, & Daley, 1997), and gender (Park & French, 2013). In the current wave of positive psychology research in the applied linguistics field, the S-FLES may prove to be a useful tool for expanding the knowledge of FLE to rival that of its negative emotion counterpart, FLCA.

The expansion of research on FLE and the S-FLES in particular should include cross-validation studies to ensure the suitability of the measure across FL learning contexts. Such studies may take the form of translated versions of the S-FLES to examine the suitability of the non-English administration of the measure. Further debate and research regarding the factor structure and introduction of a higher order FLE factor is also recommended. Additionally, invariance across gender, age, target language groups, and cultural contexts needs to be established in order to further validate the measure. The current study is limited in the fact that we used the same sample that was used in the original introduction of FLE (Dewaele & MacIntyre, 2014). In this, our study is a development of a short form that was based on the original data and not a new cross-validation. Any and all future use of the S-FLES will therefore contribute to the understanding of the validity and reliability of the measure.

It should be noted that the use of the S-FLES might not be appropriate for all research studies examining emotions in language learning. The prominence of the Teacher Appreciation subscale in the measure means the S-FLES is not appropriate for use in self-driven FL learning. Finally, the measure was developed and validated with the use of an adult FL learning sample and is therefore not recommended for use for children without prior validation studies.

Conclusion

In this study, we developed and validated a short-form measure of FLE based on the 21-item FLES. The nine-item, three-factor S-FLES was found to be a valid and reliable instrument. Use of the measure is recommended for any study examining the FLE of adolescents or adults in a FL classroom context. It is therefore with confidence that we introduce the use of this measure in the hope of furthering the research output of emotions in applied linguistics.

References

Botes, E., Dewaele, J.-M., & Greiff, S. (2021). *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

Byrne, B. M. (2005). Factor Analytic Models: Viewing the structure of an assessment instrument from three perspectives. *Journal of Personality Assessment, 85*(1), 17-32. https://doi.org/10.1207/s15327752jpa8501_02.x

Chen, T. Y., & Chang, G. B. (2004). The relationship between foreign language anxiety and learning difficulties. *Foreign Language Annals*, *37*(2), 279-289. https://doi.org/10.1111/j.1944-9720.2004.tb02200.x

Dewaele, J.-M. (2017). Psychological dimensions and foreign language anxiety. In S. Loewen & M. Sato (Eds.), *The Routledge Handbook of Instructed Second Language Acquisition* (pp. 433-450. New York, NY: Routledge

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., & Dewaele, L. (2017). The dynamic interactions in foreign language classroom anxiety and foreign language enjoyment of pupils aged 12 to 18. A pseudo-longitudinal investigation. *Journal of the European Second Language Association*, *1*(1), 12–22. https://doi.org/10.22599/jesla.6

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*, *4*(2), 273-274.

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., & MacIntyre, P.D. (2019). The predictive power of multicultural personality traits, learner and teacher variables on foreign language enjoyment and anxiety. In M. Sato & S. Loewen (Eds.), *Evidence-based second language pedagogy: A collection of Instructed Second Language Acquisition studies* (pp. 263-286). London: Routledge.

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

Dewaele, J.-M., & Li, C. (2020). Emotions in Second Language Acquisition: A critical review and research agenda. *Foreign Language World* [Chinese 外语界], 196(1), 34-49.

http://kns.cnki.net/kcms/detail/31.1040.h.20200227.1306.008.html

Dewaele, J.-M., Magdalena-Franco, A., & Saito, K. (2019). The effect of perception of teacher characteristics on Spanish EFL Learners' anxiety and enjoyment. *The Modern Language Journal*, *103*(2), 1–40. https://doi.org/10.1111/modl.12555

Dewaele, J.-M., Özdemir, C., Karci, D., Uysal, S., Özdemir, E. D., & Balta, N. (2019). How distinctive is the foreign language enjoyment and foreign language classroom anxiety of Kazakh learners of Turkish? *Applied Linguistics Review* (published online ahead of print). https://doi.org/10.1515/applirev-2019-0021

Dewaele, J.-M., Witney, J., Saito, K., & Dewaele, L. (2018). Foreign language enjoyment and anxiety: The effect of teacher and learner variables. *Language Teaching Research*, *22*(6), 676-697. https://doi.org/10.1177%2F1362168817692161

DeYoung, C. G. (2006). Higher-order factors of the Big Five in a multi-informant sample. *Journal of personality and social psychology*, *91*(6), 1138-1151. https://psycnet.apa.org/doi/10.1037/0022-3514.91.6.1138

Dörendahl, J., & Greiff, S. (2020). Are the machines taking over? Benefits and challenges of using algorithms in (short) scale construction [Editorial]. *European Journal of Psychological Assessment, 36*(2), 217–219. https://doi.org/10.1027/1015-5759/a000597

Farrell, A. M. (2010). Insufficient discriminant validity: A comment on Bove, Pervan, Beatty, and Shiu (2009). *Journal of Business Research*, *63*(3), 324–327. https://doi.org/10.1016/j.jbusres.2009.05.003

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

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

Hair, J. F., Black, W. C., Babin, B. J., Anderson, R. E., & Tatham, R. L. (2006). *Multivariate Data Analysis* (6th edition). Upper Saddle River, NJ: Prentice Hall.

