Studies in Second Language Acquisition **41** (2019), 363–387
doi:10.1017/S0272263118000311

Research Article

SECOND LANGUAGE ANXIETY AND ACHIEVEMENT

A META-ANALYSIS

Yasser Teimouri*

Georgetown University

Julia Goetze

Washington University in St. Louis

Luke Plonsky

Northern Arizona University

Abstract

Second language (L2) anxiety has been the object of constant empirical and theoretical attention for several decades. As a matter of both theoretical and practical interest, much of the research in this domain has examined the relationship between anxiety and L2 achievement. The present study meta-analyzes this body of research. Following a comprehensive search, a sample of 97 reports were identified, contributing a total of 105 independent samples ($N = 19{,}933$) from 23 countries. In the aggregate, the 216 effect sizes (i.e., correlations) reported in the primary studies yielded a mean of $r = -.36$ for the relationship between L2 anxiety and language achievement. Moderator analyses revealed effects sizes to vary across different types of language achievement measures, educational levels, target languages, and anxiety types. Overall, this study provides firm evidence for both the negative role of L2 anxiety in L2 learning and the moderating effects of a number of (non)linguistic variables. We discuss the findings in the context of theoretical and practical concerns, and we provide direction for future research.

Anxiety is arguably the most researched affective variable in the field of second language acquisition (SLA). For more than half a century, SLA researchers have been conducting studies to explore second language (L2) learners' anxiety in relation to a host of linguistic and nonlinguistic variables (for reviews, see Gkonou, Daubney, & Dewaele, 2017;

The experiment in this article earned an Open Data badge for transparent practices. The materials are available at www.iris-database.org/iris/app/home/detail?id=york%3a935928&ref=search.

*Correspondence concerning this article should be addressed to Yasser Teimouri, Georgetown University, 1421 37th Street NW, Washington, DC, 20057-1051. E-mail: yt329@georgetown.edu

Copyright © Cambridge University Press 2019

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

Horwitz, 2001, 2017; MacIntyre, 1999; 2017; MacIntyre & Gardner, 1993; Oxford, 1999). Study findings have largely attributed a negative role to anxiety in the L2 learning process, claiming it impairs L2 learners' language achievement (e.g., Aida, 1994; Gardner, Smythe, Clement, & Gliksman, 1976; Horwitz, 2017; MacIntyre, 2017; MacIntyre & Gardner, 1994a; Young, 1986), cognitive processing (e.g., MacIntyre & Gardner, 1994a, 1994b), motivation (e.g., Papi, 2010; Papi & Teimouri, 2014; Teimouri, 2017), and willingness to communicate in a second language (L2 WTC) (e.g., Hashimoto, 2002; Khajavy, MacIntyre, & Barabadi, 2018; MacIntyre & Charos, 1996).

Considering the rapid theoretical and empirical growth of L2 anxiety research following Horwitz, Horwitz, and Cope's (1986) seminal article and given the now-widespread application of research synthesis in SLA (e.g., Norris & Ortega, 2006; Oswald & Plonsky, 2010; Plonsky & Oswald, 2012), it is surprising that no meta-analysis of L2 anxiety research has been conducted to date. Although four traditional literature reviews on the topic of L2 anxiety and language achievement have been published (Horwitz, 2001; MacIntyre, 2017; MacIntyre & Gardner, 1991; Scovel, 1978), none of them has provided a systematic analysis of the relationship between the variables in question.

The aim of the current study is to present a comprehensive, systematic, and quantitative meta-analysis of the relationship between L2 anxiety and language achievement in SLA. In addition, a number of moderator variables are examined to explain variability in observed correlations. In short, this meta-analysis aims to (a) consolidate findings of previous anxiety research while (b) addressing new research questions of relevance to theory and practice that have been rarely investigated in past research but that can be investigated from a meta-analytic perspective.

REVIEW OF THE LITERATURE

Generally referred to as "the subjective feeling of tension, apprehension, nervousness, and worry associated with an arousal of the autonomic nervous system" (Spielberger, 1983, p. 1), anxiety has been traditionally conceptualized and measured as either a trait, state, or situation-specific construct in SLA (e.g., Chen, Gully, Whiteman, & Kilcullen, 2000; Horwitz, 2017; Spielberger, Gorsuch, & Lushene, 1970). Early research defined L2 anxiety in general terms as either a relatively stable personality trait across various situations, or as a transient emotional state, manifesting itself in a particular moment in time (e.g., during a high-stakes exam) (Cattell & Scheier, 1961). Over time, however, SLA researchers have increasingly adopted a situation-specific definition of anxiety, postulating that it "occur[s] consistently over time within a given situation" (MacIntyre & Gardner, 1991, p. 87), such as during language learning in a recurrent classroom setting, thereby distinguishing it from more general anxiety types (e.g., Gardner, 1985; Horwitz, 2017; Horwitz, Horwitz, & Cope, 1986). In other words, within the emerging situation-specific framework, L2 anxiety was reconceptualized in relation to different L2 settings (e.g., in a classroom, while using the target language).

Among the first situation-specific L2 anxiety constructs to be proposed in SLA was *foreign language classroom anxiety* (FLCA), which is defined 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" (Horwitz et al., 1986,

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

p. 128). Along with their domain-specific definition, Horwitz et al. (1986) introduced the Foreign Language Classroom Anxiety Scale (FLCAS), a 33-item, 5-point Likert-scale questionnaire, which measures FLCA during L2 learning. Empirical findings supported the internal consistency of the newly developed FLCAS and substantiated the authors' claim of this language-specific type of anxiety, further distinguishing it from other types of anxieties, such as communication apprehension, test anxiety, or fear of negative evaluations (Horwitz, 1986).

The introduction of FLCA as a distinct type of anxiety subsequently led to even more nuanced definitions of language-related anxieties in the field. For example, some researchers took the FLCAS as a point of departure to develop more language skill-specific, self-report scales such as the Foreign Language Reading Anxiety Scale (FLRAS) (Saito, Garza, & Horwitz, 1999), the Foreign Language Listening Anxiety Scale (FLLAS) (Elkhafaifi, 2005), and the Second Language Writing Anxiety Scale (Cheng, 2004; Cheng, Horwitz, & Schallert, 1999); Cheng's (2017) anxiety questionnaire, including all four language skills, should also be noted. Others developed self-report instruments to measure L2 learners' anxiety separately within the three stages of input, processing, and output of language learning, as proposed in several cognitive models in SLA (MacIntyre & Gardner, 1994a, 1994b). Overall, anxiety research in SLA developed a variety of nuanced ways to conceptualize and measure L2 anxiety, its characteristics, causes, and effects.

ANXIETY AND LANGUAGE ACHIEVEMENT IN FL/L2 RESEARCH

L2 researchers first began investigating the relationship between anxiety and language achievement in the 1970s. Despite the use of established measurement tools from educational psychology (e.g., Chastain, 1975; Kleinmann, 1977), these early investigations produced an array of conflicting findings. For instance, in one of the first anxiety studies in SLA, Chastain (1975) used the TAS (Test Anxiety Scale) and the MAS (Manifest Anxiety Scale) to measure anxiety and correlate the scale scores with students' final course grades in French, German, and Spanish classes. He found correlations between anxiety scores and final grades in all three languages but was unable to show consistency in the direction of the correlations within or across languages. Similarly, other researchers, such as Kleinmann (1977), turned to instruments from education, using the facilitating/debilitating dichotomy to investigate the effects of anxiety on speech skills of 39 university students of English as a second language. The study showed that debilitative anxiety led students to avoid grammatical structures, whereas facilitative anxiety did not.

The inconsistency of these findings was addressed by Scovel's (1978) seminal review of existing anxiety studies in L2 research, which is often considered a turning point in this domain (Horwitz, 2010). In his review, Scovel attributed the variability of study results to theoretical and methodological imprecisions and cautioned researchers to carefully consider and report the type of anxiety (i.e., state or trait), as well the directionality of the observed effect. Scovel's call for action and scientific rigor led Horwitz and her associates to define FLCA and to introduce FLCAS, which has resulted in more stable empirical findings and an overall more systematic investigation of anxiety in SLA (e.g., Horwitz, 2001; MacIntyre, 2017; MacIntyre & Gardner, 1994). For instance, researchers

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

have found a negative relationship between anxiety and achievement across different target languages, including Japanese, French, English, German, and Spanish (e.g., Aida, 1994; Phillips, 1992; Saito et al., 1999; Young, 1986).

