

Construction and validation of self-Esteem in the foreign language classroom scale (SEFLS)

PABLO A. CANO-JIMÉNEZ

FERNANDO D. RUBIO-ALCALÁ

Universidad de Huelva

Received: 21 January 2022 / Accepted: 15 June 2022

DOI: 10.30827/portalin.vi38.22447

ISSN paper edition: 1697-7467, ISSN digital edition: 2695-8244

ABSTRACT: This paper describes the development and standardization of a measure of self-esteem in the foreign language classroom. The reliability and validity of Self-esteem in the Foreign Language Classroom Scale (SEFLS) were determined for a sample of 252 Spanish native speakers studying a foreign language. The test reliability coefficient was $\omega = .92$. Confirmatory Factorial Analysis (CFA) yielded a 10-item three-dimension structure acceptable model fit. The three dimensions were language competence, skills performance and perceived social support. The SEFLS measure could be a useful tool for assessing students' self-esteem in the foreign language classroom. Implications, limitations, and directions for future organizational research are discussed.

Key words: Validation, Scale, Self-esteem, Self-concept, Foreign Language

Construcción y validación de una escala de autoestima en la clase de lengua extranjera (EACLE)

RESUMEN: Este artículo describe el desarrollo y la estandarización de una medida de autoestima en la clase de lengua extranjera. La confiabilidad y la validez de la Escala de Autoestima en la Clase de Lengua Extranjera (SEFLS; en inglés) fue determinada para una muestra de 252 hablantes nativos de español que estudian una lengua extranjera. El coeficiente de confiabilidad de la prueba fue $\omega = .92$. El Análisis Factorial Confirmatorio (AFC) arrojó un ajuste aceptable del modelo de estructura tridimensional de 10 ítems. Las tres dimensiones fueron competencia lingüística, desempeño de destrezas y apoyo social percibido. La medida SEFLS podría ser una herramienta útil para evaluar la autoestima de los estudiantes en las clases de idiomas. Se añaden las implicaciones, limitaciones y direcciones para una investigación futura.

Palabras clave: Validación, Escala, Autoestima, Autoconcepto, Lengua Extranjera

1. INTRODUCTION

Self-esteem, which is usually encapsulated within the affective domain (Arnold, 1999), has been shown to exert a positive correlation for language achievement in the Foreign Language Classroom (FLC) (Liu, 2012; Alrabai, 2017). Scales, questionnaires, and other

self-report tools are the most contested methods to measure the affective factors in this context (Rubio-Alcalá, 2014). However, to date, no statistically validated tool has been found that measures self-esteem in the specific situation of the FLC. Further, some studies have relied on general measures of self-esteem, for instance, using Rosenberg's scale (1965), which may not be suitable. In fact, learning a foreign language is a domain-specific experience (Arens et al., 2020) that requires distinct measurement (Mercer, 2011a). Therefore, it is of paramount importance for the scientific community to apply valid and tailor-made tools that can reach reliable results. Accordingly, we aim to inform about the construction and validation processes of a brand-new self-report tool that can serve as a spearhead to open up a new direction for interventions in the FLC.

2. SELF-ESTEEM AND THE PREVIOUS LITERATURE

Self-esteem has been utterly unexplored in the context of language learning. Apart from a couple of volumes dealing with self-esteem theory and practice (Rubio-Alcalá, 2007; Habrat, 2018), and a qualitative-based study and theory (Mercer, 2011b), the plethora of quantitative-based studies is reduced to just a few in the 20th century (e.g., Heyde-Parsons, 1979 and 1983; Anderson, 1982; Hassan, 1992).

As for the last two decades, Guban-Caisido's meta-analysis (2020) results in a list of only 20 studies. Despite the empirical evidence purported in those studies, it is still problematic the lack of use of specific measures or the use of tools that have not undergone statistical analysis for reliability and validation. Thus, some studies have used tools designed for global self-esteem: Coopersmith's Self-esteem Questionnaire, 1967; Hayati & Ostadian, 2008; Maleki & Mohammadi, 2009; Soureshjani & Naseri, 2011; Rosenberg's Self-esteem Scale, 1965; Zare & Riasati, 2012; Dev & Qiqieh, 2016; Mandokhail et al., 2018; Sorensen's Self-esteem Test, 2005; Koosha et al., 2011; and other studies have used self-designed, adapted, or indirect measures (anxiety, language attitude, language aptitude, etc.): Hassan, 1999; Pramita, 2012; Fahim & Rad, 2012; Takahashi et al., 2013; Tilfarlioglu & Delbesoglugil, 2014; Wullur, 2014; Basco & Han, 2016; Alrabai, 2017; Moriya, 2019; Satriani, 2019.

3. CONSTRUCTION OF THE SELF

Following Boateng et al.'s (2018), and Muñiz and Fonseca-Pedrero's (2019) suggestions to design and validate a scale, we devised the following stages and steps as is shown in Table 1.

