Journal of Second Language Writing 53 (2021) 100817

Contents lists available at ScienceDirect

Journal of Second Language Writing

journal homepage: www.elsevier.com/locate/jslw

Relationship between second language English writing self-efficacy and achievement: A meta-regression analysis

Ting Sun [a], Chuang Wang [b,*], Richard G. Lambert [a], Lan Liu [c]

[a] Cato College of Education, University of North Carolina at Charlotte, United States
[b] Faculty of Education, University of Macau, Macao
[c] Faculty of Hospitality and Tourism Management, Macau University of Science and Technology, Macao

ARTICLE INFO

Keywords:
Meta-analysis
L2 English writing
Self-efficacy
Effect size
HLM

ABSTRACT

The study aimed (a) to estimate the overall average effect size of the relationship between writing self-efficacy and writing achievement for first language (L1) and second language (L2) writers in English; and (b) to examine how writing in English as a L1/L2 moderates the relationships based on a meta-analysis of published journal articles and dissertations/theses. Data included 565 effect sizes from 76 studies through a rigorous process of literature searches, screening, and data coding. A two-level meta-regression model was constructed to estimate the average effect size and to examine the moderating effects of the covariates. Results revealed a medium effect size ($r = .29$) with both L1 and L2 writers, which indicated approximately 9% of the variability in English writing achievement was associated with variability in students' self-efficacy. Furthermore, writing in English as a L1/L2 was found to moderate the relationship between writing self-efficacy and writing achievement, with the effect size estimated with L2 learners ($r = .441$) being statistically significantly larger than that yielded with L1 learners ($r = .233$), after controlling for the covariates of sample size, gender, grade, statistical procedures, and publication type. Results also revealed that statistical procedure moderated effect size estimates. This meta-analysis has practical implications for heightening the awareness of English teachers in developing L2 students' self-efficacy in the writing classroom setting. It also cautions researchers against the misinterpretation and misuse of effect sizes calculated by different statistical procedures.

1. Introduction

English writing is a critical and versatile skill (Graham, 2006), which plays a pivotal role in academic success in nearly all countries (Asmari, 2013) and serves as a threshold benchmark for college admission, job application, and career promotion (National Commission on Writing, 2004). In countries where English is spoken as a second language (L2), English writing is also essential since it is an index of language learners' overall linguistic proficiency (Archibald, 2016). English writing does not only provide professional opportunities for individuals, but also is a skill requires in business, politics, and education in the globalized world (National Commission on Writing, 2004).

Writing is a thought-demanding and challenging undertaking (Anastasiou & Michail, 2013). According to the National Center for Education Statistics (2012), 20 % of eighth graders and 21 % of twelfth graders in the United States were below the basic level in

* Corresponding author at: Faculty of Education, University of Macau, Avenida da Universidade, Taipa, Macao.
E-mail addresses: tsun4@uncc.edu (T. Sun), wangc@um.edu.mo (C. Wang), rglamber@uncc.edu (R.G. Lambert), 811518650@qq.com (L. Liu).

https://doi.org/10.1016/j.jslw.2021.100817
Received 5 June 2020; Received in revised form 2 March 2021; Accepted 3 March 2021
Available online 7 May 2021

English writing and only 3% of students at both grades performed at the advanced level. Some researchers posit that this phenomenon is caused by the specificity of writing that requires not only linguistic capability such as lexical, syntactic, and discoursal competence, but also entails synthetic and analytical abilities, ideation, and logical reasoning on the part of writers (Anastasiou & Michail, 2013; Bruning et al., 2013). Additionally, some researchers attribute students' low writing achievement to teaching approaches and instructional methods, which focus primarily on cognitive knowledge.

However, writing requires metacognitive and motivational knowledge in addition to cognitive knowledge (Wischgoll, 2016). It involves other factors such as social and physical environment, motivation, working memory, long-term memory, and cognitive process (Hayes, 2000). It is a process with great intricacy and complexity that necessities both "low-level skills such as spelling, capitalization, punctuation and other conventions" (Anastasiou & Michail, 2013, p. 53) and high-level skills, to name a few, self-regulation and self-efficacy (Bruning et al., 2013). Previous research provides evidence that self-efficacy is a significant predictor of language proficiency in general (Huang et al., 2015) and writing achievement in particular (Bruning et al., 2013). Student writers with higher levels of self-efficacy are highly motivated (Zhang & Guo, 2012), display greater perseverance (Prat-Sala & Redford, 2012), expend more effort, have lower anxiety (Pajares et al., 2000; Woodrow, 2011), and therefore achieve better writing outcomes. Self-efficacy was found to have a recursive relation with academic success (Robinson et al., 2020): Writing self-efficacy promotes writing achievement; the resultant higher achievement, in turn, nurtures self-efficacy development in writing.

Although a large body of research has found significant correlations between writing self-efficacy and students' writing achievement with both English as a first language (L1) writers and L2 writers (e.g., Bruning et al., 2013; Pajares, 2003; Prat-Sala & Redford, 2012; White & Bruning, 2005; Zhang & Guo, 2012), the magnitude of the relationship is not consistent across studies. Therefore, a meta-analysis is warranted to provide an estimation of the effect size of the relationship between self-efficacy and writing achievement in L1 and L2 English writing. Writers encounter unique challenges in L2 writing (Galbraith & Vedder, 2019). Thus, a meta-analysis of the extant literature on writing self-efficacy and writing achievement in L2 is particularly important. Such a synthesis of current literature provides evidence on the effectiveness of self-efficacy interventions for enhancing writing outcomes for L2 writers. It also provides practitioners, educators, and administrators with evidence-based guidance for teaching or curriculum and program design in L2 English writing. The intents of our inquiry are (a) to estimate the overall average effect size of the relationship between English writing self-efficacy and writing achievement for L1 and L2 writers in English, and; (b) to examine the extent to which writing in English as an L1 or L2 moderates the relationship.

2. Literature review

2.1. Social cognitive theory and writing self-efficacy

Self-efficacy, defined as "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances" (Bandura, 1986, p. 391), is believed to play a facilitative role in human functioning (Bandura, 1986). According to social cognitive theory, the beliefs people have predict their behaviors such as persistence and engagement, and influence the consequences of those behaviors (Bandura, 1997). In the context of both L1 and L2 writing in English, efficacious students tend to take on challenging writing tasks and put in more effort (Usher & Pajares, 2008; Zhang & Guo, 2012). Therefore, variations in writing achievement are not only attributive to students' writing competence but also to the beliefs they have in their writing capabilities.

Self-efficacy should be differentiated from other intertwined but different concepts such as self-concept, self-esteem, and self-confidence. Self-concept was defined as "…the totality of the individual's thoughts and feelings having reference to himself as an object" (Rosenberg, 1979, p.7). Self-concept is a collective self-perception and a multidimensional construct consisting of self-efficacy, self-esteem, self-confidence, and other related constructs (Schunk, 1991). Self-esteem refers to "judgments of self-worth" (Bandura, 1997, p. 11) rather than judgments of one's capabilities. Self-confidence is a belief in one's capability in creating positive results, while self-efficacy refers to the belief of how well one can accomplish a task with the judgment of his/her own competence and the demand of the task: It is, therefore, domain- and task-specific (Bandura, 1986). Therefore, self-efficacy in writing is different from self-efficacy in other domains of language learning, and self-efficacy in one writing task may be different from that in another.

2.2. The relationship between self-efficacy and English writing achievement for L1 and L2 writers

Writing self-efficacy was defined as "students' judgments of their confidence that they possessed the various composition, grammar, usage, and mechanical skills appropriate to their academic level" (Pajares & Valiante, 2001, p.369). Voluminous research has been carried out to examine the relationship between self-efficacy and writing achievement in various languages. Statistically positive correlations between self-efficacy and writing achievement were evidenced in the languages of Chinese (Chan & Lam, 2008), German (Brunstein & Glaser, 2011), Greek (Anastasiou & Michail, 2013), Portuguese (Limpo & Alves, 2013), and Spanish (Villalón et al., 2015) for both L1 and L2 writers.

In the language of English, a large body of research was conducted on the relationship between writing self-efficacy and writing achievement for L1 writers (Pajares, Johnson, et al., 2007; Prat-Sala & Redford, 2012; Sanders-Reio et al., 2014; Wright et al., 2019). Working with 145 L1 students in English in a UK university, Prat-Sala and Redford (2012) found positive relationships between writing self-efficacy and writing scores for both freshmen and sophomores. This finding was replicated in the K-12 context. For example, with 1258 L1 writers from an elementary school, a middle school, and a high school in the United States, Pajares (2007) contended that two dimensions of writing self-efficacy (i.e., self-efficacy for basic skills and self-efficacy for composition skills) were statistically

significantly correlated with students' writing performance (r ranging from .25 to .34).