Although no systematic syntheses or meta-analyses of this domain have been conducted, we would highlight three comprehensive literature reviews that target the relationship between anxiety and L2 achievement (Horwitz, 2001; MacIntyre, 2017; MacIntyre & Gardner, 1991). MacIntyre and Gardner (1991) focus predominantly on the methodological and theoretical development of anxiety research in SLA. More specifically, the authors present and define the three existing anxiety constructs described previously (*state*, *trait*, and *situation-specific* anxiety), outline the research traditions from which these constructs originated, and provide an overview of established measurement tools for each anxiety type. While their concluding discussion of selected studies of the anxiety-achievement relationship mentions the "negative effect" of anxiety on the language learning process (p. 112), it largely omits specific statistical findings and predominantly serves to highlight the general utility of situation-specific approaches to anxiety studies in SLA.

In contrast to MacIntyre and Gardner, Horwitz (2001) concentrates on the discussion of a handful of studies to stress the relative uniformity of findings observed since the introduction of the situation-specific FLCA construct. For instance, she presents individual study findings across target languages and instructional levels (e.g., Aida, 1994; Saito & Samimy, 1996), concluding that there is evidence for a negative relationship between language anxiety and language achievement. The precise extent of this relationship, however, remains unclear. Horwitz also discusses potential sources of anxiety, such as learners' proficiency level or instructional approaches, and alludes to the promising potential of skill-specific anxiety measures, which were also developed on the assumption of situation specificity.

MacIntyre (2017) reviews L2 anxiety research and classifies studies according to the three distinct phases in the domain's historical trajectory: the confounding phase, the specialized phase, and the dynamic phase. The confounding phase represents early research on anxiety in which scholars' use of general anxiety measures generated inconsistent findings concerning the role of anxiety in the language achievement. The specialized phase refers to research built on the new reconceptualization of anxiety, promoted by Gardner (1985) and Horwitz et al. (1986). According to MacIntyre, during this phase, instruments were developed to measure anxiety within the context of L2 learning, a development that brought a sense of unity among scholars. Finally, the dynamic phase characterizes research that attempts to explore how L2 anxiety continuously interacts with a host of other factors over time in specific contexts. MacIntyre concludes his review by stating that (a) L2 anxiety has a generally negative role in L2 learning, (b) L2 anxiety is both a cause and consequence of negative performance, and (c) L2 anxiety has both internal and social dimensions.

Certainly, all these three literature reviews—each with its own specific goals—have provided informative and valuable overviews of the state of knowledge regarding different aspects of L2 anxiety. However, to form a more comprehensive and precise view of L2 anxiety and its relationship with language achievement, we intend to extend these reviews in (at least) four specific areas.

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

First, all three reviews are based on a selection of studies that is incomplete and/or at least somewhat subjectively determined. Thus, these reviews lack the systematicity necessary to gain a comprehensive view of the domain and to draw generalizable or precise conclusions. Second, although the three reviews address the use of a variety of anxiety definitions and instruments across studies (e.g., FLCAS, FLRAS), the use of a variety of different language achievement measures (e.g., course grades, language tests, GPA) in those studies is not given full and proper consideration. As a result, any inconsistencies in study findings are attributed to the anxiety instruments rather than to the equally critical achievement measures. Third, all three reviews arrive at the conclusion that there is a negative correlation between anxiety and achievement, but the precise strength of that relationship is not addressed. Fourth, none of the existing reviews systematically investigates variables that may moderate the relationship between anxiety and achievement, such as learning contexts, educational level, target language, and so forth.

OBJECTIVES OF THE CURRENT STUDY

The current study has three goals, which seek to build on and extend earlier reviews. First, we conduct a comprehensive and systematic literature search of the domain to describe the variety of methodological and reporting practices involved in research that has examined the relationship between L2 learners' anxiety and their language achievement. That is, we provide a detailed description of different types of measures used to assess both L2 anxiety and language achievement. Second, we focus on both the directionality (positive vs. negative) and magnitude of the relationship between anxiety and students' linguistic achievement measures, as reported in L2 anxiety research. Third, looking beyond the overall meta-analytic result, we investigate the roles of a number of contextual and language-related factors as potential moderators of the relationship between students' anxiety and their language achievement. With these goals in mind, we aim to answer the following research questions:

1. What are the methodological and reporting practices involved in studies of the relationship between L2 anxiety and achievement?
2. What is the direction and the size of the relationship between L2 anxiety and learners' language achievement?
3. How do different language achievement measures, learning contexts, L2 anxiety types, and educational levels moderate the relationship between language anxiety and language achievement?

METHOD

LITERATURE SEARCH

We employed a variety of approaches to identify relevant literature in the L2 anxiety domain (see Plonsky & Brown, 2015). Both published and unpublished studies were eligible for inclusion to form a more representative sample and to minimize the potential influence of publication bias in favor of studies with statistically significant results (Konstantopoulos & Hedges, 2004; Norris & Ortega, 2006). Initially, we searched several

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

online databases, including Education Resources Information Clearinghouse (ERIC), Linguistics and Language Behavior Abstracts (LLBA), ProQuest, and Google Scholar, using a combination of the following key terms: *anxiety*, *language*, *second (L2)*, *foreign (FL)*, *classroom*, *communication*, *use*, *emotion*, *affect*, and *achievement*. We manually searched the websites of five journals: S*tudies in Second Language Acquisition*, *Language Learning*, *Modern Language Journal*, *Applied Linguistics*, and *TESOL Quarterly*.

We also conducted backward and forward citation searches based on seminal and more recent articles. Finally, the reference sections of review papers and studies that have examined the relationship between L2 anxiety and language achievement were examined (e.g., Horwitz, 2001; MacIntyre, 2017; MacIntyre & Gardner, 1991). Overall, our search returned more than 700 documents relating to L2 anxiety (including duplicates).

INCLUSION/EXCLUSION CRITERIA

To determine a study's relevance to our research questions and inclusion in our sample, the title and abstract of each document were read carefully and three eligibility criteria were applied. First, a study had to contain, at a minimum, a measure of language anxiety and a measure of language achievement. Second, a study had to report on the relationship between language anxiety and language achievement expressed as a correlation or another statistic that can be transformed into a correlation (e.g., t or F). For instance, studies using multiple regression analyses or structural equation modeling as their main analytical procedures were excluded unless the correlation matrices that went into the analyses were reported. Third, a study had to have been conducted following publication of Horwitz et al.'s (1986) seminal article. This date restriction was imposed because anxiety research before 1986 was criticized mostly for its unsystematic investigation of L2 anxiety. After applying the eligibility criteria to our initial search results, a final pool of 97 reports (89 journal articles and 8 unpublished dissertations) was retained for further analysis.

CODING PROCEDURES

A coding scheme was developed to systematically capture the numerous characteristics of each study in our sample. Specifically, we determined five broad categories of features: (a) bibliographic information (e.g., authors), (b) study design (e.g., context), (c) the sample (e.g., age), (d) instrumentation (e.g., language achievement instrument, anxiety scale), and (e) effect sizes (e.g., correlations). Table 1 provides a detailed description of the coding scheme. The items in our coding scheme were designed to enable us to describe the different features of studies in this domain and to examine some of those features as potential moderators of the overall correlation in question. The coding scheme underwent several rounds of revision after it was piloted by the researchers, who each coded five randomly chosen papers from the sample database. An additional rater coded a subset of 10 studies to ensure the reliability of the obtained data. The inter-coder reliability was perfect (100%), which is unusual but not unexpected, given the low inference nature of the items in the coding scheme. The final version of the coding scheme is available on IRIS (iris-database.org; see Marsden, Mackey, & Plonsky, 2016).