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

JASP Team (2020). JASP (Version 0.13.1) [Computer software]. Amsterdam, The Netherlands: JASP.

Jin, Y., & Zhang, L. J. (2018). The dimensions of foreign language classroom enjoyment and their effect on foreign language achievement. *International Journal of Bilingual Education and Bilingualism*, 1–15. https://doi.org/10.1080/13670050.2018.1526253

Kanyongo, G. Y. (2005). The influence of reliability on four rules for determining the number of components to retain. *Journal of Modern Applied Statistical Methods*, *5*(2), 332-343.

Kenny, D. A. (2020). Measuring Model Fit. Retrieved from http://davidakenny.net/cm/fit.htm

Khajavy, G. H., MacIntyre, P. D., & Barabadi, E. (2018). Role of the Emotions and Classroom Environment in Willingness to Communicate. *Studies in Second Language Acquisition*, *40*(3), 605–624. https://doi.org/10.1017/S0272263117000304

Kline, P. (2014). *An easy guide to factor analysis*. Milton Park, UK: Taylor and Francis.

Koufteros, X., Babbar, S., & Kaighobadi, M. (2009). A paradigm for examining second-order factor models employing structural equation modeling. *International Journal of Production Economics*, *120*(2), 633-652. https://doi.org/10.1016/j.ijpe.2009.04.010

Ledesma, R. D., & Valero-Mora, P. (2007). Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. *Practical assessment, research, and evaluation*, *12*(1), 2. https://doi.org/10.7275/wjnc-nm63

Li, C. (2020). A Positive Psychology perspective on Chinese EFL students' trait emotional intelligence, foreign language enjoyment and EFL learning achievement. *Journal of Multilingual and Multicultural Development*, *41*(3), 246–263. https://doi.org/10.1080/01434632.2019.1614187

Li, C., Jiang, G., & Dewaele, J.-M. (2018). Understanding Chinese high school students' Foreign Language Enjoyment: Validation of the Chinese version of the Foreign Language Enjoyment scale. *System*, *76*, 183–196. https://doi.org/10.1016/j.system.2018.06.004

MacIntyre, P. D., Baker, S. C., Clément, R., & Donovan, L. A. (2002). Sex and age effects on willingness to communicate, anxiety, perceived competence, and L2 motivation among junior high school French immersion students. *Language Learning*, *52*(3), 537-564.

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

McCroskey, J. C., & Baer, J. E. (1985, November). *Willingness to communicate: The construct and its measurement* [Conference presentation]. Annual Meeting of the Speech Communication Association, Denver, CO.

McCroskey, J. C., & McCroskey, L. L. (1988). Self-report as an approach to measuring communication competence. *Communication Research Reports, 5*(2), 108-113. https://doi.org/10.1080/08824098809359810

McGrew, K. S. (2005). The Cattell-Horn-Carroll Theory of Cognitive Abilities: Past, Present, and Future. In D. P. Flanagan & P. L. Harrison (Eds.), *Contemporary Intellectual Assessment: Theories, Tests, and Issues* (pp. 136–181). New York, NY: The Guilford Press.

McGuire, B. E., Morrison, T. G., Hermanns, N., Skovlund, S., Eldrup, E., Gagliardino, J., Kokoszka, A., Matthews, D., Pibernik-Okanović, M., Rodríguez-Saldaña, J., de Wit, M., & Snoek, F. J. (2010). Short-form measures of diabetes-related emotional distress: The problem areas in Diabetes Scale (PAID)-5 and PAID-1. *Diabetologia*, *53*(1), 66–69. https://doi.org/10.1007/s00125-009-1559-5

Mierzwa, E. (2018). The Relationship between Foreign Language Enjoyment and gender among secondary grammar school students. *The Journal of Education, Culture, and Society*, *1*(2), 117–135.

Muris, P., Schmidt, H., Merckelbach, H., & Schouten, E. (2001). Anxiety sensitivity in adolescents: Factor structure and relationships to trait anxiety and symptoms of anxiety disorders and depression. *Behaviour Research and Therapy*, *39*(1), 89-100. https://doi-org.proxy.bnl.lu/10.1016/S0005-7967(99)00179-5

Oh, M. J. T. (1997). *Beliefs about language learning and foreign language anxiety: A study of American university students learning Japanese* (Unpublished doctoral dissertation). University of Texas at Austin, Austin, TX.

Olaru, G., Witthöft, M., & Wilhelm, O. (2015). Methods matter: Testing competing models for designing short-scale Big-Five assessments. *Journal of Research in Personality, 59*, 56-68. https://doi.org/10.1016/j.jrp.2015.09.001

Onwuegbuzie, A. J., Bailey, P., & Daley, C. E. (1997, November). *Foreign Language Anxiety among College Students* [Conference presentation]. Annual Meeting of the Mid-South Educational Research Association, Memphis, TN.