Research also examined the relationship between writing self-efficacy and writing achievement in L2 English writing (e.g., Chen & Zhang, 2019; Sun & Wang, 2020; Teng et al., 2018; Woodrow, 2011; Zabihi, 2018; Zhang & Guo, 2012). Writing self-efficacy was found to have correlational and predictive effects on writing outcomes of students in China (Chen & Zhang, 2019; Teng et al., 2018; Woodrow, 2011; Zhang & Guo, 2012), Korea (Chae, 2013), Iran (Sarkhoush, 2013; Zabihi, 2018), Ethiopia (Amogne, 2008), and Cambodia (Chea, 2012). Teng et al. (2018) found significant correlations between three dimensions of writing self-efficacy (i.e., linguistic self-efficacy, self-regulatory efficacy, and performance self-efficacy) and the writing performance of undergraduate students in China. Students with higher writing self-efficacy tended to outperformance students with lower self-efficacy. Writing self-efficacy also predicted three measures of second language writing performance (i.e., complexity, accuracy, and fluency) among 232 upper-intermediate college students in Iran (Zabih, 2018). Writing self-efficacy not only predicted writing achievement but also mediated the relationship between writing performance and other constructs such as motivation (Zhang & Guo, 2012) and writing anxiety (Woodrow, 2011).

Although meta-analyses have been applied to other factors which influence writing performance, such as instructional practices (e.g., Graham et al., 2012), computers (e.g., Goldberg et al., 2003), formative assessment (e.g., Graham et al., 2015), and written corrective feedback (e.g., Kang & Han, 2015), no meta-analysis has been conducted to examine the factor of self-efficacy in English writing. It is warranted to estimate the effect size of the relationship between writing self-efficacy and English writing achievement in L1 and L2.

In addition to the estimation of the overall average effect size, examining factors that explain the variance of true effect sizes across studies (i.e., moderator analysis) is also an important component of meta-analytic research (Chambers, 2004). Therefore, we proposed potential moderators based on empirical research and theoretical rationale in the next section.

2.3. Moderating effect of L1/L2 writing on the effect size

Although the majority of studies noted a significant relationship between writing self-efficacy and achievement in both L1 and L2 English writing, the directionality and magnitude of their relationships varied greatly. For example, a small effect size between self-efficacy and writing performance ($r = .18$) was noted with L1 students in the United States (Wilson & Trainin, 2007) as compared to a large effect size ($r = .51$) reported with L2 learners in Iran (Sarkhoush, 2013). McGovern (2004) even found a negative correlation between writing self-efficacy and writing outcomes with L1 writers. Given that L1 writers are distinct from L2 writers in terms of English proficiency, writing processes, and demands of cognitive, affective, and behavioral knowledge and skills (Silva, 1993), the extent to which self-efficacy is associated with writing outcomes is expected to vary as a function of whether English is written as the L1 or L2. Therefore, it is warranted to examine the variations of these observed effect sizes between L1 and L2 writing in English. Previous studies and theories also suggested that the relationship between writing self-efficacy and student writing performance is associated with student or effect size characteristics (sample size, gender, and grade level) and study characteristics (statistical procedures and publication type). Therefore, these covariates need to be controlled for when we examine the moderating effect of L1/L2 writing on the effect size estimates.

2.3.1. Sample size

Sample size is an important component in meta-analyses, which is used to compute the variance and weight of each primary study. Research suggested effect sizes were significantly correlated with sample sizes, regardless of disciplines (Slavin & Smith, 2009). In the review of 185 studies in education, Slavin and Smith (2009) found a significantly negative correlation ($r = -.28$) between effect size and sample size, with effect sizes ranging from .09 in studies with a large sample size ($n > 2000$) to .44 in studies with a small sample size ($n < 50$). Similarly, Cheung and Slavin (2016) examined 645 studies from reviews of preschool, reading, mathematics, and science programs and found studies with a small sample size ($n \leq 250$) had an effect size almost as twice as that yielded by studies with a large sample size ($n > 250$). Therefore, sample size needs to be controlled when examining the moderating effect of L1/L2 writing on effect size estimates.

2.3.2. Gender

Previous research suggested that the magnitude of the relationship between writing self-efficacy and writing achievement varied as a function of gender (Pajares, 2007; Pajares & Valiante, 2001; Pajares, Valiante, et al., 2007). In Pajares and Valiante's (2001) study, the correlation coefficient was .46 for girls and .38 for boys, suggesting writing self-efficacy and writing achievement displayed a stronger relationship for girls than for boys. This finding was replicated in other studies conducted by Pajares and his colleagues, revealing that larger effect sizes were always obtained with female samples (e.g., Pajares, 2007; Pajares, Valiante, et al., 2007). Some researchers attributed gender differences in writing self-efficacy and writing achievement to feminine orientation (Pajares & Valiante, 2001), which is defined as the extent to which people identify themselves with some attributes or characteristics that typically belong to females (Harter et al., 1997). Since gender differences were noted in the magnitude of the relationship between writing self-efficacy and writing achievement, gender needs to be considered when we estimate the effect size of the relationship.

2.3.3. Grade level

Research suggested that students' grade levels are associated with the relationship between writing self-efficacy and writing achievement (Prat-Sala & Redford, 2012; Zhang & Guo, 2012). Zhang and Guo (2012) conducted a study with a sample of 43 freshmen and 23 sophomores at a university in China and noted a significantly positive relationship between self-efficacy and writing proficiency

for freshmen ($r = .430$) but not for sophomores ($r = -.017$), which suggested that writing self-efficacy may have more pronounced effects on students at lower grades or on students with lower writing proficiency. Although Shell et al. (1995) claimed that students' self-efficacy increased as they progressed from primary to high schools, Limpo and Alves (2013) noted that the development in self-efficacy did not correspond with the improvement of their writing achievement. Therefore, we can postulate that grade level may interact with the effect of L1/L2 writing on the relationship between writing self-efficacy and writing achievement.

2.3.4. Statistical procedure

Researchers have been debating over aggregating studies using different statistical procedures (e.g., Rosenthal & DiMatteo, 2001). Effect sizes computed using Pearson correlation and multiple linear regression are on different metrics, which is an impediment to meta-analysts. Effect sizes obtained from different regression models may not be comparable because varying factors are included in these models. However, this problem can be solved by considering the covariates controlled for in different regression models. In simple linear regression, the effect size obtained is the same as that calculated by Pearson correlation. Inclusion of effect sizes from different statistical procedures in a meta-analysis has the advantage of obtaining a more representative sample, reducing sampling errors, increasing the statistical power, and enhancing the validity of meta-analytic results (Peterson & Brown, 2005).

2.3.5. Publication type

Research also suggested that studies with statistically significant results and larger effect sizes are more likely to be published than those with statistically insignificant results and smaller effect sizes (Borenstein et al., 2009; Rothstein et al., 2005). For example, Polanin et al. (2016) conducted a study examining the differences in effect sizes between 81 published and unpublished studies and found that the average effect size estimates in the published studies were larger than those in the unpublished studies in 0.18 standard deviations. Sampling only published documents in meta-analyses would result in the spurious elevation of the true effect sizes (Huang, 2013). Therefore, it is reasonable to assume that publication status is a covariate that needs to be controlled for when examining how L1/L2 writing moderates effect size estimates.

The purpose of the study is to estimate the effect size of the relationship between English writing self-efficacy and writing achievement for L1 or L2 writers and to examine the extent to which writing in English as a L1 or L2 moderates their relationship. Since previous studies provide evidence that the relationship between writing self-efficacy and writing achievement was associated with sample size, gender, grade level, statistical procedure, and publication type, these covariates need to be considered and controlled for in the model when examining the effect of L1/L2 writing on the effect size. Research questions that guided this study were conceptualized as follows:

- What is the magnitude of the relationship between writing self-efficacy and writing proficiency for first and second language writers in English?
- How does writing in English as a first/second language moderate the relationship between writing self-efficacy and writing proficiency?