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

TABLE 1. Variables and definitions of items in the coding scheme

Variable	Definition
Bibliographic Information	
Author(s)	Researchers who conducted and published the study
Title of publication	The title of the report
Year of publication	Year in which the study was published
Type of publication	Journal article, dissertation, book chapter, conference proceeding or presentation, unpublished manuscript, working paper
Title of publication venue	Journal or volume in which the study was published
Study Design	
Study design	Experimental or observational design
Methodology	Quantitative or mixed methodology
Study context	L2, FL, combined FL/L2, or study abroad (SA)
Country	Country in which the research was conducted
Sample	
Participants' age	Mean age or age range of the participants
Participants' nationality	Nationality of the participants
Participants' L1	Native language of the participants
Participants' L2 (target language)	Language in which achievement outcome measures were elicited
Participants' academic status	Educational level of participants (elementary school, middle school, high school, university/college, language institute)
Participant sample size	Number of participants in the study
Participants' proficiency level	Proficiency level (beginning, intermediate, advanced, multiple, or not specified)
Participants' language learning experience	Average time that participants had been learning the target language
Instrumentation (Anxiety)	
Conceptualization	State, trait, or situation-specific anxiety
Type	General, skill-specific, or other type of anxiety
Terminology	Terminology that researchers used when referring to anxiety (e.g., Foreign Language Classroom Anxiety)
Instrument	Whether the study borrowed, adapted, or developed an instrument
Instrument type	Whether the measurement tool was a questionnaire or another type of instrument
Creator of instrument	Only applicable when study borrowed an instrument
Items	Number of items in the instrument
Likert scale	Number of Likert-scale points on the instrument (e.g., 4-point Likert-scale)
Pilot	Whether the measure was piloted
Mean	The reported mean for the anxiety measure
Standard deviation (SD)	The reported SD for anxiety measure
Reliability	Whether any measures of reliability were reported for the anxiety instrument
Reliability index	If reliability was reported, what reliability index was used (e.g., alpha, test-retest)
Reliability estimate	Reliability coefficient as reported

(continued on following page)

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

TABLE 1. Variables and definitions of items in the coding scheme (continued)

Variable	Definition
Instrumentation (Language Achievement)	
Type	The type of achievement measure used in the study (course grade, language test, self-perceived competence, GPA)
Measurement source	Whether the achievement measure was adopted/adapted or newly developed
Pilot	Whether the achievement measure was piloted
Measurement level	The range of possible scores for each achievement measure
Mean	The reported mean for the achievement measure
Standard deviation (SD)	The reported SD for the achievement measure
Analysis type	Type of inferential statistical test used
Reliability	Whether measures of reliability were reported for the achievement tool
Reliability index	If reliability was reported, what reliability index was used (e.g., alpha, test-retest)
Reliability estimate	Reliability coefficient as reported
Result	
Reported effect size (r)	Observed correlation between anxiety and achievement

ANALYSIS

To address RQ1, frequency counts for different study features were calculated across the sample. Correlation coefficients (r) were used as the basis of all the analyses pertaining to RQ2 and RQ3. Because a number of studies failed to report the reliability of the instruments used in their research, it was not possible to apply psychometric corrections to observed correlations for the attenuating effects of measurement error. However, all reported effect sizes were weighted based on individual sample sizes, and confidence intervals for each effect size were calculated and presented (see Appendix A).

Many of the studies in our sample ran multiple correlations between language anxiety measures and language achievement measures. In such cases, all effect sizes pertaining to the same participants were averaged to form a single effect size for each sample. However, correlations based on independent samples within an individual study were not combined but rather treated separately (Lipsey & Wilson, 2001; Plonsky & Oswald, 2015). To answer RQ2—concerning the direction and the size of the overall relationship between language anxiety and language achievement—the weighted mean of all effect sizes along with its 95% confidence intervals was computed.

To address RQ3, which pertained to potential moderators of the strength of the relationship between L2 anxiety and language achievement, effect sizes were computed based on the groups of variables identified *a priori* (i.e., the variables in the coding scheme) to act as moderators of the relationship between anxiety and language achievement. We additionally tested the validity of our assumptions underlying the selection of moderators by conducting homogeneity analyses (Cooper, Hedges, & Valentine, 2009; Rosenthal, Cooper, & Hedges, 1994); we computed a within goodness-of-fit statistic (Q_w) to examine whether the observed variance among the set of observed effects

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

sizes exceeded the expected sampling errors. A significant Q_w statistic signals the existence of systematic variations among observed effect sizes, thereby verifying our moderator assumptions.

RESULTS

The 97 studies in our sample were conducted between 1985 and 2017 in 23 countries. A total of 105 independent samples and 19,933 individual students participated in these studies. Sample sizes ranged from 12 to 1,174 ($M = 190$; $SD = 215.5$), and a total of 216 effect sizes (r) were collected based on 105 independent samples.

The first research question addressed methodological and reporting practices. In our sample, anxiety was measured by means of 25 different questionnaires, capturing trait, state, and situation-specific anxiety types. While the majority of these questionnaires were adapted and modified from existing instruments (e.g., Horwitz et al., 1986), a handful of studies developed new instruments designed for their own research context. As expected, Horwitz et al.'s (1986) questionnaire, the FLCAS, is by far the most frequently used questionnaire in our sample, with more than half the studies using it as their main anxiety measurement tool. Out of the 129 times an anxiety questionnaire was administered, reliability was reported 84 times (65%). All studies reported alpha as the index of reliability; only one study conducted a test-retest reliability analysis. Taken together, reliability coefficients of all the language anxiety measures reached a mean of .88 (Table 2). Table 3 also shows reliability coefficients of the three most frequently used L2 anxiety measures.

Of a total of 170 different language achievement measures, 90 were language tests, 40 were course grades, 35 were self-assessments, and 5 were GPAs. Table 4 provides a detailed description of the language achievement measures. As can be seen, the measurement level of all the achievement instruments—the range of possible scores for each language achievement measure—has been reported 69 times (40.6%) in our sample. Furthermore, the mean scores of the achievement measures were only reported 43 times (25.3%), indicating a lack of reporting of basic descriptives.

Considering the characteristics of the language tests and self-assessments of language proficiency, approximately one third of the measures (34%) were adopted/adapted from other resources, and 11 measures (8.8%) were newly developed; no information was provided regarding the source of more than half of the measures. In addition, only two studies reported that they had piloted the test before its administration in the main research study. Finally, of the 125 different kinds of tests, both language proficiency tests and self-assessments, reliability coefficients for only one fifth of the measures were reported. Table 5 shows a detailed description of the reliability analyses of language achievement measures.

TABLE 2. Reliability indices of L2 anxiety measures

Instruments	*K*	Reported	Missing	Min	Max	Mean	SD
All L2 anxiety measures combined	129	84	45	.65	.97	.88	.06

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

TABLE 3. Reliability indices (Cronbach's Alpha) of the three most frequently used measures of L2 anxiety

Instruments	*K*	Reported	Missing	Min	Max	Mean	SD
Horwitz et al., 1986	67	40	27	.75	.97	.90	.05
Saito et al., 1999	11	8	3	.82	.96	.87	.04
Gardner, 1985	9	6	3	.86	.92	.88	.02

The second research question was concerned with the overall direction and magnitude of the relationship between anxiety and L2 achievement. To answer this research question, we calculated the weighted mean of all effect sizes in our sample: $r = –.36$. As an alternate take on this overall result, we can also say that language anxiety accounts for 13% of the variance in L2 achievement scores. Before arriving at this overall result, we examined the distribution of correlations for outliers (defined as standard residuals > 2.5), resulting in the exclusion of nine values. This step increased the overall result (from $r = –.33$); excluding these outliers also increased the precision of the average effect (i.e., narrowed confidence intervals), reduced sampling errors proportionally (i.e., change of Q from 1046.9 to 555.3), and ironed out some of the inconsistencies between the observed effect sizes (i.e., change of I^2 from 90.1% to 82.9%). With these improvements to the statistical model in mind, the outliers were excluded from all subsequent analyses.

TABLE 4. A description of language achievement measures used in L2 anxiety research

	Course grades		Language tests		Self-assessments		GPAs		Total	
Measurement	*k*	%	*k*	%	*k*	%	*k*	%	*k*	%
Adopted/Adapted	–	–	29	32.2	14	40	–	–	43	34
Developed	–	–	8	8.9	3	8.6	–	–	11	8.8
NR	–	–	53	58.9	18	51.4	–	–	71	57.2
Measurement level										
Reported	11	27.5	32	32.9	26	78.1	0	0	69	40.6
NR	29	72.5	58	67.1	9	21.9	5	100	101	59.4
Piloted										
Yes	–	–	2	2.2	0	0	–	–	2	2.4
No	–	–	88	97.8	35	100	–	–	123	97.6
Mean										
Reported	7	17.5	31	34.4	4	11.4	1	20	43	25.3
NR	33	82.5	59	65.6	31	88.6	4	80	127	74.7
SD										
Reported	6	15	29	32.2	3	8.6	1	20	39	22.9
NR	34	85	61	67.8	32	91.4	4	80	131	77.1
Total	40	23.3	90	53.5	35	20.1	5	3.1	170	100

Note: NR = not reported.

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

TABLE 5. Reliability estimates of language achievement measures

Measures	*K*	Reported	Missing	Median	Mean	SD
Language tests	90	10	80	.90	.85	.13
Self-perceived proficiency	35	13	22	.89	.87	.06
Total	125	23	102	.89	.86	.10

PUBLICATION BIAS

Although both published and unpublished studies were included, a funnel plot (i.e., a scatter plot of the relationship between effect size and standard error) was created to assess the presence of a publication bias. As seen in Figure 1, it could be argued that the funnel plot indicates a slight bias in our sample. However, as indicated by the additional "missing" studies on the left side, the bias is in favor of studies with smaller effects. This publication bias, though slight, can also be seen in the results concerning the overall mean of effect sizes for published articles and unpublished dissertations. To assess the effects of publication bias on the overall effect size, we employed Duval and Tweedie's (2000) trim-and-fill procedure. Based on the results of this procedure, the overall effect size decreased slightly (from $r = -.36$ to $r = -.30$) when the data for missing studies on the right side of the funnel plot were imputed. However, the overall effect size remained stable when the missing studies for the left side of the funnel were imputed.