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

Pavelescu, L. M. (2019). Motivation and emotion in the EFL learning experience of Romanian adolescent students: Two contrasting cases. *Studies in Second Language Learning and Teaching*, *9*(1), 55-82.

Pavelescu L. M. & Petrić B. (2018). Love and enjoyment in context: Four case studies of adolescent EFL learners. *Studies in Second Language Learning and Teaching*, *8*(1), 73-102.

Pawlak, M., Kruk, M., & Zawodniak, J. (2020a). Individual trajectories of boredom in learning English as a foreign language at the university level: insights from three students' self-reported experience. *Innovation in Language Learning and Teaching*. https://doi.org/10.1080/17501229.2020.1767108

Pawlak, M., Kruk, M., & Zawodniak, J. (2020b). Investigating individual trajectories in experiencing boredom in the language classroom: The case of 11 Polish students of English. *Language Teaching Research*, 1–20. https://doi.org/10.1177/1362168820914004

Phillips, E. M. (1992). The effects of language anxiety on students' oral test performance and attitudes. *Modern Language Journal*, *76*(1), 14-26.

Putnam, S. P., & Rothbart, M. K. (2006). Development of short and very short forms of the Children's Behavior Questionnaire. *Journal of Personality Assessment*, *87*(1), 102–112. https://doi.org/10.1207/s15327752jpa8701_09

Raborn, A. W., Leite, W. L. (2018). ShortForm: An R package to select scale short forms with the ant colony optimization algorithm. *Applied Psychological Measurement*, *42*(6), 516-517. https://doi.org/10.1177/0146621617752993

Rolstad, S., Adler, J., & Rydén, A. (2011). Response Burden and Questionnaire Length: Is Shorter Better? A Review and Meta-analysis. *Value in Health*, *14*(8), 1101–1108. https://doi.org/10.1016/j.jval.2011.06.003

Rosseel, Y. (2012). Lavaan: An R package for structural equation modeling. *Journal of Statistical Software*, *48*(2), 1-36. https://doi.org/10.18637/jss.v048.i02

Ryan, R. M., Connell, J. P., & Plant, R. W. (1990). Emotions in nondirected text learning. *Learning and Individual Differences*, *2*(1), 1-17. https://doi-org.proxy.bnl.lu/10.1016/1041-6080(90)90014-8

Saito, K., Dewaele, J.-M., Abe, M., & In'nami, Y. (2018). Motivation, Emotion, Learning Experience, and Second Language Comprehensibility Development in Classroom Settings: A Cross-Sectional and Longitudinal Study. *Language Learning*, *68*(3), 709–743. https://doi.org/10.1111/lang.12297

Shao, K., Nicholson, L., Kutuk, G., & Lei, F. (2020). Emotions and instructed language learning: Proposing a second language emotions and positive psychology model under the current wave of positive psychology in second language acquisition. *Frontiers in Psychology, 11*, 2142. https://doi: 10.3389/fpsyg.2020.02142

Shirvan, M. E., & Taherian, T. (2018). Longitudinal examination of university students' foreign language enjoyment and foreign language classroom anxiety in the course of general English: Latent growth curve modeling. *International Journal of Bilingual Education and Bilingualism*, 1–19. https://doi.org/10.1080/13670050.2018.1441804

Shirvan, M. E., & Talebzadeh, N. (2019). Tracing the Signature Dynamics of Foreign Language Classroom Anxiety and Foreign Language Enjoyment: A Retrodictive Qualitative Modeling. *Eurasian Journal of Applied Linguistics*, *6*(1), 23-44.

Smith, G. T., McCarthy, D. M., & Anderson, K. G. (2000). On the sins of short-form development. *Psychological Assessment*, *12*(1), 102–111. https://doi.org/10.1037/1040-3590.12.1.102

Stevanovic, D. (2014). Is the Quality of Life Enjoyment and Satisfaction Questionnaire-Short Form (Q-LES-Q-SF) a unidimensional or bidimensional instrument? *Quality of Life Research*, *23*(4), 1299–1300. https://doi.org/10.1007/s11136-013-0566-7

Tabachnick, B. G., & Fidell, L. S. (2001). *Using multivariate statistics*. Needham Heights, MA: Allyn and Bacon.

Wei, H., Gao, K., & Wang, W. (2019). Understanding the relationship between grit and foreign language performance among middle school students: The roles of foreign language

enjoyment and classroom environment. *Frontiers in Psychology*, *10*. https://doi.org/10.3389/fpsyg.2019.01508

Zaiţ, A., & Bertea, P. S. P. E. (2011). Methods for testing discriminant validity. *Management & Marketing Journal*, *9*(2), 217-224.

Zwick, W. R., & Velicer, W. F. (1986). Comparison of five rules for determining the number of components to retain. *Psychological Bulletin*, *99*(3), 432.