

Gender differences in foreign language classroom anxiety: Results of a meta-analysis

Katalin Piniel

Eötvös Loránd University, Budapest, Hungary

<https://orcid.org/0000-0001-9225-3301>

brozik-piniel.katalin@btk.elte.hu

Anna Zólyomi

Eötvös Loránd University, Budapest, Hungary

<https://orcid.org/0000-0002-9280-5775>

zolyomi.anna@btk.elte.hu

Abstract

Exploring language learners' anxiety is not a neglected area of inquiry in applied linguistics research, which can primarily be attributed to the publication of the *Foreign Language Classroom Anxiety Scale* (FLCAS), an influential instrument developed by Horwitz et al. (1986) to measure language anxiety. An ever-growing array of studies has employed the FLCAS and analyzed the underlying relationship between the focal construct and foreign language achievement, various individual difference variables and a variety of demographic variables, such as learning experiences, age, and gender. Despite the considerable number of publications, studies focusing on biographical variables and language anxiety have not been conclusive. The aim of the present meta-analysis is to analyze 48 studies that employed the FLCAS to look at the potential gender differences with respect to language anxiety. Although there is great variation in the methodological and reporting practices in the studies included, and findings show a tendency for females to experience higher foreign language anxiety, gender-related differences are not statistically significant. The results of moderator analyses showed that neither age nor target language, regional context, or, in the case of university students, their majors, influence this relationship.

Keywords: language anxiety; gender; Foreign Language Classroom Anxiety Scale; meta-analysis

1. Introduction

Foreign language anxiety has been one of the most perplexing individual variables in language learning, and as such it has been the topic of abundant research since the 1970s. Research interest has gained momentum after the publication of the *Foreign Language Classroom Anxiety Scale* (FLCAS) developed by Horwitz et al. (1986), a tool that was designed to measure language learners' levels of foreign language anxiety in the classroom context, with an emphasis on speaking (X. Zhang, 2019). An increasing number of studies have used the FLCAS to uncover the relationship between anxiety and other individual differences, such as willingness to communicate, foreign language achievement and proficiency, self-efficacy beliefs, and demographic variables, such as experiences, age, and gender. Nonetheless, very few straightforward conclusions have been drawn about these learner variables and their link to language anxiety. One of the key issues that remains to be resolved is the role of gender (Botes et al., 2020). In the present study, our aim was to investigate the relationship of language anxiety and gender by conducting a meta-analysis of already published works that have used the FLCAS and also looked at the gender of the participants. The rationale behind opting for a meta-analytic approach was that by scrutinizing existing empirical findings, it enables the researcher to draw overarching conclusions concerning a given research problem, in the present case about whether males or females tend to experience higher levels of anxiety. In what follows, we will provide a brief overview of language anxiety research, a description of the FLCAS and a narrative summary of studies on language anxiety and gender, and to justify our method of research, we will also refer to meta-analytical studies on language anxiety. Then, the methods of our meta-analysis will be described, followed by the results and discussion of our findings.

2. Literature review

2.1. Overview of foreign language anxiety research

MacIntyre (2017) synthesized literature on language anxiety along the lines of three approaches that chronologically follow one another: the confounded phase, the specialized approach, and the dynamic approach. The first two phases provide the theoretical and empirical data for our research synthesis; therefore, we will briefly summarize those phases here. That is not to say, however, that the third, dynamic phase should be neglected in terms of a concise narrative literature review on language anxiety but rather that publications subscribing to a dynamic perspective would merit a systematic synthesis of their

own due to the special nature of their approach. For this reason, we will not consider them in this paper.

According to MacIntyre (2017), the beginnings of language anxiety research can be characterized by what he called a confounded phase, where “the ideas about anxiety and their effect on language learning were adopted from a mixture of various sources without detailed consideration of the meaning of the anxiety concept for language learners” (p. 11), leading to confusion about the relationship and effect of anxiety on language learning. Mainly the works by Scovel (1978) and Kleinmann (1977), who suggested that anxiety, a construct adapted from psychology, is a quite diverse phenomenon, with complex influences on language learning, are cited from this period. It was during this era of research that scholars distinguished between debilitating and facilitating anxiety as well as trait and state anxiety. Drawing on these two lines of thought, MacIntyre (2017) claimed that the trait-state divide (Spielberger, 1966, 1983) provided more fruitful ground for applied linguists to pursue research on language anxiety. Indeed, the definition of the construct that anxiety researchers fall back on in second language acquisition studies also comes from Spielberger (1983), according to whom anxiety is “the subjective feeling of tension, apprehension, nervousness, and worry associated with an arousal of the autonomic nervous system” (p. 1), which is “a disproportionately intense reaction” to stress (Levitt, 1980, p. 30). Trait anxiety is thought of as a personality characteristic, while state anxiety is a momentary experience of inhibition (Eysenck, 1979). Once the event is appraised as potentially threatening, the person may experience state anxiety.

The end of the confounding phase and the beginning of the specialized approach in language anxiety research (MacIntyre, 2017) is marked by the inclusion of the language anxiety construct in the socio-educational model of language learning (MacIntyre & Gardner, 1991) and Horwitz et al.'s (1986) work on foreign language classroom anxiety (FLCA). Horwitz and her colleagues (1986) defined FLCA as “a distinct complex of self-perceptions, beliefs, feelings, and behaviors related to classroom language learning arising from the uniqueness of the language learning process” (p. 31). Thus, language anxiety, foreign language anxiety or FLCA (generally used interchangeably in the literature) have come to be viewed as situation-specific anxiety, comprising cumulative, repeated, momentary experiences of anxiety (state anxiety) particularly linked to the context of language learning (Dewaele, 2002, 2005; Horwitz et al., 1986; MacIntyre, 1999; MacIntyre & Gardner, 1989, 1991).

One of the main outcomes of the specialized approach phase has been the development and widespread use of the FLCAS (Horwitz et al., 1986), which has been adapted across the globe to investigate the relationship between learners' language anxiety and achievement as well as other individual difference variables

as well as more general background learner characteristics such as personality, level of proficiency, age, and gender. Since a considerable number of studies have been published in this phase using the FLCAS as a data collection tool, the present meta-analysis focuses on those that have probed into the relationship between language anxiety and gender. In the following sections, we will turn to describing the FLCAS in more detail as well as summarizing some of the key studies that fall within the specialized approach and look at the relationship between language anxiety and gender.

2.2. Measuring foreign language classroom anxiety

Although various instruments have been used in the literature for measuring language anxiety, to date, the FLCAS, developed by Horwitz et al. (1986), has probably been the most widely adapted tool across a large variety of language learning contexts. The questionnaire comprises 33 5-point Likert-scale items, with the anchors of *strongly disagree* (1) and *strongly agree* (5). Although Horwitz (2017) explicitly stated that the questionnaire was not originally intended to comprise the subscales of communication apprehension, fear of negative evaluation, and test anxiety, many studies since the publication of the FLCAS have referred to these constructs. Generally speaking, communication apprehension refers to the inhibition experienced when conversing in the foreign language, fear of negative evaluation has to do with potentially being negatively judged by the instructor or peers, and test anxiety refers to the apprehension associated with classroom assessment of learners' foreign language performance. The FLCAS includes nine negatively worded items (items 2, 5, 8, 11, 14, 18, 22, 28, 32), which are normally reversed before calculating an overall score to describe respondents' anxiety levels. Horwitz and her colleagues (Horwitz et al., 1986) have demonstrated the reliability of the questionnaire, reporting Cronbach's alpha ($\alpha = .93$ in their 1986 study) and a correlation coefficient ($r = .83$, $p < .001$) based on scores obtained from a test and a re-test using the same tool on the same sample eight weeks apart ($N = 78$).

The FLCAS has been used in many applied linguistics studies; thus, in the past few decades, quite a lot of information has become available on language anxiety and its link to other learner characteristics. However, to date, there has been a limited number of meta-analytic studies synthesizing the results of this research in a more systematic manner, as opposed to the abundant number of narrative literature reviews that have been published as part of empirical papers or as theoretical overviews summarizing work that has been done on FLCA. Therefore, the aim of this paper is to present a meta-analytic study involving the empirical findings generated by research using the full version of the FLCAS as a

tool to collect data on language learners' foreign language anxiety. An additional benefit of limiting the scope of our meta-analysis to studies on FLCA as measured by the FLCAS is that in this way we can avoid drawing on the "associations among imperfect measures of these constructs reported in primary studies" (Card, 2012, p. 147) and reduce the necessity to correct for such artifacts.