Table 1. Steps in the validation of the measuring instrument

Stages	Steps
Item development	1. Conceptualization of self-concept and self-esteem
	2. Dimensionality, hierarchy, and structure of self-concept in the FLC
	3. Elaboration of items
Scale development	4. Inter-rater validation
	5. Scale edition
	6. Pilot study
	7. Test application
Scale estimation and layout	8. Psychometric properties
	9. Scoring and interpretation of scores
	10. Final version of the scale

Item development

3.1. Conceptualization of self-concept and self-esteem

We follow Dörnyei’s (2005) definition of self-esteem as the value or result of the evaluation of self-concept. Rubio-Alcalá (2017, p. 202) takes a philological perspective to describe both constructs and coincide with Dörnyei that “self-concept refers to an entity, dimension or, ultimately and rather obviously, a *concept*, while *to esteem* is actually to estimate or give a value, consequently leading to a measurable result”. Accordingly, both constructs are the sides of the same coin and are difficult to separate (Schumann, 1997; Swain, 2013). We advert to the reader that we have chosen the umbrella term *self-esteem* for operational purposes, as we will be dealing with a measurement activity. Thus, the scale will yield a result expressed in quantity, which will point to high, moderate, or low self-esteem of specific domains of the individual’s self-concept. Thus, when we deal with dimensions we will use the term *self-concept*, and when referring to the evaluations of the dimensions we will use *self-esteem*.

3.2. Dimensionality, hierarchy, and structure of self-concept in the FLC

Mercer (2011b, p. 335) claims that “findings illustrate how self-concept is perhaps best conceived of as a complex, multilayered, multidimensional network of interrelated self-beliefs”. According to Arens et al. (2020: 3) “multidimensionality means that self-concept consists of different domain-specific facets tapping several domains of an individual’s life

and experiences". Those domains are hierarchically structured on different generality levels, and the acceptance of the different interactions of the domains result in various models: a) Higher-order factor model; (b) Marsh/Shavelson model; (c) Nested Marsh/Shavelson model; (d) Bifactor-ESEM representation; and (e) First-order factor model (cf. Arens et al., 2020). The highest level of the hierarchy is called the *general* or *global* self-concept, which is assumed to be divided into academic and non-academic self-concept on a subordinate level, and academic self-concept is further separated into verbal and non-verbal, with the specific learning fields of the curriculum such as English as a foreign language self-concept, Spanish self-concept, History self-concept, Biology self-concept, etc. While Arens et al.'s (ibidem) analyses of the models point to the direction that general academic self-concept may display a cross-sectional representation of domain-specific self-concepts, evidence was found in all the models regarding the separation between math and verbal academic self-concepts. Moreover, correlations, for instance, between verbal subjects (foreign language, native language, or history) were higher than the correlation between math and history. Marsh (1986) also found a near-zero correlation between students' verbal and math self-concepts.

Faber (2012) proved that the native language (L1) and foreign language (L2) self-perceptions were independent, and the factors were uncorrelated. Furthermore, different self-esteem loads have been found for the different practices and skills in the language classroom. For instance, Maleki & Mohammadi (2009), Kalazandeh et al. (2013), Wullur (2014), and Mandokhail et al. (2018) concurred that a significant relationship between self-esteem and speaking skills is found, and also with reading comprehension (Bagheri & Faghih, 2012; Piran, 2014), writing (Fahim & Rad, 2012), or listening comprehension (Hayati & Ostadian, 2008). Moreover, in an analysis of the specific skill of speaking, Wullur (2014) found that self-esteem has a positive and strong correlation with grammar, a positive moderate correlation with pronunciation, vocabulary, and a positive and weak correlation with fluency.

Self-concept specificity can also be found in the interaction of the individual with time dynamicity and context: trait, state and situation-specific self-esteem. Trait self-esteem, also called global self-esteem, is a relatively stable disposition that results in oriented evaluations of the situations; for instance, individuals that tend to adopt positive views of themselves regardless of time and situation. State self-esteem, also called situational self-esteem, refers to everyday and unique assessments according to the situation; for instance, a person who might experience high self-esteem after someone's feedback, and then low self-esteem following a different piece of feedback. Situation-specific self-esteem, also called intermediate self-esteem, is produced in more context-specific domains such as sports, or learning specific subjects; for instance, a learner's self-esteem may be high in math and low in the foreign language. To create the SEFLS scale we depart from a consideration of the last type of self-esteem.

Finally, self-concept structure refers to the specific and most-likely events that are evaluated and result in a level of self-esteem. In general psychology, Mruk (2013) claims that self-esteem is the result of the evaluations of the individual's competence and worthiness. In the context of foreign language learning, Heyde-Parsons (1979 and 1983), Hassan (1999), Brown (1994), Mercer (2011b) and Rubio-Alcalá (2014) make clear that self-esteem is formed by the evaluations of the performances proper of this specific context. For the

purposes of the scale and based on the previous literature, we identified four dimensions: language competence, skills performance, interpersonal relations, and learning orientations.

Language competence refers to those dispositional aspects that allow learners to accomplish academic achievement. It consists of the self-perceptions that students produce in relation to their level of self-efficacy (performance and achievement based on their perceived intellectual or academic capacities). This dimension applies globally to the academic environment, especially self-efficacy, self-confidence, although aptitude for languages and proficiency level are added as particularities of learning a language.

Skills performance refers to how students perceive themselves when they carry out activities according to the use of different skills: speaking, reading, listening, writing. Although these can occur in an integrated way with communicative, grammar or vocabulary activities, we have separated each skill in the scale to operationalize the variables.

Interpersonal relationships allude to the perceptions that learners have regarding their exchanges and interactions with classmates or teachers. It includes the sense of belonging to the group and the perceived feedback. In the foreign language classroom, where communicative activities predominate, students have to participate orally and continually expose themselves to others and to the teacher, which generates interpersonal situations that are not necessarily generated during the learning of other subjects.