3. Methodology

3.1. Literature search

In order to get a thorough search and exhaustive inclusion of relevant literature on the relationship between writing self-efficacy and writing achievement, both electronic and manual searches were adopted. First, electronic searches were conducted targeting the following computerized journal databases: Academic Search Complete, Education Research Complete, Eric, ProQuest Dissertations & Theses Global, PsycARTICLES, PsycINFO, Web of Science with citation indexes of the Science Citation Index Expanded (SCI-EXPANDED), Social Science Citation Index (SSCI), and Arts & Humanities Citation Index (A&HCI). We were using the Boolean/Phrase of "SU writing*" AND "SU (efficacy OR self-efficacy OR self-efficacy beliefs OR self-efficacy expectations OR perceived self-efficacy)". Search results were limited to: (a) scholarly (peer-reviewed) articles or theses/dissertations; (b) the time frame of 1977–2019; and (c) being written in English. We targeted articles published in peer-reviewed journals for the sake of quality control and dissertations/theses for the minimization of publication bias. Since self-efficacy was firstly conceptualized by Bandura in his seminal paper in 1977, we decided to set the time frame for searches from 1977 to 2019. Additionally, we did backward reference searching and manually searched the references of some key articles and dissertations/theses to locate relevant studies. The searches yielded 1504 peer-reviewed articles and 251 dissertations/articles, which were imported into Zotero for data storage, management, and duplication checking. We ended up with 1239 peer-reviewed articles and 208 dissertations/these after removing duplicates. Then the 1447 documents were imported into the computer software MAXQDA for screening.

3.2. Literature screening

A two-phase process was applied to determine the eligibility of the searched documents for the present meta-analysis. In the first phase, documents were screened in MAXQDA by applying the three inclusion criteria: (a) topic relevancy; (b) quantitative and empirical studies; and (c) reporting or having computable effect sizes. Topic relevancy means each study has to focus on the students' writing self-efficacy and English writing achievement, so the studies on teacher self-efficacy or self-efficacy in a language other than English or in other domains of language learning (i.e., listening, speaking, reading, general language learning) were excluded from the

present study. We excluded studies in a language other than English because the current meta-analysis focuses on English writing. In addition, as mentioned above, self-efficacy is different from self-concept, self-confidence, and self-esteem (Bandura, 1997), so we decided to exclude studies examining these constructs unless they have a sub-construct of self-efficacy under the umbrella of self-concept. For example, the first exclusion criterion was applied to a study on teacher self-efficacy (e.g., Canbulat, 2017), a study on writing self-efficacy in Spanish (e.g., Villalón et al., 2015), and a study on academic self-confidence (e.g., Pulford et al., 2018). Second, each study had to be an empirically quantitative study, so conceptual or qualitative studies were excluded, based on which, we removed 58 qualitative studies (e.g., Ruan, 2014), literature reviews (e.g., Pajares, 2003), and meta-analysis studies (e.g., Graham et al., 2012). The third inclusion criterion required that each study report effect sizes or had sufficient data to compute effect sizes. A total of 79 studies met the three criteria.

The second phase of the screening process examined studies conducted by the same researchers or the same research teams. Studies using the same data were combined to remove the duplicates of study-level information. We ended up with 76 studies after we combined two studies (i.e., Pajares, 2007; Pajares, Johnson, et al., 2007) and three other studies (i.e., Pajares & Valiante, 2001; Pajares et al., 2000, 2001). A total of 565 effect sizes were reported in these 76 studies.

Approximately 10 % of studies ($n = 150$) were independently screened in MAXQDA by a doctoral student. We obtained the initial agreement of .98, indicating satisfactory inter-rater reliability. The discrepancies were resolved through review and discussions. The six inclusion criteria (three in the literature searching phase and three in the screening phase) are presented in Table 1. The PRISMA flow diagram of literature search and screening processes are presented in Fig. 1 (Moher et al., 2009).

3.3. Data coding

Two coding protocols addressing both the sample characteristics and the statistical characteristics of all the studies were developed by the first two authors and consulted with the third author (see Appendix A). The first coding protocol included the following descriptors: (1) Identification of Studies (study ID number, author), (2) Sample Description (e.g., sample size), (3) Variables and Method (e.g., statistical procedure), and (4) Effect Size Statistics. The second coding protocol is at the article-level, including Study Identification Number, Author(s), Publication Type, Publication Year, Grade, L1 Writing, and Country of Data Collection. A subset of 10 studies (89 effect sizes) was randomly sampled and coded independently by a doctoral student to ensure the quality and consistency of coding. We obtained intra-class correlation (ICC) ranging from .95 to 1.00 for continuous variables and Kappa coefficients ranging from .97 to 1.00 for categorical variables (Orwin & Vevea, 2009), indicating satisfactory inter-rater reliability. The predictor variables that were included in the model were coded as follows.

Size refers to the sample size of students with whom each effect size was calculated;

Gender was coded as the proportion of female students in each sample;

Statistical Procedure refers to the procedure (i.e., correlation or regression) that was used to estimate an effect size. Various variables were included in the regression models of the primary studies and their differences were controlled by considering the types of these covariates. We classified these covariates into three categories of demographic variables (e.g., age, SES), affective variables (e.g., self-concept, anxiety), and achievement variables (e.g., prior writing achievement, reading achievement), based on Onwuegbuzie et al. (2000) classification.

Demographic, Affective, and Achieving were coded as 1 when a regression model controls for demographic, affective, and achievement variables, respectively. Such that correlation is the condition when the three variables were all coded as 0.

Publication Type was coded as 0 for journal articles and 1 for dissertations/theses.

Grade was coded as 1 for elementary school, 2 for middle school, 3 for high schools, and 4 for college/university. Based on the distribution of *Grade*, two dummy variables, *elementary* and *midhigh* were created. *Elementary* was coded as 1 for elementary level and *midhigh* was coded as 1 for middle/high level. College/university level was the reference category in the two variables.

L1 writing was dummy coded as 0 referring to L2 writers, and 1 referring to L1 writers. As 13 studies included both L1 and L2 writers and we could not distinguish them, we ran two models (one with the 13 studies and one without the 13 studies) to see if the results are the same.

Table 1
Inclusion and Exclusion Criteria.

Criteria	Inclusion	Exclusion
Publication Type	Articles from peer-reviewed journals or dissertation/theses.	Articles from non-peer-reviewed journals or technical report.
Publication Period	1977−2019.	Prior to 1977 or after 2019.
Language	Written in English.	Written in languages other than English.
Topic Relevancy	Student self-efficacy and its relationship with writing achievement.	Teacher self-efficacy; Self-efficacy in languages other than English; Self-efficacy in other domains of language learning.
Research Method	Empirical and quantitative studies.	Conceptual studies and qualitative studies.
Sufficient Data	Reporting effect sizes or having adequate data to compute effect sizes.	Not reporting needed effect sizes or not having adequate data to compute effect sizes.