2.3. Foreign language anxiety and gender

As already mentioned, a growing body of research has examined whether learner characteristics have an impact on foreign language learning anxiety; however, the results tend to be mixed based on gender differences across various contexts including second and foreign language learning contexts. It must be noted that throughout the present study, we refer to gender as the binary-coded biographical variable (i.e., male/female), following the positivist interpretation of the construct as appearing in quantitative studies on participants' gender and language anxiety. Specifically, a wealth of studies have found no significant gender-related differences with respect to foreign language anxiety (e.g., Aida, 1994; Dewaele, 2007, 2013a; Dewaele et al., 2008; Matsuda & Gobel, 2004; Woodrow, 2006; Yan, 1998), whereas other research endeavors have come to the conclusion that females manifest higher levels of anxiety (e.g., Arnaiz & Guillén, 2012; Briesmaster & Briesmaster-Paredes, 2015; Cheng, 2002; Dewaele et al., 2016; Donovan & MacIntyre, 2004; Öztürk & Gürbüz, 2013; Park & French, 2013). The repertoire of contradictory evidence is further endorsed by Campbell and Shaw (1994), Kitano (2001), Mejías et al. (1991), and L. J. Zhang (2001) because based on their results, males experienced higher levels of anxiety. Another intriguing aspect of this issue is when conflicting results seem to be apparent even within one specific study. For example, Elkhafaifi (2005) found no significant gender differences in listening anxiety but found significant differences in learning anxiety with females having a higher mean as compared to males. Similarly, Campbell's (1999) results indicated no significant gender differences in anxiety levels, but after two weeks of instruction males reported higher levels of anxiety. Dewaele (2013b) divided the participants of his study into two groups, and, based on the results of the first group, female students had higher anxiety scores in their third language (L3), but not in their second language (L2) and fourth language (L4). The second group, however, showed gender-related differences related to their L3 as well as L4.

It is due to the contradicting evidence concerning the link between gender and language anxiety that a meta-analysis seems to be indispensable in this domain. What is agreed upon by most researchers, however, is the undoubted complexity of foreign language anxiety. As has been concluded by Park (2013), among others, gender, language anxiety, and L2 performance exhibit an intricate

relationship with one another. Thus, the rationale for analyzing gender differences concerning foreign language anxiety lies in its multifaceted nature since “proficiency might not be the only or even the primary factor that determines the rise or decline of language anxiety” (Cheng, 2002, p. 653). In addition, the inconclusive evidence on the relationship between gender and language anxiety suggests that other variables, such as age and the learning context (including the target language, regional context and the major; cf. Horwitz, 2017) may play a determinant role in explaining the variability in the link between gender and language anxiety. For this reason, the modulating influence of these biographical characteristics should also be investigated in a meta-analysis on language anxiety and gender.

2.4. Meta-analyses on language anxiety

In order to be able to identify trends in empirical research findings, there has been a call for some years now to conduct more systematic syntheses of research in applied linguistics (Li et al., 2012; Plonsky & Oswald, 2012). Norris and Ortega (2006) in their pioneering work refer to systematic reviews as research syntheses. They make the following comment in this respect: “Research synthesis pursues systematic (i.e., exhaustive, trustworthy, and replicable) understandings of the state of knowledge that has accumulated about a given problem across primary research studies” (p. xi). According to the authors, such research can take on a variety of forms, including qualitative and quantitative research syntheses, depending on the methods and the field of study whose results are being synthesized. Since numerous papers have been published thus far on foreign language anxiety where quantitative data was gathered, a few publications have already followed suit, and presented syntheses of quantitative studies using quantitative methods. These research syntheses have been labeled as meta-analyses.

One such meta-analysis, conducted by Teimouri et al. (2019), involved 97 studies and focused on the link between language anxiety and achievement. The researchers found an overall moderate negative correlation between these two factors. The researchers also looked at whether the effect sizes differed in the case of a variety of moderator variables, such as language achievement, level of education, target languages, and types of anxiety. They found that the negative link between L2 anxiety and achievement is influenced by these variables. X. Zhang (2019) also conducted a research synthesis on language anxiety and performance; however, the author focused on performance measures that were not based on participants’ self-perceptions but rather on language course grades and language test scores. Apart from the correlation between language anxiety and performance, X. Zhang (2019) also looked at the moderating effect of other variables, such as the type of anxiety, proficiency, age, and L1-L2 distance.

This study found a moderate negative correlation between performance and language anxiety, with anxiety type, age, lexical similarity of L1 and L2 but not learners' proficiency levels, moderating this relationship. Finally, a third study, conducted by Botes et al. (2020), also investigated the link between language anxiety and achievement but considered only those studies in their meta-analysis that used the FLCAS or a translated/adapted version of it. Similarly to the previous two meta-analyses, the authors found negative correlations between achievement and FLCA. As for the moderators, neither age, nor female proportion, nor institution type were found to modulate the link between language anxiety and achievement. Nonetheless, the authors acknowledge as a limitation the fact that they have included the effect size from studies employing various adaptations (shortened versions) of the measurement tool (FLCAS), which may have influenced the outcome of the moderator analyses.

Despite the above papers presenting research syntheses, there are still very few publications that have attempted to summarize the trends emerging from the results of quantitative studies on foreign language anxiety, more specifically, what research results show us in terms of the link between language anxiety and gender. In order to fill this gap, we conducted a meta-analysis to investigate the possible connection between these two variables based on the results of quantitative studies that employed the full version of the FLCAS as a data collection instrument. Based on these aims and to fill the niche pertaining to the lack of meta-analyses concentrating on the possible relationship between language anxiety and gender, the research questions that guided our study were the following:

1. What are the methodological and reporting practices in studies of the relationship between foreign language classroom anxiety as measured by the FLCAS and gender?
2. What characterizes the foreign language classroom anxiety level of male and female language learners as measured by the FLCAS?
3. What biographical variables moderate the possible relationship between foreign language classroom anxiety and gender?

For our purposes, we chose to conduct a meta-analysis because, as already elaborated on above, it is considered to be a research technique that enables the researcher to identify trends in research outcomes by scrutinizing the results of primary empirical studies in a more objective manner. Since a few publications (e.g., Li et al., 2012; Norris & Ortega, 2006) have also started to pave the way by setting standards to be followed when conducting such studies, we intended to follow their guidelines. Li et al. (2012) views meta-analyses as parallel to conducting empirical research; hence, they claim that much of the quality of any

research synthesis depends on the systematicity in the methods used for collecting and analyzing the literature (Norris & Ortega, 2006). Therefore, in the next sections, we will describe how we went about identifying the studies to be included, the coding process, and the steps of our data analysis.

3. Methods

3.1. Inclusion criteria

Published empirical research papers that used the full (33-item) FLCAS as a data collection tool constituted the data for our meta-analysis. Journal articles published in English were collected through Google Scholar and various academic databases (i.e., EBSCO host, Web of Science, ScienceDirect, and Jstor) accessible for the researchers. It is important to note here that we did not limit our search to high profile publications in order to minimize sampling bias (Norris & Ortega, 2006; Plonsky & Oswald, 2012). In each database, a search was conducted for the expression “foreign language classroom anxiety scale” and the acronym “FLCAS.” The publications had to be more recent than 1986 (the year the FLCAS was published; see Horwitz et al., 1986) and available by May 2020 (the time of the search); the paper had to present a study using the FLCAS as a data collection tool in its complete form, in English or translated (but not abbreviated or altered in any way); the papers had to be published in English (for practical comprehensibility); and full text records had to be available to the researchers. As the final eligibility criterion, the study had to include explicit information on language anxiety in light of the gender distribution of the participants.

Keeping these inclusion criteria in mind and removing duplicates, we continued to work with 44 articles. Since two reports included more than one independent sample, as customary in such cases in meta-analyses, we decided to refer to each independent sample separately. This way, our final sample comprised $k = 48$ studies. Unfortunately, as we began coding the studies in terms of reporting practices, we realized that not all of the studies included information on the instruments’ reliability in the particular context, nor did all of them mention an effect size or sufficient information necessary to estimate an effect size. As a result, for the various analyses we conducted, we used subsamples of the $k = 48$. This is not unusual, since it has been noted by other scholars that inadequate or insufficient information in publications tends to pose a general problem for researchers conducting meta-analyses (Larsen-Hall & Plonsky, 2015). According to Larsen-Hall and Plonsky (2015), the lack of adequate information limits the number of empirical studies that can be included in a quantitative research

synthesis on a given topic, which consequently reduces the strength of conclusions that can be drawn from meta-analyses.

3.2. Coding procedure

We devised a coding scheme in order to systematize the various characteristics of the studies that comprised our sample. For our purposes, we adapted and complemented the scheme developed by Teimouri et al. (2019) because the focus of the present study was very similar. This means that we included information related to the publication of the report (i.e., author, title, journal, abstract, topic, research questions), the sample (i.e., number of participants, country, groups of participants, that is, university, high school students or adult learners, subsamples of males and females), and results (i.e., reliability of the FLCAS, reliability of the FLCAS subscales, the way anxiety levels were interpreted, means for the FLCAS, for the subscales and for the genders, *t*-test results for the comparison of the two genders, beta values from the regression analysis where gender was an independent variable, and any other analyses where gender appeared). The final coding scheme can be found in Table 1.