Orientation towards learning includes motivation, anxiety, attitude, beliefs and other situational, cognitive and affective states, which trigger a response (behaviour) of approach or rejection towards the subject and its learning. Congruently, when the state is one of approach (facilitated by positive states of motivation, interest, etc.) the alienation, effort and direction towards learning will be greater, as well as the resulting level of self-esteem. This dimension was removed in the scale development stage (see the proper section for a rationale).

In sum, our intention is to design a measuring instrument based on a theoretical basis and a model to which to ascribe. The general and substantive theories that base and define the theoretical background of the research must be clearly determined, and from this positioning, the construct is defined in operational terms so that it can be measured empirically (Soriano, 2015). According to Muñiz and Fonseca-Pedrero (2019), this allows, on the one hand, to delimit the construct and consider all its relevant dimensions and, on the other, to clearly identify its most representative behaviours. We have relied on the most significant literature of global and specific self-esteem (e.g., Rubio-Alcalá, 2007 and 2014; Mercer, 2011a; Habrat, 2018) to design a theoretical model of self-esteem in the specific context of the FLC (Rubio-Alcalá and Cano-Jiménez; sent for review).

3.3. Elaboration of items

Once conceptualization and other theoretical aspects were clarified, the next steps were deliberating and deciding the structure of the scale and the items (Fernández-Ballesteros, 2014), including the relevant aspects that form the items and the method of application of the instrument, as is shown in Table 2.

Table 2. Scale specifications

Instruments application mode	Structure of the items
- Test administration support: computerized software.	- Total number of items: the ones determined after psychometric calculations.
- Type of application: individual.	- Proportion of items for each of the dimensions identified in the definition of the construct: the ones determined after psychometric calculations and according to the literature.
- Item format: Likert-type.	- Type of answer: closed statement items.
- Distribution of the questionnaire: direct link.	- Rules for correcting and interpreting scores: score from 0 to 10, where 0 corresponds to totally disagree, and 10 to totally agree.
- Delivery of material: online.	
- Adaptation of vocabulary and accessibility to the population under study.	

The initial step for the construction of the items was to list items that would suit the theoretical model for external validity. Careful wording was made for the items to reflect as accurately as possible the behaviour of the dimension that was included in the model. Following Fernández-Ballesteros’ (2014) recommendations for procedures, we brainstormed items to provide a first draft with 111 items, which were qualified to account for heterogeneity, variability and representativeness (Martínez-Arias, 2005). In a second round, the set of items was refined and reduced to 59 items, which resulted in 41 items after a third review (7 language competence items; 15 skills performance items; 7 interpersonal relationship items; and 12 learning orientation items).

Scale development

3.4. Inter-rater validation

The draft was then sent to a group of experts to analyse content validity, and the clarity and intelligibility of the items according to the target population. The experts were 6 scholars whose professional, academic, and research experience were related to self-esteem. Table 3 lists their expertise.

Table 3. Inter-rater expertise

Expert 1	Research Methodology in Psychology
Expert 2	Research Methodology in Psychology
Expert 3	Psychology (self-esteem)
Expert 4	Psychology (self-esteem)
Expert 5	Research Methodology in Education
Expert 6	Methodology of Evaluation (with relevance in languages)

The experts acknowledged the objectives and theoretical positioning of the research (Soriano, 2015), and evaluated the following parameters: validity of the content (congruence with the scale and dimension); wording clarity; the adequacy of the language of each item with respect to the target population; and the biases that could influence responses. Additionally, space was provided for other possible observations, as well as a box to tick for modification or deletion of the items, as shown in Table 4.

Table 4. Inter-rater evaluation tool

Criteria to evaluate									
Item	Content validity (consistency)		Wording clarity		Language adequacy with respect to target population		Biases that could influence responses		Observations: If you wish to modify or delete, please indicate accordingly
	YES	NO	YES	NO	YES	NO	YES	NO	
Dimension									
Item X									

Interviews were also conducted with the experts to discuss and clarify their responses. The experts identified some items that could measure more than one specific variable and, therefore, were not valid. The language competence and skills performance dimensions were found quite robust. Regarding interpersonal relationships, the items in this block appeared to be measures of perceived social support, specifically the social support provided by classmates and the foreign language teacher. Perceived social support is likely to depend on self-concept, but it should also be a reflection of actual social support. In that sense, the experts claimed that they would not be pure indicators of self-concept. Finally, there was consensus that the block of items referring to orientation towards learning -motivation, attitude, anxiety and beliefs- were constructs or variables with their own tradition, which, although they might be related to self-concept, are different from it, and therefore, this dimension should be deleted. Consequently, the dimensions of our model are represented in Figure 1.

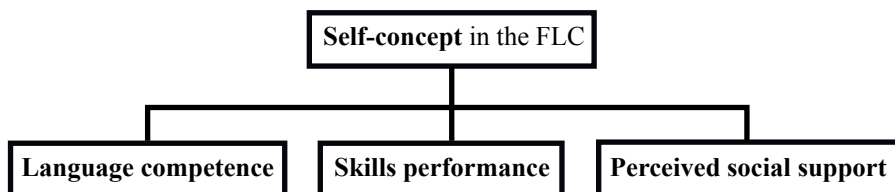


Figure 1. Dimensional model of self-concept in the FLC

The result of this stage came up with the first version of the scale with a list of 24 items.

3.5. Scale edition

The first version of the scale was presented. The items were arranged and ordered according to the dimensions. The layout of the tool was determined too, including information for the participants about the research team, the aims of the scale, a request to answer all the items with sincerity, and the instructions for completion.