The third research question was concerned with variables that may moderate the relationship between anxiety and language achievement. A test of homogeneity of effect

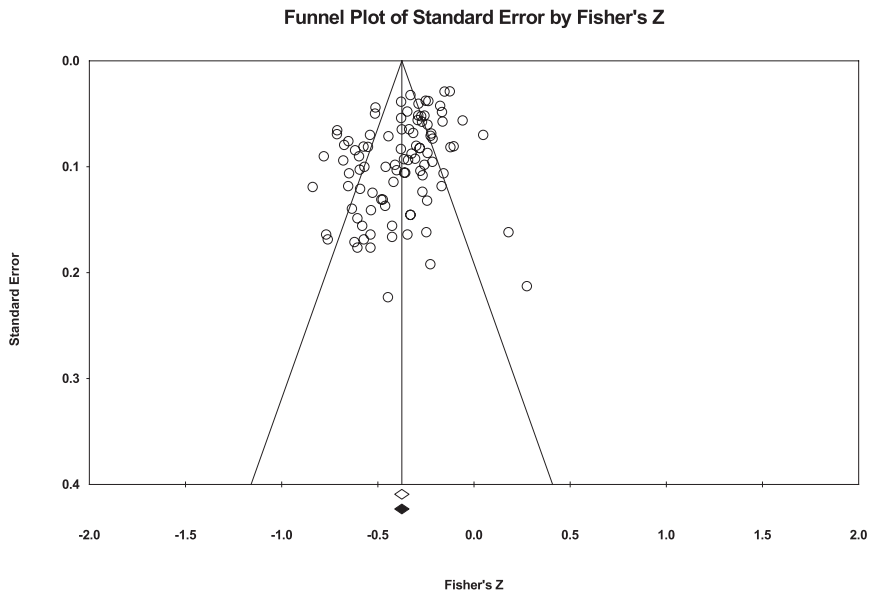


FIGURE 1. The funnel plot of unweighted effect sizes and sampling errors.

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

TABLE 6. Overall correlation between L2 anxiety and language achievement measures

				95% Confidence Intervals			
	K	*N*	M_r	Lower	Upper	*Q*	I^2
L2 achievement	96	17,403	–.36	–.39	–.33	555.27*	82.89%

Note: * = $p < .001$.

sizes was conducted to test for the presence of moderators. As can be seen in Table 6, the within goodness-of-fit statistic (Q_w) is significant, signaling the existence of systematic variations among observed effect sizes. In addition, the I^2 statistic suggests that about 83% of the observed variation among effect sizes was real and unrelated to sampling error. These statistics indicate heterogeneity of the effect sizes and justify a closer look at the potential role of additional variables (i.e., moderators) that might help explain variability in observed correlations.

To conduct the moderator analysis for this study, subgroups of effects sizes were computed based on five categories: the type of language achievement measure, study context, educational level, target language, and L2 anxiety type. Table 7 provides a detailed description of the subgrouped effect sizes in each category.

We first examined variability in the correlation between anxiety and achievement as a function of different measures of latter. The results show that the strongest relationship is observed for self-perceived measures and weakest for GPA, with course grades and objective tests falling in between.

With respect to the context of the primary studies, the correlations between language anxiety and language achievement were found to be slightly larger in second language contexts than in foreign language contexts. The moderator analysis also showed different patterns in relation to the educational levels of the students. The correlation is strongest among elementary school students and the weakest for students in private language institutes. While the strength of the relationship between language anxiety and language achievement drops from elementary school to junior high school, there is a slight upward trend from junior high school to senior high school to college. Given the limited number of studies that included samples from elementary and junior high school, this pattern of results should be interpreted with caution, however. Excluding these studies from our analyses (because of their limited power) and focusing on effect sizes of senior high school and college students, we notice a negligible increase in effect sizes from high school to college/university levels.

The results of subgroup analyses also showed a substantial difference in correlations based on target language. The negative relationship between language anxiety and language achievement is stronger when the target language is not English. Effect sizes were also found to vary in relation to the types of language anxiety. Situation-specific anxiety revealed larger correlations with achievement than trait and state anxieties. Moreover, a larger effect was found for state anxiety than for trait anxiety. These results,

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

TABLE 7. The results of moderator analyses

				95% Confidence Intervals	
	K	*N*	M_r	Lower	Upper
Achievement measure					
Self-perceived	15	3185	–.47	–.54	–.40
Course grades	33	7101	–.34	–.39	–.29
Language tests	56	1047	–.36	–.39	–.32
GPAs	5	878	–.26	–.50	–.02
Study Context					
Study abroad	2	214	–.49	–.66	–.37
Second language	21	2846	–.40	–.46	–.33
Foreign language	72	15973	–.35	–.39	–.31
Missing	1	25			
Educational Level					
Elementary	3	832	–.43	–.79	–.06
Junior high school	8	2053	–.34	–.49	–.19
Senior high school	27	5508	–.34	–.41	–.27
College	53	9811	–.37	–.42	–.32
Institute	2	518	–.25	–.65	.15
Missing	3	336			
Target L2					
English	62	15666	–.33	–.38	–.29
Non-English	34	3392	–.39	–.46	–.33
Anxiety types					
Listening	4	2133	–.46	–.65	–.27
Writing	6	2174	–.41	–.52	–.29
Speaking	73	13958	–.39	–.42	–.36
Reading	8	1371	–.38	–.47	–.29
State	2	204	–.36	–.49	–.24
Trait	8	572	–.20	–.33	–.06
Test	2	848	–.11	–.84	.62
Publication type					
Articles	88	18085	–.35	–.38	–.31
Dissertation	8	973	–.44	–.57	–.31

however, should also be interpreted cautiously due to the limited number of studies in trait and state subcategories ($k = 8$ and $k = 2$, respectively). Finally, considering skill-specific language anxieties, listening anxiety and writing anxiety showed much larger effects than reading or speaking anxiety.

DISCUSSION

A vast body of research has accumulated in the last three decades examining the relationship between anxiety and language achievement. The current study sought to provide a systematic synthesis and meta-analysis of the results of those studies. Our first research question probed methodological and reporting practices found in primary

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

reports in this domain. A variety of language anxiety instruments as well as language achievement measures were found to have been used in L2 anxiety research.

The majority of anxiety questionnaires fall under the category of situation-specific anxiety (Horwitz et al., 1986), suggesting that the domain of L2 anxiety research has consistently progressed toward its own theoretical understanding of anxiety, operationalization, and application in SLA. Moreover, most of the research studies reported the reliability coefficients of their anxiety scales, which are generally high and slightly larger than the mean reliability estimates for instruments used in SLA research (see Plonsky & Derrick, 2016). Overall, we can have a certain amount of confidence in the internal consistency of L2 anxiety instruments used in SLA research that measure the anxiety of L2 learners. Two caveats, however, should be added here. First, the reliability coefficient of a scale can be inflated when it contains a large number of items, as in the case of the 33-item FLCAS (Horwitz et al., 1986). Second, high-reliability alpha of a measure should not be interpreted as evidence of its validity. Sparks and Patton (2013), for instance, have questioned the validity of FLCAS by arguing that FLCAS is likely to be measuring students' language skills rather than their language learning anxiety (see also Teimouri, 2018).

Language achievement measures paint a different picture, however. Despite the use of various language achievement measures in primary studies, few studies have carefully examined the measurement characteristics of their achievement tests. While two thirds of studies reported reliability estimates for anxiety measures, less than one third did so for language achievement tests. Most of the studies also failed to report basic information concerning their language achievement measures, including descriptive statistics, a perennial weakness in L2 research (e.g., Larson-Hall & Plonsky, 2015; Paquot & Plonsky, 2017; Plonsky, 2013). In addition, more than half of the studies failed to state the source of their language achievement measures, that is, whether the language test was adopted/adapted or newly developed. Moreover, a very limited number of studies ($k = 2$) piloted their language tests to examine their validity or reliability before use in the main study (see Derrick, 2016). Taken together, we can see a lopsided approach toward assessing and reporting the measurement characteristics of instruments in anxiety research. Whereas validity and reliability issues underlying anxiety measures are generally attended to and thoroughly reported, a much lower standard of rigor and transparency is evident regarding the development and psychometric properties of language achievement measures in this domain.