In order to ensure trustworthiness and credibility, the coding of the studies happened in a recursive fashion, through several rounds. We coded all the data, constantly discussing and revising the codes before resolving problematic points. Once the codes were finalized, the data was ready for analysis.

Table 1 Coding scheme used to identify the main features of the papers included in the sample

Main category	Feature	Definition
Features of the report	Author	The researchers who conducted the study and published it.
	Title	The title of the paper.
	Journal	The journal in which the article was published.
	Abstract	The abstract of the article.
	Topic	The topic to which the article belongs.
	Research question(s)	The research question(s) the authors proposed.
Participants	Number	The sample size of the study.
	Nationality	The nationality of the participants.
	Target language	The foreign language (L2) of the participants.
	Academic status	The educational level of the participants (primary school, secondary school, college/university).
	Proficiency	The proficiency level of the participants (beginning, intermediate, advanced or not specified).
	Number of males Number of females	The number of male participants. The number of female participants.
FLCAS	Language of the questionnaire	The language in which the FLCAS was conducted.
	Reliability index	The internal consistency measure used for the FLCAS (e.g., Cronbach's alpha, test-retest, split-half method).
	Reliability estimate	The reported reliability coefficient for the FLCAS.

	Interpretation of anxiety level	The way the authors interpreted anxiety levels and made categories (e.g., high-anxiety, low-anxiety).
	Mean scores for the whole FLCAS	The reported mean for the FLCAS.
	Standard deviation of the FLCAS scores	The reported standard deviation for the FLCAS.
Subscales of the FLCAS	Number of the subscales	The reported number of underlying scales with factor analysis.
	Subscale labels	The labels assigned to the factors.
	Reliability index for the subscales	The internal consistency measure used for the subscales (e.g., Cronbach's alpha, test-retest, split-half method).
	Reliability estimates for the subscales	The reported reliability coefficient for the subscales.
	Mean of each subscale	The reported mean values for the subscales.
	Standard deviation for each subscale	The reported standard deviation of the subscales.
Descriptive statistics for language anxiety and gender	Mean for the FLCAS	The reported mean for males' and females' scores on the subscales of the FLCAS.
	Mean for the subscales	The reported mean for males' and females' scores on the subscales of the FLCAS.
	Standard deviation of the FLCAS scores	The reported standard deviation of males' and females' scores on the FLCAS.
	Standard deviation of the subscales	The reported standard deviation of males' and females' scores on the subscales of the FLCAS.
Inferential statistics for the analysis of the link between gender and anxiety/effect size	<i>t</i> -test (<i>t</i> statistic)	The <i>t</i> statistical value reported for paired samples <i>t</i> tests or independent samples <i>t</i> tests that are calculated to analyze gender differences in FLCAS scores.
	Regression analysis (beta)	The reported beta (β) value of regression analyses involving gender differences in FLCAS scores.
	Correlation (<i>r</i>)	The reported correlation coefficient (<i>r</i> statistic) with regard to gender differences in FLCAS scores.

3.3. Data analysis

For our investigation, for the descriptive statistics and reliability analysis needed to answer our first research question, we used the Statistical Package for Social Sciences (SPSS) version 26 software. For the computer-assisted meta-analysis necessitated by the second and third research questions, we ran the analyses with the help of the Comprehensive Meta-analysis software, version 3 (CMA; Borenstein et al., 2005). To address the first research question, we computed the overall sample size, looked at minimum and maximum values, means and standard deviations of reliability coefficients reported for the FLCAS and its subscales.

For the second research question, based on the data available for each study, effect sizes (Hedges' *g*) and their associated standard residuals were calculated, and outlier diagnosis was performed. In order to calculate effect sizes (Hedges' *g*) for the gender differences, we used reported sample sizes, *SD* values, as well as *t* and *p* values. In instances where the authors only alluded to the non-significance in the differences between the anxiety levels of males and females, based on Card's (2012) recommendation, Hedges' *g* was recorded as 0.

Where the study reported a significant difference but without providing an exact p value, following Card's (2012) guidelines, p was recorded as $p = .05$.

Tests for heterogeneity of effect sizes were run using a Q test (Lipsey & Wilson, 2001) and the degree of true heterogeneity between studies using the I^2 statistic (Borenstein et al., 2010) to see whether the variation in individual effect sizes can be attributed to between-study differences. Based on the assumption that there were between-study differences, we used a random-effects model and an aggregated effect size to check the overall relationship of language anxiety and gender. A funnel plot with a trim-and-fill test as well as the fail-safe N test served as the basis of determining publication bias. Finally, for the moderator analysis necessary to target the third research question, the categorical moderators of age group, target language, regional context and major were investigated for their modulating effect on the relationship of overall language anxiety as measured by the complete FLCAS and gender. Moderator analyses were also run for the anxiety subscales where it was possible with a minimum of $k = 10$ studies, as recommended by Higgins and Green (2008).

4. Results

The reports included in our meta-analysis ultimately comprised 48 samples with altogether $N = 10,526$ participants, where females were slightly overrepresented ($N_{\text{males}} = 4,523$; $N_{\text{females}} = 5,989$), and there was no information on participants' gender regarding 14 participants, either because they did not indicate their gender or because the empirical study did not provide clear-cut information about it. The studies were conducted between 1994 and 2019, and the total sample sizes were between 30 and 948 ($M = 219.29$; $SD = 185.10$). The sample consisted of participants from various countries, mostly from the Middle East, but other continents were also included, namely Europe, Asia, North America, South America, and Africa. For the final analysis, we included four regions to categorize the individual studies, of which 23 were from the Middle East, 12 from the Far East, eight from Europe, and five from America. One study from Ethiopia (Africa) was included in the Middle East group due to its geographical proximity to this region as well as the fact that no other studies from the middle or southern parts of Africa appeared in our sample; in this way, we avoided one study constituting a group on its own.

Half the reports analyzed the foreign language anxiety of the participants with regard to the English language ($k = 24$). Other studies focused on Japanese, Spanish, French, German, and Arabic language classroom anxiety. With a similar ratio, a considerable number of the studies included university students ($k = 25$), and a smaller proportion involved high school students and adult learners. From

the university context, 10 studies selected participants majoring in the language for which the researchers obtained FLCAS scores, while the other studies involved various programs even from non-language specialties. The proficiency of the participants was reported only in a few instances by way of grade point averages or self-reported levels of proficiency; this varied on a considerably large scale from beginner to more advanced learners.

4.1. The methodological features and reporting practices in studies on foreign language classroom anxiety and gender

The first research question focused on the methodological and reporting practices in the studies scrutinizing the relationship between language anxiety and gender as measured by the FLCAS. Overall, in terms of the data analyses and the respective reported results, the sample studies showed great variation, perhaps due to the disparity range in the publication standards of the different journals. This fragmented picture is also apparent in the presentation of our results.

First and foremost, for the $k = 48$ studies, reliability was reported 28 times (58.33%) for the whole FLCAS by the Cronbach's alpha internal consistency measure (see Table 2), while five papers (10.41%) referred to the reliability of the FLCAS subscales, and one study (2.08%) calculated alpha values for each item. It must be noted here that Horwitz et al. (1986) did not explicitly refer to the instrument consisting of these subscales. They claimed that communication apprehension, fear of negative evaluation and test anxiety were closely linked to FLCA rather than being components of it. Nonetheless, in our sample, 16 papers referred to the subscale of communication apprehension and fear of negative evaluation, while 15 studies reported data about test anxiety. Out of these, only five indicated the reliability of these subscales. Five studies reported other subscales emerging from the items referring to a kind of general (speaking/language classroom) anxiety component, which the authors most frequently labeled as "general English class anxiety." Other types of reliability measures, albeit extremely rare, also appeared in the works synthesized, including one study with a split-half method, and another with a test-retest reliability analysis for the complete instrument.

Table 2 Reliability indices (Cronbach's α) of the FLCAS and its subscales

Scales	Number of studies reporting Cronbach's α	Min.	Max.	<i>M</i>	<i>SD</i>
Complete FLCAS	28	.75	.96	0.88	0.06
Subscale: Communication apprehension	5	.72	.89	0.81	0.07
Subscale: Fear of negative evaluation	5	.62	.81	0.73	0.07
Subscale: Test anxiety	5	.71	.84	0.80	0.06

The sample studies included in the meta-analysis provided data on the overall foreign language classroom anxiety of the participating male and female subgroups, as well as the various components associated with FLCA, namely, communication apprehension, fear of negative evaluation and test anxiety. The reported language anxiety scores themselves, however, appeared on a variety of scales. That is to say, some studies interpreted the mean scores on a 1 to 5 scale, whereas others simply added up the numerical values associated with the Likert-scale responses from strongly disagree (1) to strongly agree (5). Altogether, 24 studies included the overall means for the FLCAS for both males and females, while 12 reported the mean scores for the two genders respective to the subscales of the FLCAS, namely, communication apprehension, fear of negative evaluation and test anxiety, and in a few instances the emerging scale of "general English class anxiety."