We also added the numerical values assigned to the response options. The range was from 0 (totally disagree) to 10 (totally agree). This wide range would allow us to make more complex calculations if needed, like grouping values, for instance. The total value, or maximum number, was 240 (24 items x 10). Thus, the closer the total score was to 240 the higher the student's self-esteem. The assignment of items to each of the dimensions was: language competence: items 1, 2, 3, 4, 5, 6, 7, 8; skills performance: items 9, 10, 11, 12, 13, 14, 15, 16, 17, 18; and perceived social support: items 19, 20, 21, 22, 23, 24. 6 inverted items and the corresponding reverse codification were prepared (9, 12, 18, 21, 22, 24).

3.6. Pilot study

The purpose of the pilot study was to examine the general functioning of the scale in a sample of participants with similar characteristics to the target population. This phase was relevant since it allowed detecting and correcting possible errors, as well as testing a first application of the scale in the applied context (Muñiz & Fonseca-Pedrero, 2019). The instrument was administered under the same conditions with which would be applied to the target population (Soriano, 2015). 14 English as a foreign language students (all female, 20-25 years old, and enrolled in education studies) participated in the screening stage.

Qualtrics software was used to generate the scale and was distributed electronically by sending the link to the participants. The responses were received from April 17th to May 3rd, 2021. After collecting the responses, the data was downloaded and exported to SPSS software for analysis, which was performed using qualitative and quantitative methods. On a qualitative level, participants had open questions to comment on possible difficulties in understanding the items or other inadequacies, like length and time of completion. Results were positive towards the use and completion of the scale. On a quantitative level, the metric properties were examined, that is, the reliability calculations of the scale. The purpose was to be able to apply the scale in a standard way to a certain population (Santisteban, 1990). For this reason, a standardized index of consistency or precision was employed. Although Cronbach's Alpha Coefficient is the most widely used statistical method in the literature to obtain an estimate of the reliability and internal consistency in a test, we used the Omega Coefficient, as it reaches a more precise measure of reliability in one-time applications, as reported by Ventura-León and Caycho-Rodríguez (2017). This is because the Omega Coefficient allows to include factorial loads, and it is obtained with the sum of the standardized variables. Thus, it reflects the true level of reliability and does not depend on the number of items (McDonald, 1999). The Omega coefficient was .863. According to the literature, the reliability coefficients must range between .70 and .90 for the instrument to be considered adequate. Consequently, the 24 initial items were kept for the next analysis level.

3.7. Test application

The scale was addressed to adult foreign language students (18 years old minimum). Demographic variables were included for the participants to indicate age, sex, foreign language, and study institution. The scale was administered through the institutional emails and their social networks in May and June 2021. Qualtrics software was used to prepare and manage the responses. A first screening was made to remove from the sample those participants that catered for acquiescence (tendency to accept default responses in surveys regardless of their content), resulting in 252 valid respondents. The distribution of foreign languages was English $n= 173$ (68.7%), French $n= 52$ (20.6%), Portuguese $n= 12$ (4.8%), Italian $n= 9$ (3.6%), German $n= 4$ (1.6%), Japanese $n= 1$ (0.4%) and Russian $n= 1$ (0.4%).

The language students came from different training centres: 21% from universities, 53.5% from language schools; 12.3% from other educational institutions; and 11.9% unspecified. The distribution of sex was: $n= 193$ women (76.6%); and, $n= 59$ men (23.4%). The age of the participants ranged from 18 years to over 40, with a mean of 25 years old.

Scale estimation and layout

3.8. Psychometric properties

The main psychometric properties for validation in our scale are reliability and validity. These two concepts are closely related since the first necessary condition for a test to be valid is that it has an adequate degree of reliability. A test can be reliable and meet all the requirements, but not valid, that is, it does not measure what is intended. However, for a test to be valid, it must necessarily be reliable, since unreliable scores cannot, in any case, provide a basis from which a correct validation can be carried out (López-Mezquita, 2018).

We used Confirmatory Factor Analysis (CFA) to empirically confirm the conceptual structure of the scale and to know the role that each item played in the overall set of the structure. The factorial model of self-esteem established a structure of three dimensions (language competence; skills performance; and perceived social support), as each factor saturated only on the common dimension, and the factors correlated with each other on each dimension.

We also performed CFA to evaluate the degree of fitness of the theoretical model to the empirical reality. JASP (2020) was used to calculate several indexes: the comparative fit index (CFI); the non-normed fit index (NNFI), the Tucker Lewis index (TLI); the standardized root mean square residual (SRMR); the root mean square error of approximation (RMSEA); and the goodness of fit index (GFI). We found that the model represented by items 1, 3, 4, 5, 6, 8, 10, 11, 14, 15, 19, and 20 obtained adequate adjustments, which even improved by removing item 14 “I normally understand the audios”, and item 15 “I normally understand what I read”. The combination of items 1, 3, 4, 5, 6, 8, 10, 11, 19, 20 ($R^2 \geq .60$) showed the greatest metric guarantees and maximized the psychometric properties of the model, as can be seen in Table 5.