The second research question explored the nature and strength of the relationship between learners' L2 anxiety and language achievement. As noted, anxiety research hit a number of theoretical and methodological obstacles at the early stages of its introduction to and application in SLA. Of critical significance early on was to determine the direction of the correlation—negative versus positive—generating a long-lasting discussion on the facilitating versus debilitating role of anxiety (MacIntyre, 2017). The introduction of FLCA (Horwitz et al., 1986) and the tool designed to measure it (FLCAS), however, offered a new theoretical direction for researchers working in this domain. Attributing the confusing results of past research on L2 anxiety to theoretical and methodological issues, Horwitz, along with other eminent anxiety researchers, has repeatedly argued for the overall negative effects of anxiety on students' language achievement and cautioned against the idea of facilitative anxiety (e.g., Horwitz, 2017; MacIntyre, 2017). The results

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

of this meta-analysis lend empirical support to their argument: Anxiety, overall, is negatively associated with L2 achievement.

As noted previously, it is not just the direction but the magnitude of this relationship that matters. The results of the present study revealed the correlation between language anxiety and achievement to be $r = -.36$, with 95% confidence that the correlation falls within the interval of $r = –.33$ to $r = –.39$. That is, a summary estimate collapsing across different language anxiety and achievement measures shows that L2 anxiety accounts for approximately 13% of variance on average in students' language achievement. Considering Plonsky and Oswald's (2014) guidelines for interpreting effect sizes in L2 research, a correlation of $r = -.36$ falls squarely into the range of what we would consider a medium or moderate effect. A more informative interpretation of this effect size, however, can be achieved by comparing this statistical value with the results of similar meta-analyses in related fields. In the general field of education, for instance, Seipp (1991) meta-analyzed research reports examining the relationship between students' anxiety and their academic performance (i.e., course grades and academic tests) in subjects other than language learning. The results showed a correlation of $-.21$. Likewise, Ma's (1999) synthesis of research on the relationship between students' anxiety toward mathematics and their achievement in this subject (i.e., course grades and math tests) uncovered a correlation of $-.27$. Considering these values in nearby disciplines, it can be argued that the detrimental effects of anxiety on students' achievements are stronger in the context of second/foreign language learning.

We might also consider the strength of observed effects in the present study as compared to other psychological variables associated with L2 learning and use. As we might expect, a substantially stronger (and positive) relationship to L2 achievement has been found for L2 aptitude: $r = .49$ (Li, 2016). However, the strength of the association between L2 anxiety and achievement either nearly matches or outpaces two other individual differences that have been examined frequently in primary studies and, more recently, at the meta-analytic level, as shown in Table 8: motivation (Masgoret & Gardner, 2003) and working memory (Linck, Osthus, Koeth, & Bunting, 2014).

Examining the role of individual differences in this manner provides compelling evidence for the role of learner-internal mechanisms in L2 development. In their aggregate, it would appear that the four individual difference variables for which we have meta-analytic evidence explain a total of 58% of the variance in L2 achievement.

TABLE 8. Relationships between individual differences and L2 achievement at the meta-analytic level

Variable	Correlation (r)	POV explained
Aptitude	.49	24%
Motivation	.37	14%
Anxiety	–.36	13%
Working memory	.26	7%

Notes: POV = percentage of variance in L2 achievement explained by each individual difference variable; calculated as (r^2) * 100.

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

However, this estimate is perhaps misleading on at least two accounts. First, some of the variance explained by these individual differences is likely shared, leading to a somewhat inflated overall value (see Plonsky & Ghanbar, 2018; Plonsky & Oswald, 2017, for discussions on shared variance among predictors). Most researchers working in this area accept, for example, that working memory is associated with or perhaps a component of L2 aptitude. This position is supported by Li (2016) who found very small to moderate meta-analytic correlations between working memory and aptitude, depending on how the constructs were defined and measured. Small correlations were also found between aptitude and motivation ($r = .14$) and between aptitude and anxiety ($r = –.17$). At the same time, one could also argue that observed effects in this domain underestimate true effects. The error in measures of achievement as well as psychological variables attenuates observed correlations. Correcting for measurement error and other psychometric artifacts (e.g., range restriction), though not common in L2 research, would show these and other correlations to be stronger than observed at the primary and secondary levels.

The third research question addressed the influence of moderators on the relationship between language anxiety and language achievement. The results showed a number of differences in effect sizes in relation to the use of different types of language achievement measures. Of four different types of language achievement measures identified and coded, anxiety showed the strongest relationship with self-perceived measures of L2 competence. In other words, anxiety has the strongest negative effects on L2 leaners' subjective evaluation of their own L2 proficiency. This finding is in line with the results of previous research that revealed that anxious learners are more inclined to underestimate their language proficiency than less anxious students. MacIntyre et al. (1997) found that anxious students tend to underestimate their language competence in relation to certain tasks, while objective measures of their performance in those tasks showed higher scores. Anxiety, by contrast, showed the weakest correlations with GPAs. Because GPAs represent an average score of students' achievements in different subjects that may not be related to the target language, and because the language anxiety scale is designed specifically to measure students' anxiety in relation to a second language, a weaker relationship is to be expected.

Finally, the negative effects of anxiety are almost the same in relation to both course grades and objective language tests. The use of course grades or language tests as measures of language proficiency has been a topic of debate among SLA researchers for quite some time: Some scholars advocate for the use of course grades over language tests, whereas others criticize this approach for its lack of systematicity, favoring more objective measures of language proficiency (for a review, see Brown, Plonsky, & Teimouri, 2018). Regardless of the theoretical and methodological issues underlying the use of course grades or language tests as measures of language proficiency, the results of this meta-analysis show a similar negative relationship between language anxiety and both language achievement measures. In short, as far as the negative effects of anxiety on students' L2 learning are concerned, both course grades and language tests offer ecologically valid and practical approaches for exploring such a relationship.

The results of our study did not reveal any noteworthy differences between the anxiety-achievement connection in SL and FL contexts, suggesting each context has its own unique, inherent challenges that cause L2 learners to feel anxious. The detrimental

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

effects of anxiety on students' language achievement were found to be strongest in lower-level educational settings. More specifically, elementary students seemed to suffer more from the negative effects of anxiety. It seems that younger learners are less competent in dealing with anxiety and the environmental factors that generate these negative emotions. It may be the case that, when students grow older, they become more competent in taking control of their own learning by using a range of metacognitive and affective strategies to manage their anxiety. Moreover, the language learning experiences of students may also be beneficial in managing their negative emotions in class. Students at higher educational levels clearly have had more language experiences when compared to elementary students. As a result, they are more familiar with the negative scenarios that may occur during the process of learning the target language and are thus more successful in coping with them.

However, we see an increase in students' anxiety from junior high school students to senior high school students to college students. Overall, the ups and downs of anxiety can be attributed to two confounding variables: the age of students and the inherent features of the context. In other words, as students grow older and gain more L2 experiences, they may feel less anxious. The educational context, by contrast, may create new experiences and obligations on the students that may cause more anxiety. The effect sizes were smallest in language institutes, which we might attribute to the heterogeneity of students studying the target language in such a context.

One interesting finding of this study is related to the variation in observed effects across target languages. The results suggest that the negative relationship between anxiety and achievement is weaker among learners of English. Such a finding might be due to some inherent characteristics underlying the target languages under investigation and how the L2 learners perceive and evaluate them (e.g., English vs. Chinese). Moreover, the different features underlying the context of the target language as well as cultural differences of students in proneness to anxiety might generate different emotional reactions on the part of students. Because English is by far the most researched target language in SLA, and because most of the methodological advances as well as pedagogical implications in applied linguistics have resulted from a focus on English as the target language, perhaps English has benefited more than other languages in designing more comforting, informative, and beneficial learning contexts. Furthermore, familiarity with the English language because of its international status and its widespread presence in people's daily lives globally (e.g., Internet) might also play a facilitative role in reducing anxiety of English language learners compared to learners of other languages.

Finally, the results show that anxiety more strongly affects L2 learners' listening performance than their speaking performance. This is probably due to the time pressure involved in listening. Because L2 learners need to decode an interlocutor's intended message instantly within a context, this time pressure may increase their level of anxiety and, consequently, drain their attentional resources (see Li, 2016). In addition, while L2 learners may freely modify their output using the various (non)linguistic devices at their disposal to successfully communicate their intended message, they have limited resources to decode an interlocutor's speech when miscommunication occurs. For instance, they may not be able to interrupt a conversation repeatedly to request a clarification or a repetition. However, anxiety had the lowest negative effects during reading the target language. This is somehow expected given the nature of reading. Unlike listening, students are often

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

allowed to read an L2 text at their own pace, reread parts of the text that they do not fully understand, and use other resources to aid in their comprehension.