In regard to the relationship of gender and FLCA, 26 studies used *t* tests, 17 used ANOVA, nine regression analyses with gender as one of the predictors, and in one study (despite the fact that gender is not a continuous variable), researcher(s) ran correlational analyses. For the relationship between gender and language anxiety, the effect size was only calculated in five studies, where either Cohen's *d* or the partial eta squared (η^2) was reported. Unfortunately, no studies out of the 48 reported Hedges' *g*, which is considered to be an unbiased effect size measure (Cooper et al., 2019), though for studies with a larger sample size, Cohen's *d* is very similar to Hedges' *g* (Card, 2012). We find it puzzling that the wealth of the studies failed to report the effect size which would otherwise be of crucial practical importance. In fact, while statistical significance shows that the difference between groups is not due to chance, effect size gives a lot more; it shows whether the results are practically significant (Plonsky & Oswald, 2014). This shortcoming may have become apparent due to the fact that, as an attempt to avoid publication bias in our meta-analysis, our sample was not restricted only to the top publications in the field.

4.2. Language classroom anxiety level of male and female language learners as measured by the foreign language classroom anxiety scale

As regards our second research question about the foreign language classroom anxiety levels of male and female language learners as measured by the FLCAS, we again had a variety of data to work with; hence the results are also manifold. First of all, we looked at the relationship of the overall FLCAS scores and gender. Based on the data available, we were able to calculate the standard difference in means (cf. Hedges' *g*) for 32 studies (out of which 15.63% were assigned a *g* value of 0 due to reporting only the fact that non-significant differences were found, and 3.13% reported only that significant differences were found without

providing any additional information; hence, $p = .05$ was assigned to these studies). Effect sizes and the associated standard residuals were inspected to identify outliers. Because all standard residuals were below the threshold of 2.5 (Teimouri et al., 2019), we proceeded with the analysis by keeping our subsample intact.

After this, we checked the test of heterogeneity, and upon inspecting the results, we could state that, by rejecting the null hypothesis of homogeneity, heterogeneity was present amongst the selected studies ($Q(31) = 295.94, p < .001$). This means that the observed variability in the selected 32 studies was higher than what would be expected based solely on sampling fluctuation (Card, 2012). In other words, the dispersion of the effect sizes was not only due to chance and random error, but there seemed to be real differences in the studies' effects; there appeared to be between-study differences most probably linked to the variety of contexts (regional, linguistic, age, etc.) in which the studies were conducted. A forest plot is presented to visualize the overall dispersion of effect sizes of the selected studies (see the Appendix), where the diamond shows the summary effect in light of the confidence interval (Borenstein et al., 2009). However, as the Cochran's Q value was applied for testing the null hypothesis, it was necessary to check whether the proportion of the observed variance reflected true heterogeneity in the effect sizes (Borenstein et al., 2016) using Higgins' I^2 and T^2 values. The I -squared value (I^2) was 89.53, which means that nearly 90% of the observed variance was probably true variance and was not due to sampling error. True heterogeneity or, in other words, the variance of true effects (T^2), was 0.18, and the standard deviation of true effects (T) was 0.43.

Because the test of heterogeneity was statistically significant, we opted for the random-effects model as it concentrates on the population distribution of the effect sizes as opposed to the fixed-effects model which focuses on a single effect size (Card, 2012). According to Card (2012), the random-effects model takes the standard deviation as well as the central tendency into consideration. Therefore, we analyzed the central tendencies of the effect sizes by running a random-effects model to see whether the studies in our sample provided evidence for any significant differences between the two genders' foreign language classroom anxiety level. Our results showed a negative mean effect size -0.119 with an associated statistical significance value of $p = .152$ for the random-effects model with 95% confidence interval (CI) $[-0.282, 0.044]$. This means that although the results of the pooled studies showed a tendency for females to have slightly higher overall scores on the FLCAS, this result was not statistically significant.

Following this, we also investigated the results of studies that looked at language learners' gender and the scores on the most frequently reported subscales of the FLCAS. For the communication apprehension scale, we were able to calculate Hedges' g for $k = 14$ studies, where 24.55% were assigned the g value of 0 due to

reporting only the fact that non-significant differences were found. The mean effect size was -0.096 , 95% CI $[-0.314, 0.121]$, $p = .385$. The mean effect sizes for the fear of negative evaluation scale ($k = 14$) (out of all effect sizes 24.55% were assigned the g value of 0 due to reporting only the fact that non-significant differences were found) were -0.134 , 95% CI $[-0.349, 0.081]$, $p = .221$. In the case of the test anxiety scale ($k = 13$) (out of all effect sizes, 15.38% were assigned the g value of 0 due to reporting only the fact that non-significant differences were found) the effect sizes were -0.046 , 95% CI $[-0.166, 0.075]$, $p = .457$. In each case, although the direction of the relationship seemed to indicate a higher level of anxiety in the case of female learners, once again, these results were not significant.

Finally, we have to note that in order to detect possible publication bias we created a funnel plot (i.e., a scatterplot of effect sizes). As the funnel plot output is used primarily to detect possible publication bias and not to “correct” or adjust them, we used Duval and Tweedie’s (2000) trim-and-fill method to estimate the number of missing studies (Duval, 2005). Under the random-effects model for the combined studies, the point estimate was -0.119 with 95% CI $[-0.282, 0.043]$ and, using the trim-and-fill procedure, these values were unchanged. As depicted in the funnel plot (see Figure 1), our analysis showed a slight bias towards studies with positive small effects. Following Borenstein et al.’s (2009) guidelines, we also computed Rosenthal’s fail-safe N in order to deal with this slight bias and to see how many missing studies would be needed for the p value to exceed .05. The fail-safe N was 94, which means that we would need 94 studies to nullify the effect. In the light of our analysis subsuming 48 samples, we interpreted this as meaning that there was no reason to assume that the true effect was zero.

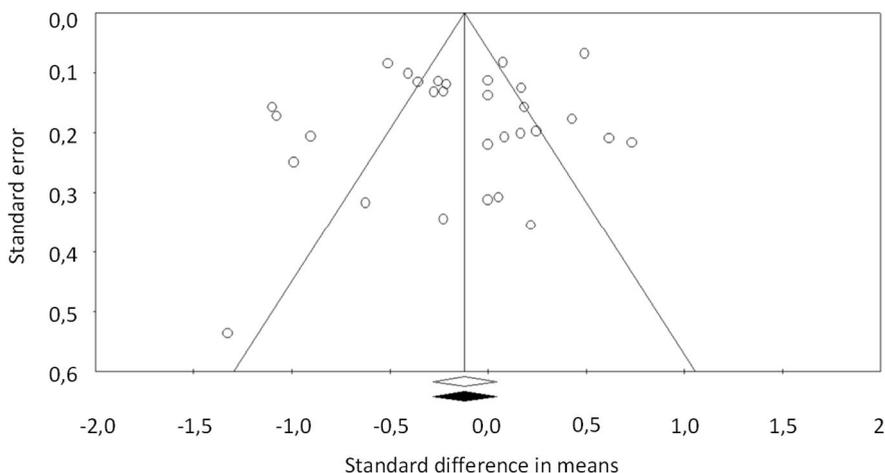


Figure 1 The funnel plot used to detect possible publication bias by the standard difference in means

4.3. Moderating influences of biographical variables on the relationship of language anxiety and gender

As for the third research question, we investigated what biographical variables moderated the relationship between language anxiety and gender. For the analysis, we looked at four possible moderators: the age group of the learners based on their school levels, the language being studied, the geographical region where the foreign language was being learnt and, in the case of university samples, the major of the participants. For each of these moderators, subgroups of effect sizes were calculated (see Table 3).