Table 5. Fit indexes of the structural models

Model	CFI	NNFI	TLI	SRMR	RMSEA	GFI	ω
Model with 24 items	.78	.75	.75	.1	.12	.72	.92
Model with 12 items	.96	.94	.94	.03	.08	.91	.93
Model with 10 items	.98	.97	.97	.02	.06	.95	.92

Maximum likelihood estimation. Robust parameter

CFI: ranges between 0 and 1, with the minimum value of .90 for the adjustment to be considered adequate

NNFI: minimum of .90 to be considered adequate; from .90 to .95 acceptable; and, >.95 optimal

TLI: <.90 generally inadequate fit; from .90 to .95 proper fit> .95 perfect fit

SRMR: <.05 good fit; and, .08 acceptable fit

RMSEA: <.06 best fit; .06 to .08 suitable fit; and> .10 poor fit between the model and the observed data

GFI: ranges between 0 and 1, considered adequate model upon .90

Ω : from .70 to .90 acceptable; and, >.90 perfect

Fit indices source: Boateng et al. (2018) and Martínez et al. (2012)

The 10-item version of SEFLS represented the most plausible and parsimonious combination of three interrelated dimensions. The factorial structure represented with all the related factors indirectly supposes the existence of a latent factor -the self-concept in the FLC- that allows us to explain, in turn, these relationships, as can be seen in Figure 2.

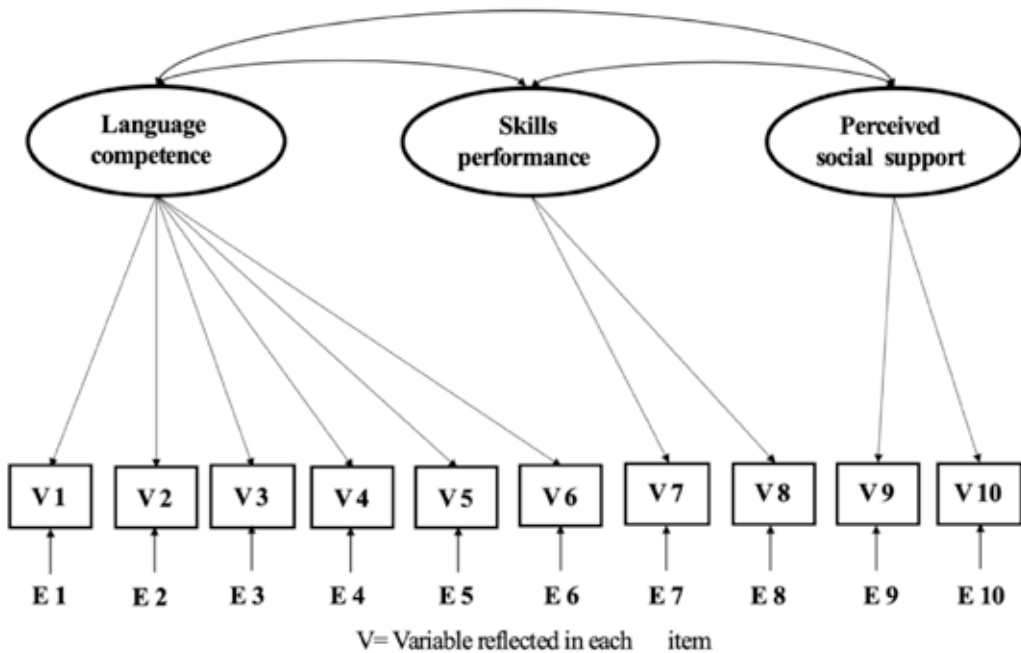


Figure 2. Structure of the correlated self-concept model for the FLC

3.9. Scoring and interpretation of scores

The total score of this scale is the sum of the scores of each of the items, which is used as an estimate of the level of self-esteem a respondent has with respect to the construct. The transformation of the scores acquires meaning when comparing them within a frame of reference, in this case, by means of the criterion referred to the level of self-esteem of the student in the FLC. Fernández-Ballesteros (2014) indicates that the scores obtained from scales referred to the criterion are those deliberately constructed to produce measurements that are directly interpretable in terms of a well-defined standard of performance. Therefore, this criterion of achievement would be the frame of reference against which the score (information) collected from the student must be interpreted. For example, in our scale the total sum of the scores is the student's total score, and, therefore, this would be the criterion with which to compare the real or direct score. Thus, if the total score of SEFLS is 100 points, 90 points is assumed to be a very high level of self-esteem while 10 points is very low. Accordingly, the higher the score, the higher self-esteem that the respondent experiences when learning a foreign language. In our scale, and after considering previous scales, five cut-off points were established: 0-20% = very low self-esteem; 21-40% = low self-esteem; 41-60% = moderate self-esteem; 61-80% = high self-esteem; and, 81-100% = very high self-esteem.

3.10. Final version of the scale

The validation process resulted in a scale formed by 10 items, as shown in Table 6 in its original language version. The language competence dimension comprised items 1-6, the skills performance dimension was formed by items 7 and 8, and items 9 and 10 composed the perceived social support dimension. The original items in Spanish have been kept, as they are the ones that have been validated. Annex 1 shows the scale translated into English.

Table 6. Original version of SEFLS (Self-esteem in the Foreign Language Classroom Scale)

Muchas gracias por tu participación en rellenar la Escala sobre Autoestima en la Clase de Lengua Extranjera. A continuación, se te presentan una serie de afirmaciones sobre las que te pedimos que expreses tu grado de acuerdo. Recuerda que no existen respuestas correctas ni incorrectas, solamente intenta responder con sinceridad. Esta escala es totalmente ANÓNIMA

Ítem	En desacuerdo <----->De acuerdo									
	1	2	3	4	5	6	7	8	9	10
En la clase de lengua extranjera...										
1	Generalmente saco buenas notas									
2	En términos generales creo que se me da bien el idioma									
3	Me considero como uno/a de los/las mejores aprendiendo el idioma									
4	Me resulta fácil aprender el idioma									
5	En general mi nivel es bueno									
6	Creo que tengo buenas cualidades para aprender el idioma									
7	Pronuncio bien									
8	Se me dan bien las actividades orales									
9	La mayoría de compañeros/as me admiten como parte de la clase									
10	Me hace sentir bien estar con mis compañeros/as									
Baremo de corrección: 0-20 puntos= muy baja autoestima; 21-40 puntos= baja autoestima; 41-60 puntos= autoestima media; 61-80 puntos= buena autoestima; y, 81-100 puntos= muy buena autoestima.										