Trait anxiety also showed a smaller effect size compared to situation-specific anxieties. This finding lends support to the use of language-specific measures to assess detrimental effects of anxiety on students' language achievement. In sum, the findings support the evolution of anxiety research from a traitlike variable to a more situation-specific construct targeting different language-skills.

LIMITATIONS AND DIRECTION FOR FURTHER RESEARCH

The present study has provided robust findings on a number of questions related to anxiety in L2 research and its relationship to L2 achievement. Nonetheless, there is limited evidence concerning the results for a number of moderators, such as for study-abroad contexts, elementary and junior high school educational levels, and anxiety types (e.g., trait and state anxiety, listening and reading anxiety). Sequencing of languages (L2– L5) (see Dewaele, 2010) might also be a significant factor influencing the strength of the relationship between anxiety and language achievement as well as the time lag involved. The results of the present study are also constrained by the exclusion of reports published in languages other than English. Despite these limitations, the results of this meta-analysis can serve to inform future empirical efforts in the domain of L2 anxiety.

Of critical significance is the need to improve reporting of statistical and psychometric features of language achievement measures. As the results revealed, few studies have carefully addressed the characteristics of the instruments used to measure language achievement. While previous research has attributed inconsistent findings in L2 anxiety research to the use of general anxiety measures (e.g., Horwitz, 2010, 2017; MacIntyre, 2017; Scovel, 1978), it is likely that inappropriate or suboptimal measures of L2 achievement may also have played a role in this regard. We also recommend that future studies use the psychometric features of their measures to interpret their results and improve the validity of their instruments. Researchers can and should also apply psychometric corrections to arrive at unattenuated estimates of their relationships of interest (see Plonsky & Derrick, 2016; Teimouri, 2017).

In addition to improving the quality of instruments and the data they yield, a combination of different language achievement measures (i.e., self-assessed, course grades, and language tests) in a study is also desirable in that it will provide a more comprehensive picture of the direction and magnitude of the relationships in question. Including both subjective (e.g., self-assessed) and objective measures of language achievement will also provide valuable sources of information. At the same time, the use of GPAs as a measure of L2 achievement, though easy to collect, is not recommended. GPAs represent students' overall academic achievement rather than their L2 achievement and therefore provide low construct validity in this domain.

There are also a number of substantive gaps in the literature. For example, future studies are needed in study abroad contexts, with elementary and junior high school learners, in language institutes, and with learners of languages other than English (see discussion in sampling practices in Plonsky, 2017). In terms of skill-specific anxiety, the vast majority of the available evidence concerns L2 speaking; much less is known about the relationship between anxiety and all other L2 skills.

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

In addition to improving instrumentation and expanding the substantive reach of L2 anxiety research, future studies might also benefit from a number of adjustments in the realm of data analysis. The statistic of choice in this domain is the correlation. Correlations are useful in that they model the relationship between two measured variables in a very straightforward fashion (Norouzian & Plonsky, 2018). However, correlation coefficients can obscure nonlinear relationships. It could be the case that the relationship between anxiety and achievements varies as a function of different levels of anxiety or achievement. For these reasons, we strongly encourage researchers in this area to examine and present their data visually by means of a scatter plot. We also suggest examining and plotting residuals based on the correlations of interest. If a nonlinear relationship is found, researchers might consider alternatives to correlation such as quantile regression. We would further encourage researchers to consider statistical models that go beyond bivariate relations. Given the multivariate nature of the associations between learner internal mechanisms and L2 achievement, it is statistically and conceptually appropriate to examine multiple relationships simultaneously (Brown, 2015; Plonsky & Oswald, 2017). Multiple regression and structural equation modeling, though not without their own analytic and interpretive challenges (see Plonsky & Ghanbar, in press), provide two approaches for doing so that might be fruitfully applied in this domain.

As noted, a situation-specific definition and measurement of anxiety in SLA was indeed a turning point in this research domain and led to much more consistent findings regarding the negative effects of anxiety. From a theoretical stance, however, language anxiety still remains an elusive concept in SLA (see Gkonou et al., 2017, Goetze, 2018; Şimşek & Dörnyei, 2017). We need theoretical frameworks wherein cognitive, social, affective, and behavioral components of language anxiety and their subtle interrelationships can be examined and explained in detail. Furthermore, examining the links between anxiety and other negative emotions, such as fear, shame, embarrassment, and sadness—which are usually swept under the cover of language anxiety—can elucidate the ambiguous nature of language anxiety (Teimouri, 2018).

CONCLUSION

This meta-analysis was conducted to examine the relationship between language anxiety and L2 achievement. Overall, the findings indicate that anxiety has a moderate, negative association with achievement. As concluded by MacIntyre (2017) in his review of L2 anxiety research, we can put to bed, once and for all, whether anxiety is best conceptualized as a negative predictor of language achievement. Moreover, the negative effects of anxiety vary concerning the context, educational levels of students, and the target language. At the same time, we hope that the present study will motivate and provide direction to future efforts in this dynamic domain of L2 research, such as novel theoretical perspectives and robust measurement tools.

SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit https://doi.org/10.1017/S0272263118000311

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

REFERENCES

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*, 155–168.

Brown, A., Plonsky, L., & Teimouri, Y. (2018). The use of course grades as metrics in L2 research: A systematic review. *Foreign Language Annals*, *51*, 763–778.

Brown, J. D. (2015). Why bother learning advanced quantitative methods in L2 research? In L. Plonsky (Ed.), *Advancing quantitative methods in L2 research* (pp. 9–20). New York, NY: Routledge.

Cattell, R. B., & Scheier, I. H. (1961). *The meaning and measurement of neuroticism and anxiety*. New York: Ronald Press.

Chastain, K. (1975). Affective and ability factors in second language learning. *Language Learning*, *25*, 153–161.

Chen, G., Gully, S. M., Whiteman, J. A., & Kilcullen, R. N. (2000). Examination of relationships among trait-like individual differences, state-like individual differences, and learning performance. *Journal of Applied Psychology*, *85*, 835–847.

Cheng, Y. S. (2004). A measure of second language writing anxiety: Scale development and preliminary validation. *Journal of Second Language Writing*, *13*, 313–335.

Cheng, Y. S. (2017). Development and preliminary validation of four brief measures of L2 language-skill-specific anxiety. *System*, *68*, 15–25.

Cheng, Y. S., Horwitz, E. K., & Schallert, D. L. (1999). Language anxiety: Differentiating writing and speaking components. *Language Learning*, *49*, 417–446.

Cooper, H., Hedges, L. V., & Valentine, J. C. (Eds.). (2009). *The handbook of research synthesis and meta-analysis*. New York: Russell Sage Foundation.

Derrick, D. J. (2016). Instrument reporting practices in second language research. *TESOL Quarterly*, *50*, 132–153.

Dewaele, J. (2010). *Emotions in multiple languages*. London: Palgrave Macmillan.

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.

Elkhafaifi, H. (2005). Listening comprehension and anxiety in the Arabic language classroom. *Modern Language Journal*, *89*, 206–220.

Gardner, R. (1985). *Social psychology and second language learning: The role of attitudes and motivation*. London, UK: Edward Arnold.

Gardner, R. C., & MacIntyre, P. D. (1993). A student's contributions to second-language learning. Part II: Affective variables. *Language Teaching*, *26*, 1–11.

Gardner, R. C., Smythe, P. C., Clément, R., & Gliksman, L. (1976). Social and psychological factors in second language acquisition. *Canadian Modern Language Review*, *32*, 198–213.

Gkonou, C., Daubney, M., & Dewaele, J. M. (2017). *New insights into language anxiety: Theory, research and educational implications*. Bristol, UK: Multilingual Matters.

Goetze, J. (2018). *Linking cognition and emotion: An appraisal study of foreign language teacher anxiety* (Unpublished doctoral dissertation). Washington, DC: Georgetown University.

Hashimoto, Y. (2002). Motivation and willingness to communicate as predictors of reported L2 use: The Japanese ESL context. *University of Hawai'i Second Language Studies Paper*, *20*, 29–69.

Horwitz, E. K. (2001). Language anxiety and achievement. *Annual Review of Applied Linguistics*, *21*, 112–126.

Horwitz, E. K. (2010). Foreign and second language anxiety. *Language Teaching*, *43*, 154–167.

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. 31–50). Bristol, UK: Multilingual Matters.

Horwitz, E. K., Horwitz, M. B., & Cope, J. (1986). Foreign language classroom anxiety. *Modern Language Journal*, *70*, 125–132.