Table 3 The results of the moderator analyses with random-effects models for the complete FLCAS

	<i>K</i>	<i>N</i>	<i>M</i>	<i>SE</i>	95% CI		<i>z</i>	<i>p</i>
					<i>LL</i>	<i>UL</i>		
<i>Age level</i>								
Elementary	1	260	0.170	0.479	-0.769	1.110	0.356	.722
High school	2	505	-0.100	0.345	-0.776	0.576	-0.291	.771
University	24	5,439	-0.138	0.103	-0.339	0.064	-1.338	.181
High school and university	1	355	0.000	0.476	-0.934	0.934	0.000	1.000
Adult learners	4	296	-0.142	0.267	-0.665	0.382	-0.530	.596
<i>Total</i>	32	6,855	-0.120	0.089	-0.295	0.055	-1.346	.178
<i>Target language</i>								
English	24	6,097	-0.148	0.097	-0.339	0.043	-1.520	.128
Arabic	1	233	-0.274	0.459	-1.175	0.626	-0.597	.550
Japanese	1	96	0.085	0.486	-0.868	0.038	0.175	.861
Blank	6	429	0.009	0.211	-0.404	0.422	0.043	.966
<i>Total</i>	32	6,855	-0.119	0.085	-0.287	0.048	-1.397	.162
<i>Regional context</i>								
Middle East	18	3,536	-0.195	0.116	-0.422	0.033	-1.673	.094
Far East	6	2,298	0.084	0.193	-0.294	0.461	0.435	.663
Europe	6	692	-0.124	0.211	-0.538	0.290	-0.586	.558
America	2	329	-0.105	0.340	-0.771	0.562	-0.307	.758
<i>Total</i>	32	6,855	-0.110	0.100	-0.306	0.086	-1.100	.271
<i>Major</i>								
English	10	3,874	0.004	0.172	-0.333	0.341	0.024	.981
Other	11	1,204	-0.213	0.155	-0.516	0.091	-1.373	.170
Blank	3	361	-0.310	0.314	-0.926	0.306	-0.987	.324
<i>Total</i>	24	5,439	-0.139	0.116	-0.366	0.087	-1.206	.228

Note. CI = confidence interval, *LL* = lower limit, *UL* = upper limit

Based on our analyses, we could not establish that any of the variables under scrutiny moderate the relationship of language anxiety and gender. This

means that the link between gender and learners' levels of language anxiety did not depend on their age, the target language, the regional context, or the major studied at university. As seen from Table 3, in our sample of studies, there were clearly underrepresented groups in terms of the learners' age group, the target language, the regional context, and university students' majors as most studies were conducted in the university context with English as a target language. From our analyses, it appears that the European and American continents were also underrepresented.

We also looked at studies' results reporting participants' data on language anxiety based on the three subscales (i.e., communication apprehension, fear of negative evaluation and test anxiety) in order to determine the modulating influences of the biographical variables. Although the samples for these were not very large, the number of total studies were above the recommended minimum of $k = 10$ (Higgins & Green, 2008). Tables 4-6 summarize the results of the moderator analyses for these subscales.

Table 4 The results of the moderator analyses with random-effects models for the subscale of communication apprehension

	<i>K</i>	<i>N</i>	<i>M</i>	<i>SE</i>	95% CI		<i>z</i>	<i>p</i>
					<i>LL</i>	<i>UL</i>		
<i>Age</i>								
Elementary	1	260	0.192	0.431	-0.652	1.036	0.446	.656
High school	3	1,065	-0.100	0.251	-0.591	0.391	-0.398	.690
University	10	2,219	-0.126	0.144	-0.409	0.156	-0.877	.381
<i>Total</i>	14	3,544	-0.096	0.120	-0.331	0.140	-0.796	.426
<i>Regional context</i>								
Middle East	7	1,437	-0.021	0.166	-0.347	0.305	-0.128	.898
Far East	4	1,644	-0.108	0.210	-0.520	0.304	-0.512	.609
Europe	3	463	-0.271	0.269	-0.799	0.257	-1.006	.314
<i>Total</i>	14	3,544	-0.096	0.117	-0.326	0.134	-0.815	.415
<i>Major</i>								
English	3	274	0.135	0.275	-0.405	0.674	0.489	.625
Other	6	1,904	-0.227	0.183	-0.586	0.132	-1.240	.215
Blank	1	41	-0.253	0.544	-1.319	0.814	-0.464	.642
<i>Total</i>	10	2,219	-0.117	0.164	-0.438	0.204	-0.717	.473

Note. CI = confidence interval, *LL* = lower limit, *UL* = upper limit

Table 5 The results of the moderator analyses with random-effects models for the subscale of fear of negative evaluation

	<i>K</i>	<i>N</i>	<i>M</i>	<i>SE</i>	95% CI		<i>z</i>	<i>p</i>
					<i>LL</i>	<i>UL</i>		
<i>Age</i>								
Elementary	1	260	0.287	0.381	-0.459	1.032	0.753	.451
High school	3	1,065	0.10	0.222	-0.425	0.444	0.043	.966
University	10	2,219	-0.231	0.129	-0.483	0.021	-1.794	.073
<i>Total</i>	14	3,544	-0.081	0.158	-0.390	0.229	-0.511	.609
<i>Regional context</i>								
Middle East	7	1,437	-0.148	0.169	-0.480	0.183	-0.877	.381
Far East	4	1,644	-0.162	0.214	-0.582	0.258	-0.756	.450
Europe	3	463	-0.051	0.274	-0.588	0.485	-0.187	.852
<i>Total</i>	14	3,544	-0.134	0.120	-0.368	0.100	-1.122	.262
<i>Major</i>								
English	3	274	0.011	0.273	-0.524	0.545	0.039	.969
Other	6	1,904	-0.339	0.182	-0.696	0.017	-1.864	.062
Blank	1	41	-0.196	0.541	-1.257	0.865	-0.362	.717
<i>Total</i>	10	2,219	-0.222	0.156	-0.527	0.083	-1.429	.153

Note. CI = confidence interval, LL = lower limit, UL = upper limit

Table 6 The results of the moderator analyses with random-effects models for the subscale of test anxiety

	<i>K</i>	<i>N</i>	<i>M</i>	<i>SE</i>	95% CI		<i>z</i>	<i>p</i>
					<i>LL</i>	<i>UL</i>		
<i>Age</i>								
Elementary	1	260	0.033	0.215	-0.388	0.453	0.153	.878
High school	3	1,065	0.064	0.125	-0.181	0.308	0.511	.609
University	9	2,062	-0.104	0.083	-0.267	0.058	-1.258	.209
<i>Total</i>	13	3,387	-0.036	0.076	-0.186	0.114	-0.472	.637
<i>Regional context</i>								
Middle East	6	1,280	-0.067	0.104	-0.272	0.137	-0.647	.518
Far East	4	1,644	-0.021	0.110	-0.237	0.194	-0.192	.848
Europe	3	463	-0.038	0.163	-0.357	0.281	-0.233	.816
<i>Total</i>	13	3,387	-0.044	0.069	-0.179	0.090	-0.644	.520
<i>Major</i>								
English	3	274	0.040	0.172	-0.297	0.376	0.230	.818
Other	5	1,747	-0.153	0.111	-0.370	0.063	-1.387	.166
Blank	1	41	-0.220	0.399	-1.003	0.563	-0.550	.582
<i>Total</i>	9	2,062	-0.103	0.091	-0.281	0.074	-1.139	.255

Note. CI = confidence interval, LL = lower limit, UL = upper limit

Based on our findings, we cannot claim that the biographical variables included in our study modulated the relationship of gender and language anxiety as measured by the subscales of the FLCAS. Although most of the time the data suggested that females tended to have higher levels of anxiety, the differences failed to reach significance. More importantly, gender may denote a more complex construct than researchers following the positivist paradigm originally thought and it may thus be an oversimplification to investigate the differences (or lack thereof) in language anxiety by merely comparing females' and males' FLCAS scores.

5. Discussion

Although there are recent meta-analyses that have examined the relationship between FLCA and language achievement (e.g., Botes et al., 2020; Teimouri et al., 2019; X. Zhang, 2019), our research synthesis aimed at examining a relatively neglected area of systematic review, namely, the possible connection between foreign language anxiety, measured by the FLCAS and an important demographic variable, that is, gender. As for our first research question, the results of our systematic review showed considerable variation with respect to the methodological practices as well as reporting the results in studies focusing on the relationship between foreign language classroom anxiety, as measured by the FLCAS and gender. This raises many issues in terms of research quality assurance. Based on our inclusion criteria, all the studies we looked at employed the complete version of the FLCAS as a data collection instrument; unfortunately, however, many authors seemed to have taken the tool and its psychometric qualities (especially its consistency) for granted and almost half of them failed to report the results of reliability checks for the given contexts. It is important that, when translating or using an instrument, even if it is a well-established one, the reliability of that particular version in a particular context should be ensured and accounted for (Derrick, 2015; Larsen-Hall & Plonsky, 2015). When the authors did check the reliability of the instrument, it was most often done by relying on the Cronbach's alpha internal consistency measure, while other reliability analyses were scarcely used (e.g., split-half method, test-retest reliability check) (cf. Derrick, 2015). The main issue with only reporting Cronbach's alpha is that it does not address unidimensionality and misunderstandings around its interpretation also abound (Hoekstra et al., 2018).

When it comes to the instrument, it was also interesting to see that some authors referred to the complete instrument and used it as an overall measure of foreign language classroom anxiety, while others looked at the (supposed) underlying factors, either by using other researchers' previous groupings or referring to the misconceived notion that the questionnaire purports to measure these distinct constructs of communication apprehension, fear of negative evaluation,

and test anxiety (Horwitz, 2017). Instead, for the validity argument and in order to justify the interpretation of the responses indicating learners' language anxiety levels, statistical analyses (e.g., factor analysis) of the data from a given sample would have been more useful (Park, 2014). We also found that researchers applied a vast array of statistical procedures involving paired and independent samples *t*-tests or analyses of variance (ANOVAs). However, quite surprisingly, no studies used hierarchical cluster analysis to group participants and shed light on foreign language anxiety patterns, which would add more to our understanding with respect to learner profiles (as suggested by Horwitz, 2017) and to analyzing specific learner types (Csizér & Dörnyei, 2005).