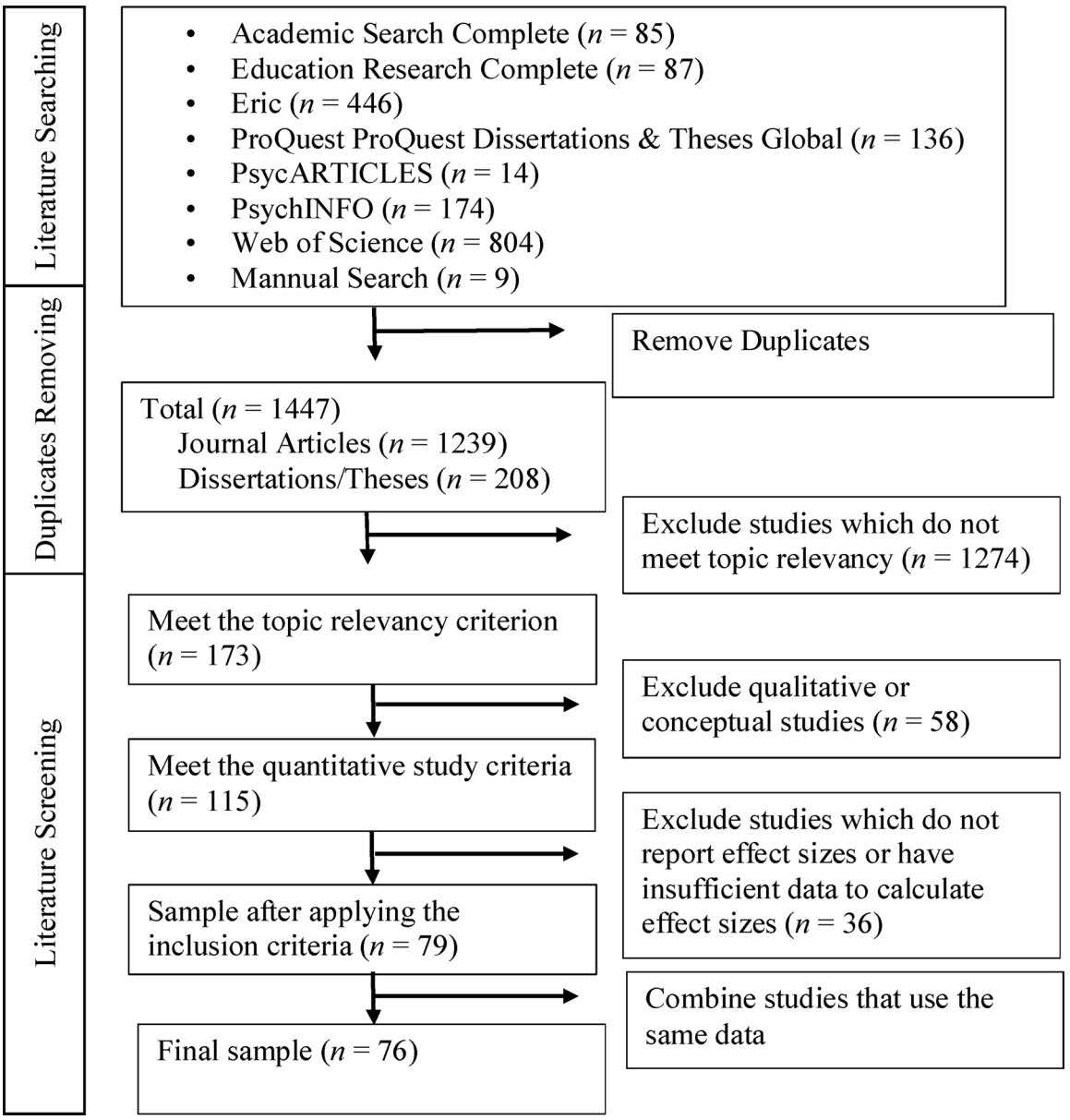

Fig. 1. PRISMA Flow Diagram of Literature Search and Screening.

3.4. Publication Bias

The issue of publication bias arises when the studies collected in a meta-analysis are not representative of the studies in the population (Rothstein et al., 2005). The following three approaches were explored to address this problem of publication bias. First, we included both published studies and unpublished studies (i.e., dissertations and thesis) to minimize the probability of publication bias. Second, since comprehensive and thorough research is the best way to deal with publication bias (Borenstein et al., 2009), we searched seven databases and citation indexes with a time span of more than 50 years. Third, we conducted Classic Fail-Safe *N* and Orwin's Fail-Safe *N* to assess the likelihood of the presence of publication bias (Borenstein et al., 2009). The Classic Fail-Safe *N* revealed that 6407 additional studies with an effect size of zero were needed to yield a nonsignificant result, while Orwin's Fail-Safe *N* showed that 370 missing studies with an effect size of zero were needed to bring the cumulative effect size to a trivial level of .05. Although there is no criterion for evaluating these values, Rosenthal suggested that if the Classic Fail-Safe *N* is large than $5k+10$ with *k* being the number of primary studies, publication bias is not a serious concern (Becker, 2005). Therefore, all these approaches provided evidence that the current meta-analysis is robust to publication bias.

3.5. Missing values

There were 13 missing values for the variable of the proportion of female students. We adopted the method of mean imputation, as Pigott (2012) suggested it as a viable method in handling missing values. Among the potential strategies to deal with missing data (e.g., listwise deletion, pairwise deletion, mean imputation, and regression imputation), listwise deletion and pairwise deletion have the risk of decreasing the power of the analysis, whereas regression imputation is biased in the estimation of standard errors (Pigott, 2012). Although mean imputation tends to underestimate the variance of this variable, the impact of this underestimation is minimal due to the small proportion of missing values (13 out of 565, which is the total number of effect sizes extracted).

3.6. Data analysis

The effect sizes used in this meta-analysis are Pearson's correlation coefficient, *r*. Pearson correlation, a measure of the linear bivariate relation between two variables, was recommended as the effect size index for synthesizing correlational studies by Borenstein et al. (2009). Studies that report the standardized beta coefficient (β) were also included in the present meta-analysis as suggested by Peterson and Brown (2005) since bivariate correlation coefficient equals the standardized regression coefficient in a bivariate regression model. The difference between *r* and β in multivariate regression models was accounted for by indicating the types of variables included in the regression models (i.e., demographic, affective, or achievement variables). *r* extracted from primary studies was transformed into Fisher's z metric to normalize the sampling distribution of observed effect sizes because meta-analysis assumes the normal distribution of observed effect sizes. Furthermore, *r*-to-z transformation removes the dependence of sampling variances of observed effect sizes on the unknown correlation parameter (ρ), which will lead to a more accurate estimation of variances (Borenstein et al., 2009). The *r*-to-z transformation formula was noted as follows.

$$z = 0.5 \times \ln\left(\frac{1+r}{1-r}\right) \tag{1}$$

After the analysis, estimated z values were transformed back to *r* for purposes of reporting and interpreting results using the following formula. The notations used here were borrowed from Borenstein et al. (2009).

$$r = \frac{e^{2z}-1}{e^{2z}+1} \tag{2}$$

A meta-regression analysis using a hierarchical linear model (HLM) was employed for data analysis. An HLM model was adopted because the data for meta-analysis were structured hierarchically: effect sizes were nested within studies (Raudenbush & Bryk, 2002). Multi-level meta-regression models can aid in estimating the average effect sizes across studies and predicting the effects of multiple covariates on the effect sizes at both within-study and between-study levels simultaneously.

Two two-level HLM models using R package *lme4* and the estimation method of Maximum Likelihood were estimated (Bates et al., 2015). The unconditional model was constructed first to estimate the overall average effect size and variance components. This model was an intercept-only model and included no predictors for both Level 1 and Level 2. Second, a conditional model with all the predictors entered was constructed to estimate the moderating effects of these predictors on the effect size. Level 1 predictors included sample size, the proportion of female students, and types of variables included in regression models (i.e., demographic, affective or achievement), which were variables associated with effect size characteristics within studies. The method of Pearson correlation was the reference variable against which each regression method was compared. Level 2 predictors include publication type, grade (*elementary* and *midhigh*), and L1 writing, which were variables related to study-level characteristics. The conditional model was a random intercept and fixed slope model because empirical evidence suggested that level-2 predictors had moderating effects on the effect size (the intercept), as illustrated in the literature review section. Our knowledge of whether level-2 predictors impacting level-1 slopes (the effect of level-1 predictors on the effect size) is limited so we fixed all the level-1 slopes. All the level-1 predictors were group-mean centered to produce an intercept value that equals the average of within-study effect sizes. All level-2 predictors were grand-mean centered. The meta-regression HLM was represented as follows.

Table 2
Frequency Distribution of Studies by Publication Type, Research Setting, L1/L2 Writing, and Statistical Procedures.

Variables		Frequency	Percentage
Publication Type	Journal Articles	55	72.4
	Dissertation/Thesis	21	27.6
Grade	College	47	61.8
	Middle/High	17	22.4
	Elementary	14	18.4
L1/L2 Writing	L2 Writing	18	23.7
	L1 Writing	58	76.3
Method	Regression Control for DEV	72	12.7
	Regression Control for AFV	102	18.1
	Regression Control for ACV	50	8.8
	Correlation	413	73.1

Note. DEV = demographic variables, AFV = affective variables, ACV = achievement variables.

Level 1:

$$Effect\ Size_{ij} = \beta_{0j} + \beta_{1j}*(SampleSize_{ij}) + \beta_{2j}*(ProportionFemale_{ij}) + \beta_{3j}*(Regression\_Demographic_{ij}) + \beta_{4j}*(Regression\_Affective_{ij}) + \beta_{5j}*(Regression\_Achieve_{ij}) + r_{ij}$$

Level 2:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}*(\boldsymbol{Type\_DisserThesis_j}) + \gamma_{02}*(\boldsymbol{Grade\_elementary_j}) + \gamma_{02}*(\boldsymbol{Grade\_midhigh_j}) + \gamma_{03}*(\boldsymbol{L1Writing_j}) + u_{0j}$$

$$\beta_{pj} = \gamma_{p0}; p = 1,\ 2,\ 3,\ 4$$

4. Results

This meta-regression synthesized 565 effect sizes nested within 76 primary studies. Descriptive statistics of the independent variables are presented in Table 2. Sample sizes of the 565 effect sizes ranged from 8 to 2037, with a mean of 245.03 ($SD = 247.77$). The average proportion of female students was 57.98 % ($SD = 0.22$). Effect sizes were predominantly calculated using Pearson correlation ($n = 413$, 73.1 %). Of these 76 studies, the majority of studies ($n = 55$, 72.4 %) were published journal articles. There were 18 studies (23.7 %) on L2 writing. There were 14 (18.4 %) studies conducted with elementary school students and 17 (22.4 %) studies conducted with middle and high school students.