4. CONCLUSIONS

The construction and validation of a scale is a dynamic and open process, consisting of a set of steps aimed at providing the scores obtained from the scale with a coherent theoretical interpretation in relation to a well-identified and delimited context of use. In this sense, the study on self-esteem in the FLC is initiated just before the construction of the instrument, and they guide and accompany its development and validity throughout it, ensuring that the interpretations are supported by sufficient accumulation of evidence that confers the necessary metric and theoretical quality to the measuring instrument. Thus, the validation of an instrument is the process of accumulating evidence for the interpretation of the scores in the context in which it is applied (Martínez-Arias, 2005). Therefore, the object of validation is not the instrument, but the interpretation of its scores in relation to a specific objective or use. In our case, designing and validating an instrument that measures the construct of the self-concept in the FLC means, on the one hand, expanding the set of existing measurement instruments in psychology and, on the other, thriving on the understanding of this complex and multidimensional phenomenon (Guzmán et al., 2011).

Governed by ethical research procedures, the researchers affirm that the results presented are free from the outcome and confirmatory biases. The first step has been to conform a model starting from an explicit definition and the theoretical foundation in which the construct is enclosed. Self-esteem in the FLC has been described as a situation-specific experience that needs distinct treatment, and, accordingly, particular tools for its measurement. Yeung & Wong (2004) found out that self-esteem is specific for even different languages (namely, Mandarin, Cantonese, and English). Further, Lau, Yeung, Jin & Low (1999) evidenced that skill-specific self-perceptions of speaking, reading, writing, and global English, were different constructs, although the relations between the facets could form a higher-order construct or global English self-concept. A profound study of the literature has resulted in the identification of four main dimensions that can play a critical role in the learner's self-esteem measurement: language competence, skills performance, perceived social support, and learning orientation. These dimensions were considered from different angles: 1) modality: cognitive, affective and behavioural; 2) stimulus: internal or external; 3) specificity: global/trait, state, specific/task; 4) domain: factual, ideal, normative; 5) point of view: own, of others. Inter-raters' indications advised to remove the learning orientation dimension, as the factors encapsulated (anxiety, motivation, beliefs, etc.) would be of a separate entity. The model then emerged with a structure of three dimensions: language competence, skills performance, and perceived social support. After the pilot study, we performed statistical operations that resulted in a reliability coefficient of $\omega = .92$. Confirmatory Factorial Analysis (CFA) has yielded a 10-item three-dimension structure acceptable model fit.

Despite its contributions, this study has certain limitations. Regarding the sample, the study was addressed to adult students, and we do not have evidence of how SEFLS would work for primary, secondary, or older adult students. The scope also should be limited to foreign language students and the foreign language classroom, that is, monolingual learners that are learning a language that is not used in that context, and in educational settings, respectively. Further, the version of SEFLS has been done in Spanish, and other versions are needed in other languages.

Regarding the methodology, although some items were reversed, the validation process determined that the items that should remain in the final version of the scale were all positive. Although this might lead to acquiescence bias, we presume it will not happen, as the scale has only 10 items and the wording is very simple.

Future research is needed to replicate SEFLS in other contexts with different samplings and languages. Furthermore, general or global measures of self-esteem (e.g., Rosenberg, 1965) should be studied to establish what types of correlations exist with SEFLS.

In short, we hope that our work will contribute significantly to the field of research on self-concept and self-esteem in the FLC, and that researchers and language practitioners use SEFLS to have a better understanding of the phenomenon to cater to better educational practices.

5. REFERENCES

- Alrabai, F. (2017). Self-esteem of Saudi learners and its relationship to their achievement in English as a foreign language. *English Linguistics Research*, 6(4), 1-12. <https://doi.org/10.5430/elr.v6n4p1>
- Anderson, P.L. (1982). Self-esteem in the foreign language: A preliminary investigation. *Foreign Language Annals*, 5(2), 109-114.
- Arens, A.K., Jansen, M., Preckel, F., Schmidt, I., & Brunner, M. (2020). The structure of academic self-concept: A methodological review and empirical illustration of central models. *Review of Educational Research*, 91(1), 34-72. <https://doi.org/10.3102/0034654320972186>
- Arnold, J. (1999). *Affect in language learning* (1st ed.). Cambridge University Press.
- Bagheri, M.S., & Faghieh, M. (2012). The relationship between self-esteem, personality type and reading comprehension of Iranian EFL students. *Theory and Practice in Language Studies*, 2(8), 1641-1650. <https://doi.org/10.4304/tpls.2.8.1641-1650>
- Basco, L.M., & Han, S. (2016). Self-esteem, motivation, and anxiety of Korean university students. *Journal of Language Teaching and Research*, 7(6), 1069-1078. <https://doi.org/10.17507/jltr.0706.02>
- Boateng, G.O., Neilands, T.B., Frongillo, E.A., Melgar-Quíñonez, H. R., & Young, S.L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6(146), 1-18. <https://doi.org/10.3389/fpubh.2018.00149>
- Brown, H.D. (1994). *Principles of language learning and teaching* (3rd ed.). Prentice Hall Regents.
- Coopersmith, S. (1967). *The antecedents of self-esteem*. Freeman.
- Dev, S., & Qiqieh, S. (2016). The relationship between English language proficiency, academic achievement and self-esteem of non-native-English-speaking students. *International Education Studies*, 9(5), 147-155. <https://doi.org/10.5539/ies.v9n5p147>
- Dörnyei, Z. (2005). *The psychology of the language learner*. Routledge. <https://doi.org/10.4324/9781410613349>
- Faber, P. (2012). *A cognitive linguistics view of terminology and specialized language*. De Gruyter Mouton.
- Fahim, M., & Rad, S.K. (2012). The relationship between self-esteem and paragraph writing of Iranian EFL learners. *Psychology (Irvine, Calif.)*, 3(1), 24-29. <https://doi.org/10.4236/psych.2012.31004>
- Fernández-Ballesteros, R. (2014). *Evaluación psicológica* (1st ed.). Ediciones Pirámide.