Khajavy, G. H., MacIntyre, P., & Barabadi, E. (2018). Role of the emotions and classroom environment in willingness to communicate: Applying doubly latent multilevel analysis in second language acquisition research. *Studies in Second Language Acquisition*, *40*, 605–624.

Kleinmann, H. (1977). Avoidance behaviour in adult second language acquisition. *Language Learning*, *27*, 93–107.

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

Konstantopoulos, S., & Hedges, L. V. (2004). Meta-analysis. In D. Kaplan (Ed.), *The SAGE handbook of quantitative methodology for the social sciences* (pp. 281–297). New York: Sage.

Larson-Hall, J., & Plonsky, L. (2015). Reporting and interpreting quantitative research findings: What gets reported and recommendations for the field. *Language Learning*, *65*, 127–159.

Li, S. (2016). The construct validity of language aptitude: A meta-analysis. *Studies in Second Language Acquisition*, *38*, 801–842.

Linck, J. A., Osthus, P., Koeth, J. T., & Bunting, M. F. (2014). Working memory and second language comprehension and production: A meta-analysis. *Psychonomic Bulletin & Review*, *21*, 861–883.

Lipsey, M. W., & Wilson, D. B. (2001). *Practical meta-analysis*. Thousand Oaks, CA: Sage.

Ma, X. (1999). A meta-analysis of the relationship between anxiety toward mathematics and achievement in mathematics. *Journal for Research in Mathematics Education*, *30*, 520–540.

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). Boston, MA: McGraw-Hill.

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*. Bristol, UK: Multilingual Matters.

MacIntyre, P. D., & Charos, C. (1996). Personality, attitudes, and affect as predictors of second language communication. *Journal of Language and Social Psychology*, *15*, 3–26.

MacIntyre, P. D., & Gardner, R. C. (1991). Methods and results in the study of anxiety and language learning: A review of the literature. *Language Learning*, *41*, 85–117.

MacIntyre, P. D., & Gardner, R. C. (1994a). The subtle effects of language anxiety on cognitive processing in the second language. *Language Learning*, *44*, 283–305.

MacIntyre, P. D., & Gardner, R. C. (1994b). The effects of induced anxiety on three stages of cognitive processing in computerized vocabulary learning. *Studies in Second Language Acquisition*, *16*, 1–17.

MacIntyre, P. D., Noels, K. A., & Clément, R. (1997). Biases in self-ratings of second language proficiency: The role of language anxiety. *Language Learning*, *47*, 265–287.

Marsden, E., Mackey, A., & Plonsky, L. (Eds.). (2015). *Advancing methodology and practice: The IRIS repository of instruments for research into second languages*. New York: Routledge.

Marsden, E., Mackey, A., and Plonsky, L. (2016). The IRIS Repository: Advancing research practice and methodology. In E. Marsden, A. Mackey, & L. Plonsky (Eds.), *Advancing methodology and practice: The IRIS Repository of instruments for research into second languages* (pp. 1–21). New York: Routledge.

Masgoret, A. M., & Gardner, R. C. (2003). Attitudes, motivation, and second language learning: A meta-analysis of studies conducted by Gardner and associates. *Language Learning*, *53*, 123–163.

Norouzian, R., & Plonsky, L. (2018). Correlation and simple linear regression in applied linguistics. In A. Phakiti, P. De Costa, L. Plonsky, & S. Starfield (Eds.), *The Palgrave handbook of applied linguisitics research methodology* (pp. 395-421). London: Palgrave Macmillan.

Norris, J. M., & Ortega, L. (2006). The value and practice of research synthesis for language learning and teaching. In J. M. Norris & L. Ortega (Eds.), *Synthesizing research on language learning and teaching* (pp. 3–50). Philadelphia, PA: John Benjamins.

Oswald, F. L., & Plonsky, L. (2010). Meta-analysis in second language research: Choices and challenges. *Annual Review of Applied Linguistics*, *30*, 85–110.

Oxford, R. L. (1999). Anxiety and the language learner: New insights. In J. Arnold (Ed.), *Affect in language learning* (pp. 68–86). New York: Cambridge University Press.

Papi, M. (2010). The L2 motivational self system, L2 anxiety, and motivated behavior: A structural equation modeling approach. *System*, *38*, 467–479.

Papi, M., & Teimouri, Y. (2014). Language learner motivational types: A cluster analysis study. *Language Learning*, *64*, 493–525.

Paquot, M., & Plonsky, L. (2017). Quantitative research methods and study quality in learner corpus research. *International Journal of Learner Corpus Research*, *3*, 61–94.

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

Plonsky, L. (2013). Study quality in SLA: An assessment of designs, analyses, and reporting practices in quantitative L2 research. *Studies in Second Language Acquisition*, *35*, 655–687.

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

Plonsky, L. (2017). Quantitative research methods in instructed SLA. In S. Loewen & M. Sato (Eds.), *The Routledge handbook of instructed second language acquisition* (pp. 505–521). New York: Routledge.
Plonsky, L., & Brown, D. (2015). Domain definition and search techniques in meta-analyses of L2 research (Or why 18 meta-analyses of feedback have different results). *Second Language Research*, *31*, 267–278.
Plonsky, L., & Derrick, D. J. (2016). A meta-analysis of reliability coefficients in second language research. *The Modern Language Journal*, *100*, 538–553.
Plonsky, L, & Ghanbar, H. (2018). Multiple regression in L2 research: A methodological synthesis and guide to interpreting R^2 values. *Modern Language Journal*, *102*, 713–731.
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). West Sussex, UK: Blackwell.
Plonsky, L., & Oswald, F. L. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning*, *64*, 878–912.
Plonsky, L., & Oswald, F. L. (2015). Meta-analyzing second language research. In E. Marsden, A. Mackey, & L. Plonsky (Eds.), *Advancing quantitative methods in second language research* (pp. 106–128). New York: Routledge.
Plonsky, L., & Oswald, F. L. (2017). Multiple regression as a flexible alternative to ANOVA in L2 research. *Studies in Second Language Acquisition*, *39*, 579–592.
Rosenthal, R., Cooper, H., & Hedges, L. V. (1994). Parametric measures of effect size. In H. Cooper & V. Hedges (Eds.), *The handbook of research synthesis* (pp. 231–244). New York: Russell Sage Foundation.
Saito, Y., & Samimy, K. K. (1996). Foreign language anxiety and language performance: A study of learner anxiety in beginning, intermediate, and advanced-level college students of Japanese. *Foreign Language Annals*, *29*, 239–249.
Saito, Y., Horwitz, E. K., & Garza, T. J. (1999). Foreign language reading anxiety. *Modern Language Journal*, *83*, 202–218.
Seipp, B. (1991). Anxiety and academic performance: A meta-analysis of findings. *Anxiety Research*, *4*, 27–41.
Scovel, T. (1978). The effect of affect on foreign language learning: A review of the anxiety research. *Language Learning*, *28*, 129–142.
Şimşek, E., & Dörnyei, Z. (2017). Anxiety and L2 self-images: The 'anxious self'. In C. Gkonou, M. Daubney, & J.-M. Dewaele (Eds.), *New insights into language anxiety: Theory, research and educational implications* (pp. 51–69). Bristol, UK: Multilingual Matters.
Sparks, R. L., & Patton, J. (2013). Relationship of L1 skills and L2 aptitude to L2 anxiety on the Foreign Language Classroom Anxiety Scale. *Language Learning*, *63*, 870–895.
Spielberger, C. D., Gorsuch, R. L., & Lushene, R. E. (1970). *Manual for the state-trait anxiety inventory*. Palo Alto, CA: Consulting Psychologists Press.
Teimouri, Y. (2017). L2 selves, emotions, and motivated behaviors. *Studies in Second Language Acquisition*, *39*, 681–709.
Teimouri, Y. (2018). Differential roles of shame and guilt in L2 learning: How bad is bad? *The Modern Language Journal*, *102*, 632–652.
Young, D. J. (1986). The relationship between anxiety and foreign language oral proficiency ratings. *Foreign Language Annals*, *19*, 439–445.