Meta-analysis assumes a normal distribution of the observed effect sizes (Borenstein et al., 2009). The distribution of Fishers' z was indicated that the effect sizes were approximately normally distributed with a few outliers on the upper end of the distribution. The unconditional model revealed that the overall average of z-transformed effect sizes was 0.30 ($r = .29$). The within-study variance component was 0.024, and the between-study variance component was 0.046. The intraclass correlation coefficient (ICC) was 0.66, suggesting that about 66 % of the variation in the z-transformed effect sizes was between studies. Thus, it was deemed reasonable to proceed with the two-level HLM.

The estimates of the predictors in the conditional model are presented in Table 3. Note that the values in this Table 3 are intentionally presented with more decimal places because the coefficients of some variables (e.g., sample size) were very small. Sample size was a positive and significant predictor of effect size estimates, $t(473.94) = 2.71$, $p < .01$. Regression controlling for demographic variables did not have a significant effect on effect sizes, $t(473.94) = -1.38$, $p > .05$. However, regression controlling for affective variables significantly predicted effect sizes, $t(473.94) = -3.34$, $p = .001$. This estimate of -0.08 suggested that effect sizes estimated in regression models after controlling for affective variables were 0.08 smaller than effect sizes estimated in other models. Likewise, regression controlling for achievement variables tended to yield effect size estimates, which were statistically significantly smaller than other methods, $t(473.94) = -2.66$, $p < .01$.

Among level-2 predictors, L1 writing was a significant predictor of effect sizes, $t(63.78) = -3.67$, $p < .001$. Studies with L2 writers were 0.24 larger in effect size estimates than studies with L1 writers. There were no statistically significant differences in effect size estimates between elementary students and students in other levels of schooling, $t(56.37) = 0.70$, $p > .05$. Neither were differences between middle and high school students and students of other levels in effect size estimates, $t(56.69) = 1.61$, $p > .05$. Contrary to our expectation, publication status was not a significant predictor of the magnitude of effect sizes, $t(58.86) = -0.83$, $p > .05$.

The effect sizes under certain conditions were also estimated. Table 4 presents both the z-transformed estimates, and r estimates that were calculated using Eq. (2). We did not include the conditions for the continuous variables but employed the average sample size and proportion of female students to calculate these estimates. Effect size estimates did not vary greatly across different conditions for the predictors of publication type and grade level. However, there were appreciable differences in effect size estimates for the factors of whether writing in L1 or L2 and statistical methods. The effect size estimate (r) for studies on L2 writing was 0.441, whereas that for studies on L1 writing was 0.233. The effect size estimate using correlation (0.289) was also noticeably larger than those estimated in regression models after controlling for affective variables (0.196) or achievement variables (0.177).

Table 3
Predictors of Effect Size Magnitude.

	Estimate	*SE*	*t*
Level 1—Within-Study Characteristics			
Sample Size	0.000149**	0.000055	2.710
Proportion of Females	−0.011701	0.040176	−0.291
Regression_Demographic	−0.053200	0.038666	−1.376
Regression_Achievement	−0.096598**	0.036281	−2.662
Regression_Affective	−0.083147***	0.024896	−3.340
Level 2—Between-Study Characteristics			
Intercept	0.267472***	0.029178	9.167
Publication Type_dissertation/thesis	−0.046408	0.056286	−0.825
Elementary	0.046278	0.065823	0.703
Midhigh	0.099471	0.061689	1.612
L1 Writing	−0.236479***	0.064369	−3.674

Note. *$p < .05$, **$p < .01$, ***$p < .001$.

Table 4
Model-Based Effect Size Estimates.

Effect Size Estimates	z_r Transform	r
Overall Estimate	0.267	0.261
Publication Type		
Journal Articles	0.292	0.284
Dissertation/Theses	0.246	0.241
Grade Level		
Elementary	0.299	0.290
Middle/High	0.344	0.331
College	0.229	0.225
L1/L2 Writing		
L2 Writing	0.473	0.441
L1 Writing	0.237	0.233
Method		
Pearson Correlation	0.298	0.289
Regression Controlling for DEV	0.221	0.217
Regression Controlling for AFV	0.199	0.196
Regression Controlling for ACV	0.179	0.177

Note. DEV = demographic variables, AFV = affective variables, ACV = achievement variables.

The results of the 63 studies after the removal of the 13 studies with both L1 and L2 writers are presented in Appendix B. The direction and significance of the estimates for all predictors in the two levels were consistent with those of the overall set of data, which provided evidence for the robustness of our results.

5. Discussion

The purposes of the present study were: (a) to estimate the overall average effect size of the relationship between English writing self-efficacy and writing achievement with L1 and L2 writers in English; and (b) to examine the extent to which writing in English as an L1 or L2 impacts the strength of the relationship. The hierarchical linear meta-regression indicated a statistically significant and positive relationship ($r = .29$) between writing self-efficacy and English writing achievement with both L1 and L2 writers. According to Cohen' (1988) benchmark, effect size is small when $r = .10$, medium when $r = .30$, and large when $r = .50$. The effect size estimate in the present study indicated a medium-sized effect for the relationship between writing self-efficacy and students' writing achievement, which confirmed findings from a plethora of previous studies (e.g., Bruning et al., 2013; Pajares, 2007; Pajares & Valiante, 2001; Prat-Sala, & Redford, 2012; Teng et al., 2018; Woodrow, 2011; Wright et al., 2019; Zhang & Guo, 2012). The average correlation of .29 indicated approximately 9% of the variability in writing achievement is associated with the variability in students' self-efficacy, with higher writing achievement associated with higher self-efficacy. This moderate effect size does not negate the substantive significance of self-efficacy if we take the content area into consideration. Lipsey and Wilson (1993) reviewed 302 meta-analyses of psychological, educational, and behavioral treatment research and found an overall average effect size of $r = .25$ (converted from $d = .50$). Plonsky and Oswald (2014) also suggested an average effect size of .21 in social psychology. This implies in the psychological, educational, and behavioral field, the effect size (r) of .29 is far from being trivial, and writing self-efficacy has practical significance to students' writing achievement.

The current study reported statistically significant differences between L1 writing ($r = .233$) and L2 writing ($r = .441$) in effect size estimates, with a larger effect size with L2 writers. One explanation might be that self-efficacy plays a more facilitative role in writing achievement among low self-efficacious students. Given L1 learners have already had higher levels of self-efficacy beliefs (Scholz et al., 2002), the effect of self-efficacy on writing is diminished because of the ceiling effect. Since self-efficacy is more strongly associated with writing achievement for L2 students, it is imperative that English language teachers in English as a foreign/second language (EFL/ESL) contexts employ effective strategies to nurture L2 writers' self-efficacy.

The present study also reported a positive relationship between sample size and effect size magnitude, which contradicted the observation that effect sizes and sample sizes were negatively correlated (Slavin & Smith, 2009). Methodologists contend that studies with smaller sample sizes tend to report larger effect sizes (Lipsey & Wilson, 1993; Slavin & Smith, 2009). This is possibly due to publication bias since studies have little chance of being published if they do not find statistically significant results (Rothstein et al., 2005), and statistical significance for studies with smaller sample sizes can only be detected when effect sizes are large (Slavin & Smith, 2009). The discrepancy between the present study and the previous research can be attributed to the fact that the present study encompassed both studies published in journal articles and unpublished dissertations and theses, and the incorporation of the latter discount the influence of publication bias on effect sizes. This finding suggested that including both published and unpublished studies in systematic reviews is a viable alternative to the minimization of publication bias. Further research is recommended to examine what factors (e.g., publication status, sampling methods, research design) might moderate the relationship between sample sizes and effect sizes in order to obtain conclusive evidence.

This study failed to detect the moderating effect of gender on effect sizes, which suggested that the magnitude of the relationship between writing self-efficacy and writing achievement was the same for females and males. Previous research noted that significant gender differences in writing self-efficacy and writing achievement in favor of female students were nullified when feminine

orientation was controlled (Pajares & Valiante, 2001). Feminine orientation is defined as the extent to which people identify themselves with some attributes or characteristics that typically belong to females (Harter et al., 1997). Writing was believed to be a stereotypically female activity (Pajares & Valiante, 2001). Therefore, it is gender orientation rather than gender that moderates the relationship between writing self-efficacy and writing achievement. Future research is recommended to add gender orientation to the HLM model and investigate its effect on effect sizes.