- Guban-Caisido, D.A.D. (2020). Self-esteem and language learning: Empirical evidence from the past two decades. *Teaching and Learning English in Multicultural Contexts (TLEMC)*, 4(2), 95-106.
- Guzmán, F.J.B., Fermán, I.A.T., Gómez, C.B., Torres, A.A.B., & Calderón, M.M. (2011). Diseño de un instrumento para evaluar autoestima en jóvenes. *Medicina Salud y Sociedad*, 1(2), 1-17.
- Habrat, A. (2018). *The role of self-esteem in foreign language learning and teaching*. Springer.
- Hassan, B.A. (1992). Foreign language self-esteem and the oral communicative ability of university students. *Journal of the Faculty of Education - Mansoura University*, 19, 1-19.
- Hassan, B.A. (1999). The relationship of writing apprehension and self-esteem to the writing quality and quantity of EFL university students. *Journal of the Faculty of Education - Mansoura University*, 39, 1-36.
- Hayati, A.M., & Ostadian, M. (2008). The relationship between self-esteem and listening comprehension of EFL students. *Published Research, Faculty of Letters and Humanities, Shahid Chamran University of Ahvaz, Iran*, 1-13.
- Heyde-Parsons, A. (1979). *The Relationship between self-esteem and the oral production of a second language* [Doctoral Dissertation]. University of Michigan, Ann Arbor, United States.
- Heyde-Parsons, A. (1983). The relationship between self-esteem and the acquisition of French. *Second Language Acquisition Studies*, 175-187.
- JASP Team. (2020). JASP (Version 0.14.1) [Computer software].
- Kalazandeh, G.A., Mahnegar, F., Hassannejad, E., & Bakhtiarvand, M. (2013). The influence of EFL students' self-esteem on their speaking skills. *The International Journal of Language Learning and Applied Linguistics World*, 2(2), 77-84.
- Koosha, B., Ketabi, S., & Kassaian, Z. (2011). The effects of self-esteem, age and gender on the speaking skills of intermediate university EFL learners. *Theory and Practice in Language Studies*, 1(10), 1328-1337. <https://doi.org/10.4304/tpls.1.10.1328-1337>
- Lau, I.C., Yeung, A.S., Jin, P., & Low, R. (1999). Toward a hierarchical, multidimensional English self-concept. *Journal of Educational Psychology*, 91(4), 747-755. <https://doi.org/10.1037/0022-0663.91.4.747>
- Liu, M. (2012). Predicting effects of personality traits, self-esteem, language class risk-taking and sociability on Chinese university EFL learners' performance in English. *Journal of Second Language Teaching & Research*, 1(1), 30-57. <https://doi.org/10.5420/jsltr.01.01.3318>
- López-Mezquita, M.T. (2018). *La evaluación de la competencia léxica: Test de vocabulario, su fiabilidad y validez*. [Unpublished Doctoral Dissertation]. University of Granada, Spain. <https://digibug.ugr.es/handle/10481/4571>
- Maleki, N., & Mohammadi, A. (2009). Self-esteem and anxiety of Iranian EFL learners in oral communication tasks. *Journal of Linguistic Studies*, 2(2), 47-54.
- Mandokhail, S., Khan, F.R., & Malghani, M. (2018). Impact of ESL learners' self-esteem on their oral proficiency. *International Journal of English Linguistics*, 8(3), 210-222. <https://doi.org/10.5539/ijel.v8n3p210>
- Marsh, H.W. (1986). Verbal and math self-concepts: an internal external frame of reference model. *American Educational Research Journal*, 23(1), 129-149. <https://doi.org/10.3102/00028312023001129>
- Martínez, E.R., García-Alandete, J., Sellés, P., Bernabé, G., & Soucase, B. (2012). Análisis factorial confirmatorio de los principales modelos propuestos para el Purpose-In-Life test en una muestra de universitarios españoles. *Acta Colombiana De Psicología*, 15(1), 67-76.
- Martínez-Arias, M.R. (2005). *Psicometría: Teoría de los tests psicológicos y educativos*. Síntesis.
- McDonald, R.P. (1999). *Test theory. A unified treatment* (1st ed.). Psychology Press. <https://doi.org/10.4324/9781410601087>