Another noteworthy aspect is that little attention has been devoted to reporting the *p* value appropriately to indicate statistical significance (or the lack thereof). According to the American Psychological Association (2020), researchers ought to report the exact *p* value unless $p < .001$ (pp. 180, 204). Many studies in our investigation failed to report the *p* value, and this practice is not at all beneficial for meta-analysts because the researchers may have to leave out complete studies which would otherwise be important for the analysis, or they would have to work with the least favorable level of significance (Card, 2012).

Unfortunately, none of the 48 samples relied on an associated Hedges' *g* as the unbiased effect size measure, and Cohen's *d* was reported only in a handful of studies; what is more, only 32 mentioned data that could be used to calculate Hedges' *g*. It is important to note that statistical significance only tells us that we can reject the null hypothesis, while the effect size shows us practical importance (Card, 2012; Plonsky & Oswald, 2014). Therefore, reporting the effect size is indispensable in understanding the practical significance of the results. Overall, we can say that our findings are in line with Teimouri et al.'s (2019) conclusions, who claimed that "we can see a lopsided approach toward assessing and reporting the measurement characteristics of instruments in anxiety research" (p. 376). As a less important issue, reporting practices also showed inconsistency in terms of referring to learners' language anxiety levels. Although the same measure was used, the scores were difficult to compare directly because in some cases the average of the responses on the Likert-scale items was given, whereas in others the authors provided a sum of the responses to individual items.

With respect to our second research question about what characterizes the FLCA level of male and female language learners as measured by the FLCAS, we can state that despite the tendency for females to manifest slightly higher anxiety, this result was not statistically significant, both with respect to the whole instrument and its suggested subscales. In a previous meta-analysis, Botes et al. (2020) arrived at similar results, reporting that the link between language anxiety and achievement was not moderated significantly by the proportion of female learners.

It must also be mentioned that the construct of gender nowadays is increasingly interpreted in its social context rather than as a binary biographical variable (e.g., Dewaele et al., 2016). Therefore, Dewaele et al. (2016) forewarned researchers that “before speculating on possible reasons for differences between women and men (or the absence of them), there is reason to investigate how large the differences between . . . men and women really are, especially when it comes to language learning” (p. 42). Naturally, as the data in the studies included in this meta-analysis was based on the binary-coded gender variable (male/female), we cannot make conclusions about FLCA and gender as a social variable. This points to the inherent complexity of the construct and to the importance of its cautious interpretation. Although the binary interpretation of gender has been dominant for centuries, this construct may be more complex than it appears at first sight.

Finally, to see what biographical variables seem to moderate the relationship between foreign language classroom anxiety and gender, we relied on the most frequently reported demographic variables, namely, the age group of the learners, the language being studied, the geographical region where the L2 is being learnt, and the major of the university participants. Based on the results of the analyses run on our sample, we could not conclude any modulating influence of these variables on the link between gender and foreign language anxiety. Therefore, we cannot say that age, target language, regional context, or, in the case of university students, their majors play a discernible role concerning the relationship between FLCA and gender.

6. Conclusions

In the present study, we set out to examine the association between foreign language classroom anxiety and gender by conducting a meta-analysis on research utilizing the full 33-item version of the FLCAS and tapping into the link between language anxiety levels and gender. More precisely, we looked at the reporting practices of these studies, the magnitude of the link between language anxiety and gender, and various biographical variables that may modulate this relationship.

First of all, we found that great variation exists in the methodological and reporting practices of the studies despite the relatively small number of eligible research endeavors. The authors of these papers generally relied on Cronbach's alpha internal consistency measure to check the reliability of the instrument, but a considerable number of them failed to report effect sizes. We saw various statistical procedures being employed to analyze foreign language anxiety differences, as measured by the FLCAS, albeit multivariate statistical methods were

scarcely used. The results of our research synthesis indicate that while females showed a tendency to manifest slightly more foreign language classroom anxiety, this result was not statistically significant; therefore, based on the present meta-analysis, we can say that gender does not seem to be linked to differences in FLCA levels as measured by the FLCAS. Additionally, based on moderator analyses, we could not draw any conclusions as to the variables of age, regional context, target language, and study major influencing the link between language anxiety and gender.

Moving on to the limitations of our meta-analysis, we must highlight that the number of studies involved in the final analysis was rather small. This, however, might be accounted for by the fact that, unfortunately, many studies reported missing data or focused on analyzing the responses to individual items rather than scales (i.e., dimensions) subsuming more items. The issue of missing data when conducting fairly large-scale meta-analyses is also highlighted by Larsen-Hall and Plonsky (2015), who state that “omitted statistics – or, more precisely, the authors who omitted them – are responsible in some cases for rendering massive amounts of data un-meta-analyzable and therefore unavailable to contribute to already limited efforts to aggregate findings across studies” (p. 133).

While we acknowledge that educators should ultimately raise learners’ awareness of foreign language anxiety and assist them in combating this negative emotion rather than worrying about measurement issues and statistical procedures (Horwitz, 2017), we believe that the role of researchers is to provide evidence and backing concerning the language learning-related phenomenon under investigation. In order for this information to be interpreted in a valid and reliable fashion (which in turn would allow us to draw overarching and valid conclusions by way of meta-analytic studies), we think that it is important to ensure quality in not only high profile publications but at the level of individual empirical studies as well. Apart from quality control, we find it noteworthy to mention that more meta-analytic studies should be conducted on the role that language anxiety plays in the process of language learning, perhaps by focusing on other biographical and contextual variables.

Acknowledgement

The first author was supported by the NKFIH – 129149 research grant. The second author is a member of the MTA-ELTE Foreign Language Teaching Research Group and was supported by the Research Program for Public Education Development of the Hungarian Academy of Sciences.

References¹

- *Abood, M. H., & Abu-Melhim, A-R. H. (2015). Examining the effectiveness of group counseling in reducing anxiety for Jordanian EFL learners. *Journal of Language Teaching and Research*, 6(4), 749-757. <https://doi.org/10.17507/jltr.0604.06>
- *Abood, M. H., & Ahouari-idri, N. (2017). The effect of group counselling based on the modification of negative self-statements on reducing gender-biased foreign language anxiety among Ajloun National University students. *Journal of Educational and Psychological Studies*, 11(4), 730-735. <https://doi.org/10.24200/jeps.vol11iss4pp730-735>
- *Aida, Y. (1994). Examination of Horwitz, Horwitz, and Cope's construct of foreign language anxiety: The case of students of Japanese. *Modern Language Journal*, 78(2), 155-168. <https://doi.org/10.1111/j.1540-4781.1994.tb02026.x>
- *Al-Shuaibi, J., Ayman M. H-M., & Saleh, N. A. (2014). Foreign language anxiety among students studying foreign languages. *Life Science Journal*, 11(8), 197-203.
- *Amengual-Pizzaro, M. (2018). Foreign language classroom anxiety among English for specific purposes (ESP) students. *International Journal of English Studies*, 18(2), 145-159. <https://doi.org/10.6018/ijes/2018/2/323311>
- American Psychological Association. (2020). *Publication manual of the American Psychological Association* (7th ed.). American Psychological Association. <https://doi.org/10.1037/0000165-000>
- *Amiri, M., & Ghonsooly, B. (2015). The relationship between English learning anxiety and the students' achievement on examinations. *Journal of Language Teaching and Research*, 6(4), 855-865. <https://doi.org/10.17507/jltr.0604.20>
- Arnaiz, P., & Guillén, P. (2012). Foreign language anxiety in a Spanish university setting: Interpersonal differences. *Revista de Psicodidáctica*, 17(1), 5-26.
- *Bensalem, E. (2019). The relationship between foreign language classroom anxiety and background variables in a multilingual context. *International Journal of English Linguistics*, 9(6), 249-256. <https://doi.org/10.5539/ijel.v9n6p249>
- Borenstein, M., Hedges L. V., Higgins, J. P. T., & Rothstein, H. R. (2005). Comprehensive meta-analysis (version 2.2.027) [Computer software].
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.
- Borenstein, M., Hedges, L. V., Higgins, J. P. T., & Rothstein, H. R. (2010). A basic introduction to fixed-effect and random-effects models for meta-analysis. *Research Synthesis Methods*, 1(2), 97-111. <https://doi.org/10.1002/jrsm.12>

¹ References marked with an asterisk indicate studies included in the meta-analysis.