This study failed to find a moderating effect of grade level on the correlation between writing self-efficacy and writing achievement, which was consistent with Cheung and Slavin's (2016) study. Cheung and Slavin analyzed 581 studies and found no statistically significant differences in effect sizes between elementary and secondary students. Multon et al. (1991) noted that younger students might not have an accurate estimate of their self-efficacy beliefs, and this discordance between their self-efficacy beliefs and actual achievement would weaken the relationship. However, the present study did not support this notion and indicated that students' writing self-efficacy developed proportionately with the improvement of their writing skills and abilities across educational settings. Since research suggested a curvilinear pattern of writing self-efficacy development (Valiante, 2001), the categorization of all grade levels into three groups might have concealed the pattern of effect size change across each grade level. Further research should be recommended to examine the trajectory of students' writing self-efficacy development and effect size change across grade levels using cross-lagged models.

This study noted statistically significant differences in effect sizes as a function of statistical procedures, with Pearson correlation yielding larger effect sizes than regression. This result was expected because Pearson correlation represents the relationship between the two variables of interest only, whereas the slopes in regression (i.e., β) represents the common variance between the two variables of interest when other covariates are controlled at the same level. The effect sizes reported in regressions were adjusted for the correlation between writing self-efficacy and other predicting variables and for the correlation between writing achievement and these variables (Kline, 2016), which may explain the attenuation in effect sizes. The statistically significantly smaller effect sizes reported in regression models after controlling for affective variables and achievement variables may convey two messages. First, writing self-efficacy tends to covary with affective or achievement variables (Pajares et al., 2000). Second, there are significant correlations between writing achievement and other affective and achievement variables (Zhang & Guo, 2012). Our results also suggest that demographic variables did not moderate the relationship between writing self-efficacy and writing achievement.

To our surprise, this study failed to detect the moderating effect of publication status on the relationship between self-efficacy and writing achievement, which is inconsistent with the contention that published studies tended to report larger effect sizes than the unpublished studies (Borenstein et al., 2009; Polanin et al., 2016; Rothstein et al., 2005). However, this finding supported Huang's (2013) study, which found that publication status did not impact the effect sizes. We further found that although there were no significant differences between published and unpublished studies in average effect size magnitude, unpublished studies are more heterogeneous than published studies in effect size estimates, implying that although publication status may not be a moderating factor, published studies are more reliable in the estimation of effect sizes than unpublished studies.

6. Limitation and future research

The first limitation arises from the inclusion of only journal articles and dissertations/theses in the present meta-analysis. Other grey literature such as conference abstracts/presentations, books and book chapters, unpublished technical reports, or white papers is suggested to be included for purposes of obtaining more representative data and further minimizing publication bias. Second, the two variables (writing self-efficacy and writing achievement) were not operationalized identically across the 76 studies. Some studies employed researcher/teacher-developed measures while others employed standardized measures. Future studies are called for to account for measurement errors in meta-analyses. Moreover, this study focused on English language writing self-efficacy and excluded studies of other languages. As the motivation to learn English and languages other than English (LOTEs) diverges (Dörnyei & Al-Hoorie, 2017), the effect size of the relationship between self-efficacy and writing achievement in LOTEs deserves investigation. Future research is also suggested to estimate the effect size with bilinguals in EFL/ESL contexts or to examine the moderating effect of students' characteristics such as the dominant language spoken at home.

7. Conclusion and implication

Social cognitive theory posits that affective factors such as self-efficacy are essential to human functioning (Bandura, 1986, 1997). The importance of self-efficacy to students' academic achievement and language proficiency has been discussed for decades (Huang et al., 2015). This study contributes to the literature in the writing domain by addressing the statistical and practical significance of this construct to writing outcomes. The present meta-analysis provides evidence to corroborate the contention that self-efficacy in writing is positively associated with students' writing performance. The meta-analytic results provide classroom teachers, practitioners, and policymakers with evidence-based guidance in formulating and implementing appropriate interventions to enhance students' self-efficacy in writing.

The current meta-analysis has pedagogical implications. The current study provided evidence that the relationship between writing self-efficacy and English writing achievement was stronger for L2 writers ($r = .441$). This implies that writing self-efficacy plays a more significant role in L2 writing. Therefore, it is imperative for L2 teachers to heighten their awareness of nurturing students' self-efficacy in writing. Bandura (1997) summarized four sources for the development and nurture of self-efficacy beliefs, namely, mastery or enactive experience, vicarious experience, social persuasion, and physiological and emotional states. Therefore, classroom teachers are encouraged to boost students' writing self-efficacy through various strategies. For example, they need to provide more resources and

create more opportunities for students to achieve success in writing, provide opportunities for students to learn from their peers through group work and modeling, offer encouraging and positive feedback as they work on challenging tasks, and care for negative feelings students may have in the process of writing. The medium to large effect size of the relationship between self-efficacy and English writing achievement with L2 writers identified by the current meta-analysis also calls for the need to develop and implement appropriate interventions to improve L2 writers' self-efficacy in writing. Literature suggests that writing self-efficacy can be enhanced by the self-regulated strategy development (SRSD) intervention (MacArthur et al., 2013), a criterion-based instruction incorporating explicit instruction of strategies, knowledge, and self-regulation procedures (Harris et al., 2015). Since SRSD was developed for L1 writers and examined predominantly in L1 English writing, more studies on SRSD in L2 writing are needed to inform program designers, educators, and classroom teachers of adaptations in the design and implementation of the intervention in the L2 English writing context after consideration of unique features of L2 writing.

The current study also has implications for primary researchers and research methodologists. Evidence from the current meta-analysis showed that effect sizes vary across statistical procedures and regression models. We found that effect sizes tend to attenuate in regression models after controlling for affective variables and achievement variables, which implied that these variables covary with writing self-efficacy or writing achievement. Therefore, these covariates need to be examined before investigating the relationship between writing self-efficacy and writing achievement. This implication is important for researchers to uncover the spurious relationship between writing self-efficacy and writing achievement.

Declaration of Competing Interest

The authors report no declarations of interest.

Acknowledgments

Support for this research was provided in part by the Research & Development Grant for Chair Professor of the University of Macau (CPG2020-00017-FED). The opinions expressed do not necessarily reflect the position or policy of the University of Macau, and no official endorsement should be inferred.

Appendix A. Codebook

Level-1 Codebook

Section A: Identification of Studies

1 Study Identification Number (ID)
2 Author(s) (author)

Section B: Sample Description

1 Sample Size (samsize)
2 Proportion of Female Students (porfe)
3 Sample Description (samdes)

Section C: Variables and Method

1 Variable1: Writing Self-Efficacy (seleff)
2 Variable2: Writing Achievement (wriachi)
3 Writing Outcome Measure Name (wriname)
4 Writing Self-Efficacy Measure (effiname)
5 Statistical procedure (procedure)
  1 Pearson correlation
  2 Regression
6 11. Demographic (demo)
  1 Not included
  2 Included
7 Affective (cogni)
  1 Not included
  2 Included
8 Achieving (achi)
  1 Not included
  2 Included

Section D: Effect Size Statistics

1 Type of Effect Sizes (ES_Type)
  1 Pearson correlation coefficient *r*
  2 Standardized regression coefficient β
  3 Other type of effect sizes (Specify)
2 Effect Size Statistics (ES)

Level-2 Codebook

1 Study Identification Number (ID)
2 Author(s) (author)
3 Publication Type (type)
  1 Journal article
  2 Dissertation/thesis
4 Publication Year (year)
5 Grade (grade)
  1 Elementary school
  2 Middle school
  3 High school
  4 College/University
6 L1 Writing
  1 L2 writing
  2 L1 writing
7 Country of Data Collection (counties)

Appendix B. Predictors of Effect Size Magnitude after Removing Studies with both L1 and L2 Writers

	Estimate	*SE*	*t*
Level 1—Within-Study Characteristics			
Sample Size	0.000111*	0.000057	1.960
Proportion of Females	−0.025280	0.051770	−0.488
Regression_Demographic	−0.035630	0.041610	−0.856
Regression_Achievement	−0.115900**	0.041430	−2.798
Regression_Affective	−0.091970***	0.027230	−3.377
Level 2—Between-Study Characteristics			
Intercept	0.262400***	0.033250	8.795
Publication Type_dissertation/thesis	−0.053910	0.069150	−0.780
Elementary	0.041980	0.081030	0.518
Midhigh	0.096810	0.074630	1.297
L1 Writing	−0.222000**	0.075160	−2.954

Note. *$p < .05$, **$p < .01$, ***$p < .001$.

Appendix C. Supplementary data

Supplementary material related to this article can be found, in the online version, at doi:https://doi.org/10.1016/j.jslw.2021.100817.