- Mercer, S. (2011a). Language learner self-concept: Complexity, continuity and change. *System*, 39(3), 335-346. <https://doi.org/10.1016/j.system.2011.07.006>
- Mercer, S. (2011b). *Towards an understanding of language learner self-concept* (1st ed.). Springer Science & Business Media.
- Moriya, R. (2019). In the midst of emotion and identity: Investigating trajectories of learners' self-esteem from psychological and sociocultural perspectives. *Learner Development Journal*, 1(3), 8-25.
- Mruk, C.J. (2013). Defining self-esteem as a relationship between competence and worthiness: How a two-factor approach integrates the cognitive and affective dimensions of self-esteem. *Polish Psychological Bulletin*, 44(2), 157-164. <https://doi.org/10.2478/ppb-2013-0018>
- Muñiz, J., & Fonseca-Pedrero, E. (2019). Diez pasos para la construcción de un test. *Psicothema*, 31(1), 7-16. <https://doi.org/10.7334/psicothema2018.291>
- Piran, N.A. (2014). The relationship between self-concept, self-efficacy, self-esteem and reading comprehension achievement: Evidence from Iranian EFL learners. *International Journal of Social Sciences and Education*, 5(1), 58-66.
- Pramita, G.A.P. (2012). The contribution of self-esteem and language learning strategies to the students' English proficiency for second-year students of SMA Negeri 7 Denpasar. *Journal Penelitian*, 1(1), 10-13.
- Rosenberg, M. (1965). *Society and the adolescent self-image*. Princeton University Press.
- Rubio-Alcalá, F.D. (2007). *Self-esteem and foreign language learning*. Cambridge Scholars Publishing.
- Rubio-Alcalá, F.D. (2014). Self-esteem and self-concept in foreign language learning. In S. Mercer & M. Williams (Eds.), *Multiple perspectives on the self in SLA* (pp. 41-58). Multilingual Matters.
- Rubio-Alcalá, F.D. (2017). The links between self-esteem and language anxiety and implications for the classroom. In C. Gkonou, M. Daubney & J. M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 200-218). Multilingual Matters.
- Santisteban, C. (1990). *Psicometría. Teoría y práctica en la construcción de tests*. Norma.
- Satriani, M. (2019). The correlation between self-esteem and speaking performance in Indonesia. *Teaching and Learning English in Multicultural Contexts (TLEMC)*, 3(1), 8-14.
- Schumann, J.H. (1997). *The neurobiology of affect in language*. Malden, Mass: Blackwell.
- Sorensen, M. (2005). *The self-esteem institute self-esteem questionnaire*. <https://pdf4pro.com/view/the-self-esteem-institute-self-esteem-questionnaire-269c8f.html>
- Soriano, A.M. (2015). Diseño y validación de instrumentos de medición. *Diá-Logos*, (14), 19-40. <https://doi.org/10.5377/dialogos.v0i14.2202>
- Soureshjani, K.H., & Naseri, N. (2011). An investigation into the relationship between self-esteem, proficiency level, and the reading ability of Iranian EFL language learners. *Journal of Language Teaching and Research*, 2(6), 1312-1319. <https://doi.org/10.4304/jltr.2.6.1312-1319>
- Swain, M. (2013). The inseparability of cognition and emotion in second language learning. *Language Teaching*, 46(2), 195-207.
- Takahashi, A., Takahashi, H., 高橋, 歩., & 高橋, 秀. (2013). Learners' self-esteem and its relationships with motivation for learning English, self-perceived and actual English proficiency. *新潟大学言語文化研究*, 18, 1-12.
- Tilfarlioglu, F.Y., & Delbesoglugil, A.B.Ö. (2014). Questioning academic success through self-regulation, self-esteem and attitude in foreign language learning (a case study). *Theory and Practice in Language Studies*, 4(11), 2223-2230. <https://doi.org/10.4304/tpls.4.11.2223-2230>
- Ventura-León, J.L., & Caycho-Rodríguez, T. (2017). El coeficiente omega: Un método alternativo

para la estimación de la confiabilidad. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud*, 15(1), 625-627.

Wullur, B.G. (2014). Relationship between self-esteem and speaking skill of SMA students. *Teaching and Learning English in Multicultural Contexts (TLEMC)*, 3(1), 8-14.

Yeung, A.S., & Wong, E.K.P. (2004). Domain specificity of trilingual teachers' verbal self-concepts. *Journal of Educational Psychology*, 96(2), 360-368. <https://doi.org/10.1037/0022-0663.96.2.360>

Zare, P., & Riasati, M.J. (2012). The relationship between language learning anxiety, self-esteem, and academic level among Iranian EFL learners. *Pertanika Journal of Social Sciences and Humanities*, 20(1), 219-225.

6. APPENDIX

Annex 1: SEFLS (Self-esteem in the Foreign Language Classroom Scale): Items translated into English (our translation)

Thank you very much for filling out Self-Esteem in the Foreign Language Classroom Scale (SE-FLS). You will read a series of statements to express your degree of agreement. Remember that there are no right or wrong answers, just try to answer honestly. This scale is ANONYMOUS

Item	disagree <----->agree									
	1	2	3	4	5	6	7	8	9	10
<i>In the foreign language classroom...</i>										
1	I generally get good grades									
2	In general terms, I think I am good at the language									
3	I consider myself as one of the best learning the language									
4	I find it easy to learn the language									
5	In general my level is good									
6	I think I have good qualities to learn the language									
7	I pronounce well									
8	I am good at oral activities									
9	Most classmates accept me as part of the class									
10	I feel good with my classmates									

Correction scale: 0-20 points = very low self-esteem; 21-40 points = low self-esteem; 41-60 points = moderate self-esteem; 61-80 points = high self-esteem; and, 81-100 points = very high self-esteem.