- Borenstein, M., Higgins, J. P. T., Hedges, L. V., & Rothstein, H. R. (2016). Basics of meta-analysis: I^2 is not an absolute measure of heterogeneity. *Research Synthesis Methods*, 8(1), 5-18. <https://doi.org/10.1002/jrsm.1230>
- Botes, E., Dewaele, J.-M., & Greiff, S. (2020). The foreign language classroom anxiety scale and academic achievement: An overview of the prevailing literature and a meta-analysis. *Journal for the Psychology of Language Learning*, 2(1), 26-56. <https://doi.org/10.52598/jpll/2/1/3>
- *Briesmaster, M., & Briesmaster-Paredes, J. (2015). The relationship between teaching styles and NNPSETs' anxiety levels. *System*, 49, 145-156. <https://doi.org/10.1016/j.system.2015.01.012>
- Campbell, C. M. (1999). Language anxiety in men and women: Dealing with gender difference in the language classroom. In D. J. Young (Ed.), *Affect in foreign language and second language learning: a practical guide to creating a low-anxiety classroom atmosphere* (pp. 191-215). McGraw Hill.
- Campbell, C. M., & Shaw, V. M. (1994). Language anxiety and gender differences in adult second language learners: Exploring the relationship. In C. A. Klee (Ed.), *Faces in a crowd: The individual learner in multi-section courses* (pp. 47-80). Heinle & Heinle.
- *Capan, S. A., & Simsek, H. (2012). General foreign language anxiety among EFL learners: A survey study. *Frontiers in Language and Teaching*, 3, 116-124.
- Card, N. A. (2012). *Applied meta-analysis for social science research*. The Guilford Press.
- *Cheng, Y. (2002). Factors associated with foreign language writing anxiety. *Foreign Language Annals*, 35(6), 647-656. <https://doi.org/10.1111/j.1944-9720.2002.tb01903.x>
- *Cocorada, E., & Maican, A. (2013). A study of foreign language anxiety with Romanian students. *Bulletin of the Transilvania University of Braşov*, 6(2), 9-18.
- Cooper, H., Hedges, L. V., & Valentine, J. C. (2019). *The handbook of research synthesis and meta-analysis* (3rd ed.). Russell Sage Foundation.
- *Çubuçuğu, F. (2008). A study of the correlations between self efficacy and foreign language learning anxiety. *Journal of Theory and Practice in Education*, 4(1), 148-158.
- *Cui, J. (2011). Research on high school students' English learning anxiety. *Journal of Language Teaching and Research*, 2(4), 875-880. <https://doi.org/10.4304/jltr.2.4.875-880>
- Csizér, K., & Dörnyei, Z. (2005). Language learners' motivational profiles and their motivated learning behavior. *Language Learning*, 55(4), 613-659. <https://doi.org/10.1111/j.0023-8333.2005.00319.x>
- *Debreli, E., & Demirkan, S. (2016). Sources and levels of foreign language speaking anxiety of English as a foreign language university students with

- regard to language proficiency and gender. *International Journal of English Language Education*, 4(1), 49-62. <https://doi.org/10.5296/ijele.v4i1.8715>
- Derrick, D. J. (2015). Instrument reporting practices in second language research. *TESOL Quarterly*, 50(1), 132-153. <https://doi.org/10.1002/tesq.217>
- Dewaele, J.-M. (2002). Psychological and sociodemographic correlates of communicative anxiety in L2 and L3 production. *International Journal of Bilingualism*, 6(1), 23-38. <https://doi.org/10.1177/13670069020060010201>
- Dewaele, J.-M. (2005). Investigating the psychological and emotional dimensions in instructed language learning: Obstacles and possibilities. *Modern Language Journal*, 89(3), 367-380. <https://doi.org/10.1111/j.1540-4781.2005.00311.x>
- Dewaele, J.-M. (2007). The effect of multilingualism, sociobiographical, and situational factors on communicative anxiety and foreign language anxiety of mature language learners. *International Journal of Bilingualism*, 11(4), 391-409. <https://doi.org/10.1177/13670069070110040301>
- Dewaele, J.-M. (2013a). *Emotions in multiple languages* (2nd ed.). Palgrave Macmillan.
- *Dewaele, J.-M. (2013b). The link between foreign language classroom anxiety and psychoticism. Extraversion and neuroticism among adult bi- and multilinguals. *Modern Language Journal*, 97(3), 670-684. <https://doi.org/10.1111/j.1540-4781.2013.12036.x>
- Dewaele, J.-M., Petrides, K. V., & Furnham, A. (2008). The effects of trait emotional intelligence and sociobiographical variables on communicative anxiety and foreign language anxiety among adult multilinguals: A review and empirical investigation. *Language Learning*, 58(4), 911-960. <https://doi.org/10.1111/j.1467-9922.2008.00482.x>
- Dewaele, J.-M., MacIntyre, P. D., Boudreau, C., & Dewaele, L. (2016). Do girls have all the fun? Anxiety and enjoyment in the foreign language classroom. *Theory and Practice of Second Language Acquisition*, 2(1), 41-63. <https://journals.us.edu.pl/index.php/TAPSLA/article/view/3941/3090>
- *Dogan, Y., & Tuncer, M. (2016). Examination of foreign language classroom anxiety and achievement in foreign language in Turkish university students in terms of various variables. *Journal of Education and Training Studies*, 4(5), 18-29. <https://doi.org/10.11114/jets.v4i5.1337>
- Donovan, L., & MacIntyre, P. D. (2004). Age and sex differences in willingness to communicate, communication apprehension, and self-perceived competence. *Communication Research Reports*, 21(4), 420-427. <https://doi.org/10.1080/08824090409360006>
- Duval, S. (2005). The trim and fill method. In H. R. Rothstein, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis: Prevention, assessment and adjustments* (pp. 127-144). John Wiley & Sons.

- Duval, S., & Tweedie, R. (2000). Trim and fill: A simple funnel-plot-based method of testing and adjusting for publication bias in meta-analysis. *Biometrics*, *56*, 455-463. <https://doi.org/10.1111/j.0006-341x.2000.00455.x>
- *Elaldi, S. (2016). Foreign language anxiety of students studying English language and literature: A sample from Turkey. *Educational Research and Reviews*, *11*(6), 219-228. https://doi.org/_10.5897/ERR2015.2507
- *Elkhafaifi, H. (2005). Listening comprehension and anxiety in the Arabic language classroom. *Modern Language Journal*, *89*(2), 206-220. <https://doi.org/10.1111/j.1540-4781.2005.00275.x>
- Eysenck, M. W. (1979). Anxiety, learning, and memory: A reconceptualization. *Journal of Research in Personality*, *13*(4), 363-385. [https://doi.org/10.1016/0092-6566\(79\)90001-1](https://doi.org/10.1016/0092-6566(79)90001-1)
- *Farjami, H., & Amerian, M. (2012). Relationship between EFL learners' perceived social self-efficacy and their foreign language classroom anxiety. *Journal of English Language Teaching and Learning*, *4*(10), 77-103.
- *Gerencheal, B. (2016). Gender differences in foreign language anxiety at an Ethiopian university: Mizan-Tepi University third year English major students in focus. *African Journal of Education and Practice*, *1*(1), 1-16. <https://files.eric.ed.gov/fulltext/ED582461.pdf>
- *Ghorban Dordinejad, F., & Moradian Ahmadabad, R. (2014). Examination of the relationship between foreign language classroom anxiety and English achievement among male and female Iranian high school students. *International Journal of Language Learning and Applied Linguistics World*, *6*(4), 446-460. <https://doi.org/10.1007/s10936-015-9371-5>
- Higgins, J. P. T., & Green, S. (Eds.). (2008). *Cochrane handbook for systematic reviews of interventions: Cochrane book series*. Wiley-Blackwell.
- *Hismanoglu, M. (2013). Foreign language anxiety of English language teacher candidates: A sample from Turkey. *Procedia – Social and Behavioral Sciences*, *93*, 930-937. <https://doi.org/10.1016/j.sbspro.2013.09.306>
- Hoekstra, R., Vugteveen, J., Warrens, M. J., & Kruyen, P. M. (2018). An empirical analysis of alleged misunderstandings of coefficient alpha. *International Journal of Social Research Methodology*, *22*(4), 351-364. <https://doi.org/10.1080/13645579.2018.1547523>
- Horwitz, E. K. (2017). On the misreading of Horwitz, Horwitz and Cope (1986) and the need to balance anxiety research and the experiences of anxious language learners. In C. Gkonou, M. Daubney, & J.-M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 31-48). Multilingual Matters. <https://doi.org/10.21832/9781783097722-003>
- Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal*, *70*(2), 125-132. <https://doi.org/10.2307/327317>