References

Amogne, D. (2008). *An investigation of the correlation among efficacy sources, students' self- efficacy and performance in reading and writing skills: Bahir Dar University in focus (Unpublished master's thesis)*. Addis Ababa, Ethiopia: Addis Ababa University.
Anastasiou, D., & Michail, D. (2013). Exploring discourse between self-efficacy and writing performance among low-literate adult students. *Learning Disabilities: A Contemporary Journal, 11*(1), 53–87.
Archibald, A. (2016). *Writing in a second language*. Retrieved from. Center for Language Linguistics and Area Studies https://www.llas.ac.uk/resources/gpg/2175.html.
Asmari, A. A. (2013). Investigating of writing strategies, writing apprehension, and writing achievement among Saudi EFL-major students. *International Education Studies, 6*(11), 130–143. https://doi.org/10.5539/ies.v6n11p130.
Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.
Bandura, A. (1997). *Self-efficacy: The exercise of control*. Freeman.

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software, 67*(1), 1–48. https://doi.org/10.18637/jss.v067.i01.
Becker, B. J. (2005). Failsafe *N* or file-drawer number. In H. R. Rothsetin, A. J. Sutton, & M. Borenstein (Eds.), *Publication bias in meta-analysis prevention, assessment and adjustments* (pp. 111–125). Wiley.
Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to meta-analysis*. John Wiley & Sons.
Bruning, R., Dempsey, M., Kauffman, D. F., McKim, C., & Zumbrunn, S. (2013). Examining dimensions of self-efficacy for writing. *Journal of Educational Psychology, 105*(1), 25–38. https://doi.org/10.1037/a0029692.
Brunstein, J., & Glaser, C. (2011). Testing a path-analysis mediation model of how self-regulated writing strategies improve fourth graders' composition skills" A randomized controlled trial. *Journal of Educational Psychology, 103*(4), 922–938. https://doi.org/10.1037/a0024622.
Canbulat, A. N. (2017). Classroom teacher candidates' perceptions of teacher self-efficacy in developing students' reading, writing and verbal skills: Scale development study. *Educational Research and Reviews, 12*(16), 789–800. https://doi.org/10.5897/err2017.3305.
Chae, S. E. (2013). Effects of self-efficacy and interest on college students' L2 writing strategy use and performance. *English Teaching, 68*(4), 43–64. https://doi.org/10.15858/engtea.68.4.201312.43.
Chambers, E. A. (2004). An introduction to meta-analysis with articles from the Journal of Educational Research (1992-2002). *The Journal of Educational Research, 98*(1), 35–45. https://doi.org/10.3200/joer.98.1.35-45.
Chan, J. C., & Lam, S. (2008). Effects of competition on students' self-efficacy in vicarious learning. *The British Journal of Educational Psychology, 78*, 95–108. https://doi.org/10.1348/000709907X185509.
Chea, S. (2012). *The relationships among writing self-efficacy, writing goal orientation, learning strategies, and writing achievement (Publication No. 1513414) [Master's thesis, Northern Illinois University]*. ProQuest Dissertations Publishing.
Chen, J., & Zhang, L. J. (2019). Assessing student-writers' self-efficacy beliefs about text revision in EFL writing. *Assessing Writing, 40*, 27–41. https://doi.org/10.1016/j.asw.2019.03.002.
Cheung, A. C., & Slavin, R. E. (2016). How methodological features affect effect sizes in education. *Educational Researcher, 45*(5), 283–292. https://doi.org/10.3102/0013189x16656615.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Erlbaum.
Dörnyei, Z., & Al-Hoorie, A. H. (2017). The motivational foundation of learning languages other than global English: Theoretical issues and research directions. *The Modern Language Journal, 101*(3), 455–468. https://doi.org/10.1111/modl.12408.
Galbraith, D., & Vedder, I. (2019). Methodological advances in investigating L2 writing processes: Challenges and perspectives. *Studies in Second Language Acquisition, 41*, 633–645. https://doi.org/10.1017/S0272263119000366.
Goldberg, A., Russell, M., & Cook, A. (2003). The effect of computers on student writing: A meta-analysis of studies from 1992 to 2002. *The Journal of Technology, Learning, and Assessment, 2*(1), 1–55. Available from http://www.jtla.org.
Graham, S. (2006). Writing. In P. Alexander, & P. Winne (Eds.), *Handbook of educational psychology* (pp. 457–478). Erlbaum.
Graham, S., Hebert, M., & Harris, K. R. (2015). Formative assessment and writing: A meta-analysis. *The Elementary School Journal, 115*(4), 523–547. https://doi.org/10.1086/681947.
Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology, 104*(4), 879–896. https://doi.org/10.1037/a0029185.
Harris, K. R., Graham, S., & Adkins, M. (2015). Practice-based professional development and Self-Regulated Strategy Development for Tier 2, at-risk writers in second grade. *Contemporary Educational Psychology, 40*, 5–16. https://doi.org/10.1016/j.cedpsych.2014.02.003.
Harter, S., Waters, P., & Whitesell, N. (1997). Lack of voice as a manifestation of false self- behavior among adolescents: The school setting as a stage upon which the drama of authenticity is enacted. *Educational Psychologist, 32*, 153–173. https://doi.org/10.1207/s15326985ep3203_2.
Hayes, J. R. (2000). A new framework for understanding cognition and affect in writing. In R. Indrisano, & J. R. Squire (Eds.), *Perspectives on writing: Research, theory, and practice* (pp. 6–44). International Reading Association.
Huang, C. (2013). Gender differences in academic self-efficacy: A meta-analysis. *European Journal of Psychology of Education, 28*, 1–35. https://doi.org/10.1007/s10212-011-0097-y.
Huang, P., Wang, C., & Xie, L. (2015). Self-efficacy beliefs and self-regulated learning strategies of Chinese students learning English as a second language and as a foreign language. In C. Wang, W. Ma, & C. Martin (Eds.), *Chinese education from the perspectives of American educators: Lessons learned from study-abroad experiences* (pp. 179–197). Information Age Publishing.
Kang, E. Y., & Han, Z. (2015). The efficacy of written corrective feedback in improving L2 written accuracy: A meta-analysis. *The Modern Language Journal, 99*(1), 1–18. https://doi.org/10.1111/modl.12189.
Kline, R. B. (2016). *Principles and practice of structural equation modeling*. Guilford.
Limpo, T., & Alves, R. A. (2013). Modeling writing development: Contribution of transcription and self-regulation to Portuguese students' text generation quality. *Journal of Educational Psychology, 105*, 401–413. https://doi.org/10.1037/a0031391.
Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *The American Psychologist, 48*, 1181–1209. https://doi.org/10.1037/0003-066x.48.12.1181.
MacArthur, C. A., & Philippakos, Z. A. (2013). Self-regulated strategy instruction in developmental writing: A design research project. *Community College Review, 41*(2), 176–195. https://doi.org/10.1177/0091552113484580.
McGovern, A. R. (2004). *The influence of mental simulation in an expressive writing context on academic performance and the moderating influence of self-efficacy, optimism, and gender (Publication No. 3162697) [Doctoral dissertation, Oklahoma State University]*. ProQuest Dissertations Publishing.
Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & The PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *Annals of Internal Medicine, 151*(4), 264. https://doi.org/10.7326/0003-4819-151-4-200908180-00135.
Multon, K., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology, 38*(1), 30–38. https://doi.org/10.1037/0022-0167.38.1.30.
National Center for Education Statistics. (2012). *The nation's report card: Writing 2011 (NCES 2012-470)*. Retrieved from. Washington, D.C: Institute of Education Sciences, U.S. Department of Education https://nces.ed.gov/nationsreportcard/pdf/main2011/2012470.pdf.
National Commission on Writing. (2004). *Writing: A ticket to work…or a ticket out: A survey of business leaders*. September, Retrieved from http://www.writingcommission.org/report/html.
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.
Orwin, R. G., & Vevea, J. L. (2009). Evaluating coding decisions. In H. Cooper, L. V. Hedges, & J. C. Valentine (Eds.), *The handbook of research synthesis* (pp. 177–203). Russell Sage Foundation.
Pajares, F. (2003). Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading and Writing Quarterly, 19*(2), 139–158. https://doi.org/10.1080/10573560308222.
Pajares, F. (2007). Empirical properties of a scale to assess writing self- efficacy in school contexts. *Measurement and Evaluation in Counseling and Development, 39*(4), 239–249. https://doi.org/10.1080/07481756.2007.11909801.
Pajares, F., & Valiante, G. (2001). Gender differences in writing motivation and achievement of middle school students: A function of gender orientation? *Contemporary Educational Psychology, 26*, 366–381. https://doi.org/10.1006/ceps.2000.1069.
Pajares, F., Britner, S. L., & Valiante, G. (2000). Relationship between achievement goals and self-beliefs of middle school students in writing and science. *Contemporary Educational Psychology, 25*, 406–422. https://doi.org/10.1006/ceps.1999.1027.
Pajares, F., Harley, J., & Valiante, G. (2001). Response format in writing self-efficacy assessment: Greater discrimination increases prediction. *Measurement and Evaluation in Counseling and Development, 33*, 214–221. https://doi.org/10.1080/07481756.2001.12069012.