APPENDIX A

			95% CI		
K	Study name	*r*	Lower	Upper	Weight %
1	Alsowat, 2016	–0.28	–0.37	–0.18	1.1

(continued on following page)

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

(continued)

K	Study name	*r*	95% CI Lower	95% CI Upper	Weight %
2	Awan et al., 2010	–0.27	–0.42	–0.12	1.02
3	Bashosh et al., 2013*	0.81	0.70	0.88	0.86
4	Coulombe, 2001 (S1)	–0.21	–0.34	–0.07	1.047
5	Coulombe, 2001 (S2)	–0.44	–0.63	–0.21	0.86
6	GhorbanDordinejad & Nasab, 2013	–0.36	–0.46	–0.24	1.07
7	Göçer, 2014	–0.06	–0.17	0.05	1.09
8	Kelsen & Liang, 2012*	0.21	–0.03	0.42	0.9
9	Knell & Chi, 2012	–0.57	–0.66	–0.46	1.04
10	Aida, 1994	–0.38	–0.54	–0.19	0.95
11	Anyadubalu, 2010	–0.28	–0.38	–0.18	1.09
12	Bailey et al., 2000	0.05	–0.09	0.19	1.06
13	Bernaus & Gardner, 2008	–0.24	–0.31	–0.17	1.13
14	Chakrabarti & Sengupta, 2012	–0.36	–0.50	–0.21	1.02
15	Chen & Lin, 2009*	–0.76	–0.83	–0.67	0.99
16	Cheng et al., 1999	–0.33	–0.41	–0.24	1.11
17	Chen & Chang, 2004	–0.50	–0.61	–0.37	1.02
18	Sarwat, 2012	–0.29	–0.43	–0.14	1.02
19	Zaman & Semul, 2013	–0.53	–0.65	–0.39	0.99
20	Chiang, 2010	–0.22	–0.34	–0.08	1.1
21	Djigunovi & Legac, 2008	–0.43	–0.63	–0.18	0.84
22	Dogan & Tuncer, 2016	–0.23	–0.30	–0.16	1.13
23	Elkhafaifi, 2005	–0.61	–0.69	–0.52	1.1
24	Ghorbandordinejad & Ahmadabad, 2016	–0.47	–0.55	–0.39	1.11
25	Woodrow, 2006	–0.24	–0.34	–0.12	1.08
26	Al Asmari, 2013	–0.42	–0.53	–0.29	1.05
27	Abu-Rabia, 2004	–0.48	–0.65	–0.27	0.88
28	Bekleyen, 2009	–0.53	–0.68	–0.33	0.89
29	Cakici, 2016	–0.26	–0.36	–0.15	1.09
30	Lien, 2016	–0.25	–0.34	–0.15	1.1
31	Liu, 2013	–0.55	–0.65	–0.42	1.01
32	Djigunovic, 2004	0.18	–0.15	0.47	0.76
33	Ganschow & Sparks, 1996	–0.52	–0.62	–0.39	1.02
34	Hewitt & Stephenson, 2012	–0.49	–0.70	–0.20	0.75
35	Jafarigohar & Behrooznia, 2012	–0.21	–0.38	–0.02	0.98
36	Jee, 2014*	0.59	–0.05	0.89	0.36
37	Jee, 2016	–0.45	–0.63	–0.21	0.86
38	Liu & Ni, 2015	–0.12	–0.18	–0.07	1.14
39	Machida, 2001	–0.39	–0.54	–0.21	0.97
40	MacIntyre & Gardner, 1994	–0.53	–0.66	–0.37	0.96
41	Marcos-Llinás & Garau, 2009*	0.18	0.00	0.34	1.01
42	McCroskey et al., 1985 (S1)	–0.36	–0.42	–0.29	1.13
43	McCroskey et al., 1985 (S2)	–0.36	–0.45	–0.26	1.09
44	Olivares-Cuhat, 2010	–0.65	–0.80	–0.41	0.75
45	Rastegar & Karami, 2015	–0.57	–0.71	–0.39	0.9
46	Salehi & Marefat, 2014	–0.22	–0.35	–0.08	1.05
47	Shariati & Bordbar, 2009	–0.17	–0.38	0.07	0.9

(continued on following page)

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

(continued)

			95% CI		
K	Study name	*r*	Lower	Upper	Weight %
48	Sparks et al., 2009	–0.56	–0.72	–0.34	0.83
49	Tanielian, 2014	–0.16	–0.25	–0.07	1.11
50	Valizadeh & Alavinia, 2013	–0.59	–0.68	–0.47	1.03
51	Wu, 2011	–0.16	–0.35	0.06	0.94
52	Yana, 2012	–0.32	–0.55	–0.04	0.81
53	Zhao et al., 2013*	0.45	0.29	0.59	0.98
54	Young, 1986	–0.24	–0.47	0.02	0.86
55	Atasheneh & Izadi, 2012*	0.47	0.24	0.65	0.86
56	Azher et al., 2010	–0.27	–0.42	–0.12	1.02
57	Chen & Chang, 2009	–0.26	–0.45	–0.05	0.94
58	Dalkilic, 2001	–0.59	–0.70	–0.45	0.98
59	Gardner & MacIntyre, 1993	–0.34	–0.51	–0.14	0.94
60	Horwitz, 1986 (S1)	–0.49	–0.71	–0.17	0.71
61	Horwitz, 1986 (S2)	–0.54	–0.75	–0.22	0.71
62	Liu, 2016	–0.15	–0.21	–0.09	1.14
63	Kao & Craigie, 2010*	–0.79	–0.85	–0.70	0.96
64	Lu & Liu, 2011	–0.32	–0.37	–0.26	1.14
65	Phillips, 1992	–0.40	–0.63	–0.11	0.78
66	Gardner et al., 1997	–0.51	–0.65	–0.35	0.96
67	MacIntyre & Gardner, 1989	–0.43	–0.57	–0.25	0.97
68	MacIntyre et al., 1997	–0.55	–0.75	–0.27	0.73
69	Park & Lee, 2005	–0.31	–0.46	–0.15	1.01
70	Saito & Samimy, 1996 (S1)	–0.24	–0.39	–0.07	1.01
71	Saito & Samimy, 1996 (S2)	–0.39	–0.57	–0.19	0.92
72	Saito & Samimy, 1996 (S3)	–0.52	–0.71	–0.26	0.78
73	Sanchez-Herrero, A., & Sanchez, M. P., 1992	–0.49	–0.59	–0.38	1.06
74	Szyszka, 2011	–0.54	–0.72	–0.30	0.8
75	Wang & Li, 2011	–0.35	–0.52	–0.15	0.95
76	Alkhateeb, 2014	–0.35	–0.50	–0.18	0.99
77	Atef-Vahid & Kashani, 2011	–0.64	–0.80	–0.39	0.73
78	Budin, 2014*	0.36	0.23	0.48	1.05
79	Fadillah, 2010	–0.12	–0.28	0.04	1.02
80	Liu, 2007	–0.17	–0.25	–0.09	1.12
81	Rodriguez, 1995	–0.57	–0.70	–0.41	0.94
82	Fernandez, 2011	–0.49	–0.67	–0.24	0.83
83	Tianjian, 2010	–0.32	–0.43	–0.20	1.07
84	Tittle, 1997 (S1)	–0.22	–0.55	0.17	0.66
85	Tittle, 1997 (S2)	–0.24	–0.52	0.08	0.76
86	Tittle, 1997 (S3)	–0.42	–0.72	0.02	0.57
87	Xiu Yan & Wang, 2012	–0.32	–0.55	–0.04	0.81
88	Bosmans & Hurd, 2016	0.27	–0.16	0.61	0.6
89	Chan & Wu, 2004	–0.28	–0.35	–0.20	1.12
90	Anderson, 2012	–0.40	–0.64	–0.09	0.74
91	Price, 1988	–0.25	–0.42	–0.06	0.97
92	Tallon, 2006	–0.61	–0.69	–0.52	1.05
93	Teemant, 1997	–0.52	–0.72	–0.22	0.73

(continued on following page)

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press

(continued)

K	Study name	*r*	95% CI Lower	95% CI Upper	Weight %
94	Wang, 2011	–0.27	–0.36	–0.17	1.1
95	Yang, 1993	–0.33	–0.59	–0.01	0.75
96	Zhao, 2009	–0.65	–0.74	–0.54	0.99
97	Argaman & Abu-Rabia, 2002	–0.26	–0.47	–.02	0.89
98	Arnaiz & Guillen, 2012	–0.30	–0.42	–0.18	1.06
99	Chanprasert & Wichadee, 2015	–0.33	–0.48	–0.15	0.99
100	Dewaele & Ip, 2013	–0.68	–0.79	–.53	0.9
101	Kamarulzaman et al., 2013	–0.29	–0.45	–.12	0.99
102	Mills et al., 2006	–0.27	–0.45	–.07	0.95
103	Mills et al., 2007	–0.16	–0.27	–.05	1.09
104	Razak et al., 2017	–0.25	–0.26	.06	1.03
105	Shaq et al., 2013	–0.47	–0.53	–0.40	1.12

Note: The starred references were excluded because they were found as outliers (standard residuals > 2.5.

https://doi.org/10.1017/S0272263118000311 Published online by Cambridge University Press