- *Huang, H.-T. D. (2018). Modeling the relationships between anxieties and performance in second/foreign language speaking assessment. *Learning and Individual Differences, 36*, 44-56. <https://doi.org/10.1016/j.lindif.2018.03.002>
- *Karabiyik, C., & Özkan, N. (2017). Foreign language anxiety: A study at Ufuk University preparatory school. *Journal of Language and Linguistic Studies, 13*(2), 667-680.
- Kitano, K. (2001). Anxiety in the college Japanese language classroom. *Modern Language Journal, 85*(4), 549-566. <https://doi.org/10.1111/0026-7902.00125>
- Kleinmann, H. H. (1977). Avoidance behavior in adult second language acquisition. *Language Learning, 27*(1), 93-107. <https://doi.org/10.1111/j.1467-1770.1977.tb00294.x>
- Larsen-Hall, J., & Plonsky, L. (2015). Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning, 65*, 127-159. <https://doi.org/10.1111/lang.12115>
- *Latif, N. A. B. (2015). A study on English language anxiety among adult learners in Universiti Teknologi Malaysia (UTM). *Procedia – Social and Behavioral Sciences, 208*, 223-232. <https://doi.org/10.1016/j.sbspro.2015.11.198>
- Levitt, E. (1980). *The psychology of anxiety* (2nd ed.). Lawrence Erlbaum.
- Li, S., Shintani, N., & Ellis, R. (2012). Doing meta-analysis in SLA: Practice, choice, and standards. *Contemporary Foreign Language Studies, 384*(12), 1-17.
- *Lileikienė, A., & Danilevičienė, L. (2016). Foreign language anxiety in student learning. *Baltic Journal of Sport & Health Science, 3*(102), 18-23. <https://doi.org/10.33607/bjshs.v3i102.61>
- Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Sage Publications.
- MacIntyre, P. D. (1999). Language anxiety: A review of the research for language teachers. In D. J. Young (Ed.), *Affect in foreign language and second language learning: A practical guide to creating a low-anxiety classroom atmosphere* (pp. 24-45). McGraw-Hill College.
- MacIntyre, P. D. (2017). An overview of language anxiety research and trends in its development. In C. Gkonou, M. Daubney, & J.-M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 11-30). Multilingual Matters. <https://doi.org/10.21832/9781783097722-003>
- MacIntyre, P. D., & Gardner, R. C. (1989). Anxiety and second-language learning: Toward a theoretical clarification. *Language Learning, 39*(2), 251-275. <https://doi.org/10.1111/j.1467-1770.1989.tb00423.x>
- MacIntyre, P. D., & Gardner, R. C. (1991). Language anxiety: Its relationship to other anxieties and to processing in native and second languages. *Language Learning, 41*(4), 513-534. <https://doi.org/10.1111/j.1467-1770.1991.tb00691.x>

- *Matsuda, S., & Gobel, P. (2004). Anxiety and predictors of performance in the foreign language classroom. *System*, 32, 21-36.
- Mejías, H., Applebaum, R. L., Applebaum, S.J., & Trotter, R. T. (1991). Oral communication apprehension and Hispanics: An exploration of oral communication apprehension among Mexican American students in Texas. In E. K. Horwitz & D. J. Young (Eds.), *Language anxiety: From theory and research to classroom implications* (pp. 87-97). Prentice Hall.
- *Merzin, A., & Alandal, A. (2015). An investigation of anxiety among elementary school students towards foreign language learning. *Studies in Literature and Language*, 10(6), 1-11. <https://doi.org/10.3968/7180>
- Norris, J. M., & Ortega, L. (2006). *Synthesizing research on language learning and teaching*. John Benjamins Publishing Company.
- *Onwuegbuzie, A. J., Bailey, P., & Daley, C. E. (2000). Cognitive, affective, personality, and demographic predictors of foreign-language achievement. *The Journal of Educational Research*, 94(1), 3-15. <https://doi.org/10.1080/00220670009598738>
- *Özer, S. (2019). An investigation of attitude, motivation and anxiety levels of students studying at a faculty of tourism towards vocational English course. *Journal of Language and Linguistic Studies*, 15(2), 560-577. <https://doi.org/10.17263/jlls.586246>
- Öztürk, G., & Gürbüz, N. (2013). The impact of gender on foreign language speaking anxiety and motivation. *Procedia – Social and Behavioral Sciences*, 70, 654-665. <https://doi.org/10.1016/j.sbspro.2013.01.106>
- Park, G.-P. (2014). Factor analysis of the foreign language classroom anxiety scale in Korean learners of English as a foreign language. *Psychological Reports*, 115(1), 261-275. <https://doi.org/10.2466/28.11.PR0.115c10z2>
- *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>
- Plonsky, L., & Oswald, F. L. (2012). How to do a meta-analysis. In A. Mackey & S. M. Gass (Eds.), *Research methods in second language acquisition: A practical guide* (pp. 275-295). Wiley-Blackwell. <https://doi.org/10.1002/9781444347340.ch14>
- Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, 64(4), 878-912. <https://doi.org/10.1111/lang.12079>
- *Ra, J., & Rhee, K. J. (2018). Detection of gender related DIF in the foreign language classroom anxiety scale. *Educational Sciences: Theory & Practice*, 18(1), 47-60. <https://doi.org/10.12738/estp.2018.1.0606>
- *Rastegar, M., Akbarzadeh, M., & Heidari, N. (2012). The darker side of motivation: Demotivation and its relation with two variables of anxiety among

- Iranian EFL learners. *International Scholarly Research Notices*, 2012(1), 1-8. <https://doi.org/10.5402/2012/215605>
- *Rodríguez, Y., Delgado, V., & Colón, J. M. (2009). Foreign language writing anxiety among preservice EFL teachers. *Lenguas Modernas*, 33, 21-31.
- *Sankueana, W. (2018, July 14-15). Foreign language classroom anxiety of Thai high school students [Conference presentation]. 11th International Conference on Language, Literature, Culture and Education.
- Scovel, T. (1978). The effect of affect on foreign language learning: A review of the anxiety research. *Language Learning*, 28(1), 129-142. <https://doi.org/10.1111/j.1467-1770.1978.tb00309.x>
- Spielberger, C. D. (1966). Theory and research on anxiety. In C. D. Spielberger (Ed.), *Anxiety and behavior* (pp. 3-20). Academic Press.
- Spielberger, C. D. (1983). *Manual for the state-trait anxiety inventory (STAI Form Y)*. Consulting Psychologist Press.
- *Sultan, S. (2012). Students' perceived competence affecting level of anxiety in learning English as a foreign language. *Pakistan Journal of Psychological Research*, 27(2), 225-239.
- *Taghinezhad, A., Abdollahzadeh, P., Dastpak, M., & Rezaei, Z. (2016). Investigating the impact of gender on foreign language learning anxiety of Iranian EFL learners. *Modern Journal of Language Teaching Methods*, 6(5), 418-426.
- *Tanielian, A. R. (2017). Foreign language anxiety among first-year Saudi university students. *The International Education Journal: Comparative Perspectives*, 16(2), 116-130.
- Teimouri, Y., Goetze, J., & Plonsky, L. (2019). Second language anxiety and achievement: A meta-analysis. *Studies in Second Language Acquisition*, 41(2), 363-387. <https://doi.org/10.1017/S0272263118000311>
- *Wang, Y., & Jingna, L. (2011, May 27-29). The interference of foreign language anxiety in the reading comprehension of agricultural engineering students [Conference presentation]. 2011 International Conference on New Technology of Agricultural Engineering.
- *Wei, J., & Yodkamlue, B. (2012). The Chinese Bouyei College students' classroom anxiety in foreign language learning: A survey study. *International Journal of English Linguistics*, 2(2), 75-90. <https://doi.org/10.5539/ijel.v2n2p75>
- Woodrow, L. (2006). Anxiety and speaking English as a second language. *RELC Journal*, 37(3), 308-328. <https://doi.org/10.1177/0033688206071315>
- *Yamini, M., & Tahriri, A. (2006). On the relationship between foreign language classroom anxiety and global self-esteem among male and female students at different educational levels. *Iranian Journal of Applied Linguistics*, 9(1), 101-129.

- Yan, X. (1998). *An examination of foreign language classroom anxiety: Its sources and effects in a college English program in China* [Unpublished doctoral dissertation]. The University of Texas at Austin.
- *Zasiekina, L., & Zhuravlova, O. (2019). Acculturating stress, language anxiety and procrastination of international students in the academic settings. *Psycholinguistics*, 26(1), 126-40. <https://doi.org/10.31470/2309-1797-2019-26-1-126-140>
- *Zhang, L. J. (2001). ESL students' classroom anxiety. *Teaching and Learning*, 21(2), 51-62.
- Zhang, X. (2019). Foreign language anxiety and foreign language performance: A meta-analysis. *Modern Language Journal*, 103(4), 763-781. <https://doi.org/10.1111/modl.12590>
- *Zhang, Z.-H. (2019). An empirical study on foreign language classroom anxiety of college students. *Sino-US English Teaching*, 16(9), 376-386. <https://doi.org/10.17265/1539-8072/2019.09.003>

APPENDIX

The forest plot for the random effects model of FLCAS based on the standard difference in means