Pajares, F., Johnson, M. J., & Usher, E. L. (2007). Sources of writing self-efficacy beliefs of elementary, middle, and high school students. *Research in the Teaching of English, 42*(1), 104–120.
Pajares, F., Valiante, G., & Cheong, Y. F. (2007). Writing self-efficacy and its relation to gender, writing motivation, and writing competence: A developmental perspective. In G. Rijlaarsdam, P. Boscolo, & S. Hidi (Eds.), *Studies in writing: Writing and motivation* (Vol. 19, pp. 141–159). Elsevier.
Peterson, R. A., & Brown, S. (2005). On the use of Beta coefficients in meta-analysis. *The Journal of Applied Psychology, 90*(1), 175–181. https://doi.org/10.1037/0021-9010.90.1.175.
Pigott, T. D. (2012). *Advances in meta-analysis*. Springer-Verlag.
Plonsky, L., & Oswald, F. (2014). How big is "big"? Interpreting effect sizes in L2 research. *Language Learning, 64*(4), 1–35. https://doi.org/10.1111/lang.12079.
Polanin, J. R., Tanner-Smith, E. E., & Hennessy, E. A. (2016). Estimating the difference between published and unpublished effect sizes: A meta-review. *Review of Educational Research, 86*, 207–236. https://doi.org/10.3102/0034654315582067.
Prat-Sala, M., & Redford, P. (2012). Writing essays: Does self- efficacy matter? The relationship between self-efficacy in reading and in writing and undergraduate students' performance in essay writing. *Educational Psychology, 32*(1), 9–20. https://doi.org/10.1080/01443410.2011.621411.
Pulford, D. P., Woodward, B., & Taylor, W. E. (2018). Do social comparisons in academic settings relate to gender and academic self-confidence? *Social Psychology of Education, 21*(3), 677–690. https://doi.org/10.1007/s11218-018-9434-1.
Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods*. Sage.
Robinson, K. A., Perez, T., White-Levatich, A., & Linnenbrink-Garcia, L. (2020). Gender differences and roles of two science self-efficacy beliefs in predicting post-college outcomes. *Journal of Experimental Education*. https://doi.org/10.1080/00220973.2020.1808944. Advance online publication.
Rosenberg, M. (1979). *Conceiving the self*. Basic Books.
Rosenthal, R., & DiMatteo, M. R. (2001). Meta-analysis: Recent developments in quantitative methods for literature reviews. *Annual Review of Psychology, 52*, 59–82. https://doi.org/10.1146/annurev.psych.52.1.59.
Rothstein, H. R., Sutton, A. J., & Borenstein, M. (2005).). *Publication bias in meta-analysis: Prevention, assessment and adjustments*. John Wiley & Sons.
Ruan, Z. (2014). Metacognitive awareness of EFL student writers in a Chinese ELT context. *Language Awareness, 23*(1–2), 76–91. https://doi.org/10.1080/09658416.2013.863901.
Sanders-Reio, J., Alexander, P. A., Reio, T. G., Jr., & Newman, I. (2014). Do students' beliefs about writing relate to their writing self-efficacy, apprehension, and performance. *Leaning and Instruction, 33*, 1–11. https://doi.org/10.1016/j.learninstruc.2014.02.001.
Sarkhoush, H. (2013). Relationship among Iranian EFL learners' self-efficacy in writing, attitude towards writing, writing apprehension and writing performance. *Journal of Language Teaching and Research, 4*(5), 1126–1132. https://doi.org/10.4304/jltr.4.5.1126-1132.
Scholz, U., Doña, B. G., Sud, S., & Schwarzer, R. (2002). Is general self-efficacy a universal construct? Psychometric findings from 25 countries. *European Journal of Psychological Assessment, 18*, 242–251. https://doi.org/10.1027//1015-5759.18.3.242.
Schunk, D. H. (1991). Self-efficacy and academic motivation. *Educational Psychologist, 26*(3–4), 207–231. https://doi.org/10.1080/00461520.1991.9653133.
Shell, D. F., Colvin, C., & Bruning, R. H. (1995). Self-efficacy, attributions, and outcome expectancy mechanisms in reading and writing achievement: Grade-level and achievement-level differences. *Journal of Educational Psychology, 87*, 386–398. https://doi.org/10.1037/0022-0663.87.3.386.
Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly, 27*(4), 657–677. https://doi.org/10.2307/3587400.
Slavin, R., & Smith, D. (2009). The relationship between sample sizes and effect sizes in systematic reviews in education. *Educational Evaluation and Policy Analysis, 31*(4), 500–506. https://doi.org/10.3102/0162373709352369.
Sun, T., & Wang, C. (2020). College students' writing self-efficacy and writing self-regulated learning strategies in learning English as a foreign language. *System, 90*. https://doi.org/10.1016/j.system.2020.102221, 202221.
Teng, L. S., Sun, P. P., & Xu, L. (2018). Conceptualizing writing self-efficacy in English as a foreign language contexts: Scale validation through structural equation modeling. *TESOL Quarterly, 52*(4), 911–942. https://doi.org/10.1002/tesq.432.
Usher, E. L., & Pajares, F. (2008). Sources of self-efficacy in school: Critical review of the literature and future directions. *Review of Educational Research, 78*(4), 751–796. https://doi.org/10.3102/0034654308321456.
Valiante, G. (2001). *Writing self-efficacy and gender orientation: A developmental perspective (Publication No. 3009481) [Doctoral dissertation, Emory University]*. ProQuest Dissertations Publishing.
Villalón, R., Mateos, M., & Cuevas, I. (2015). High school boys' and girls' writing conceptions and writing self-efficacy beliefs: What is their role in writing performance? *Educational Psychology, 35*(6), 653–674. https://doi.org/10.1080/01443410.2013.836157.
White, M. J., & Bruning, R. (2005). Implicit writing beliefs and their relation to writing quality. *Contemporary Educational Psychology, 30*(2), 166–189. https://doi.org/10.1016/j.cedpsych.2004.07.002.
Wilson, K. M., & Trainin, G. (2007). First-grade students' motivation and achievement for reading, writing, and spelling. *Reading Psychology, 28*(3), 257–282. https://doi.org/10.1080/02702710601186464.
Wischgoll, A. (2016). Combined training of one cognitive and one metacognitive strategy improves academic writing skills. *Frontiers in Psychology, 7*, 1–13. https://doi.org/10.3389/fpsyg.2016.00187.
Woodrow, L. (2011). College English writing affect: Self-efficacy and anxiety. *System, 39*(4), 510–522. https://doi.org/10.1016/j.system.2011.10.017.
Wright, K. L., Hodges, T. S., & McTigue, E. M. (2019). A validation program for the self-beliefs, writing-beliefs, and attitude survey: A measure of adolescents' motivation toward writing. *Assessing Writing, 39*, 64–78. https://doi.org/10.1016/j.asw.2018.12.004.
Zabihi, R. (2018). The role of cognitive and affective factors in measure of L2 writing. *Written Communication, 35*(1), 32–57. https://doi.org/10.1177/0741088317735836.
Zhang, Y., & Guo, H. (2012). A study of English writing and domain-specific motivation and self-efficacy of Chinese EFL learners. *Journal of Pan-Pacific Association of Applied Linguistics, 16*(2), 101–121.

Ting Sun is a doctoral student at the Cato College of Education in the University of North Carolina at Charlotte. Her research interests include meta-analysis and English writing self-efficacy beliefs and strategies. She has published over 10 articles in peer-reviewed journals.

Chuang Wang is Distinguished Professor at the Faculty of Education in the University of Macau. His research interests include statistics, English language self-efficacy beliefs, and use of self-regulated learning strategies. He has published over 100 articles in peer-reviewed journals.

Richard G. Lambert is Professor and Director of the Center for Educational Measurement and Evaluation at the Cato College of Education in the University of North Carolina at Charlotte. His research interests include statistics, meta-analysis, teacher stress, and early childhood education. He has published over 100 articles in peer-reviewed journals.

Lan Liu is a master student at the Faculty of Hospitality and Tourism Management, Macau University of Science and Technology. Her research interests include English language learner's self-efficacy beliefs and online education.