RELATIONSHIP BETWEEN WRITING SELF-EFFICACY AND WRITING ACHIEVEMENT: EVIDENCE FROM AN EMPIRICAL STUDY AND A META-ANALYSIS

by

Ting Sun

A dissertation submitted to the faculty of
The University of North Carolina at Charlotte
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in
Educational Research, Measurement, and Evaluation

Charlotte

2021

Approved by:

_____
Dr. Richard G. Lambert

_____
Dr. Chuang Wang

_____
Dr. Xiaoxia Newton

_____
Dr. Lan Quach Kolano

ABSTRACT

TING SUN. Relationship between Writing Self-Efficacy and Writing Achievement: Evidence from an Empirical Study and a Meta-Analysis. (Under the co-direction of Dr. RICHARD G. LAMBERT and Dr. CHUANG WANG)

This three-article format dissertation aims to provide evidence of the effect size of the relationship between writing self-efficacy and writing achievement in an empirical study and a meta-analysis, and compare three methods, the averaging method, three-level meta-analysis (3LM), and robust variance estimation (RVE), in handling multiple dependent effect sizes within studies in doing the meta-analysis. In the first article, data were collected from college students in an English as a foreign language (EFL) context. The second article extracted 575 effect sizes from 76 primary studies, which were also used in the third article. Results revealed a medium to large observed effect size (r = .47) with EFL students and a medium true effect size (r = .29) of the relationship between writing self-efficacy and writing achievement. Writing in English as a first language (L1) or a second language (L2) was found to moderate the relationship between writing self-efficacy and writing achievement, with the effect size estimated with L2 learners (r = .44) being statistically significantly larger than that yielded with L1 learners (r = .23). Results also revealed that statistical procedure moderated effect size estimates. The third article found that while the averaging method, 3LM, and RVE yielded similar estimates of the average effect size and standard error, the averaging method resulted in a much smaller variance of true effect sizes. 3LM produced the smallest variance estimate when the number of primary studies is small. This dissertation provides evidence of the observed and true effect sizes of the relationship between writing self-efficacy and writing achievement. It also informs meta-analysts of the performance of the three methods in handling dependent effect sizes in doing meta-analyses.

ACKNOWLEDGEMENTS

I would like to extend my heartfelt appreciation to my academic advisers and dissertation co-chairs, Dr. Richard G. Lambert, and Dr. Chuang Wang, for their tremendous guidance and continuous encouragement through each stage of the dissertation process. Your insightful feedback inspired me to think critically and elevated my dissertation to a higher level. Without your support, this dissertation would not have been able to complete, and the first two articles would not have been accepted by the refereed journals *System* and *Journal of Second Language Writing*.

My sincere appreciation also goes to the members on the dissertation committee: Dr. Xiaoxia Newton and Dr. Lan Quach Kolano, for their constructive comments and suggestions. I am extremely grateful for the time, patience, and intellectual contributions that they offer to the improvement of my dissertation.

I am also indebted to the faculty members in the ERME program, Dr. Sandra Dika, Dr. Claudia Flowers, Dr. Stella Kim, Dr. Jae Hoon Lim, and Dr. Carl Westine, for their instruction and supportive efforts, which aid me in the transformation from a novice researcher to an established scholar. My thanks also go to my friends in the ERME program, Kristin, Liz, Tong, Tuba, and Yi, for their cordial friendship and endless assistance, which have largely facilitated my research processes.

Last but not least, I would like to express my affectionate thanks to my family, who has extended much consideration to me over the years. Thanks to my father, Yanfeng Sun, and my mother, Yuqin Zhang, for loving me unconditionally. Thanks to my sister, Ying Sun, for taking care of my parents while I was studying for my Ph.D. in the United States. It is their unceasing caring and love that has motivated me to complete my dissertation.

DEDICATION

This dissertation is dedicated to my husband, Shengqiang Chen, and my daughter, Yixuan Chen, who have been a constant source of support, encouragement, and happiness.

TABLE OF CONTENTS

LIST OF TABLES ix
LIST OF FIGURES x
LIST OF ABBREVIATIONS xi
INTRODUCTION 1
Statement of the Problem 3
Research Purposes and Research Questions 6
Overview of Study: The Three Articles 6
The First Article 6
The Second Article 7
The Third Article 8
Significance of the Study 9
Limitations 10
CHAPTER 1 [ARTICLE I]: COLLEGE STUDENTS' WRITING SELF-EFFICACY AND WRITING SELF-REGULATED LEARNING STRATEGIES IN LEARNING ENGLISH AS A FOREIGN LANGUAGE 11
Theoretical Framework 13
Literature Review 15
Writing Self-Efficacy 15
Writing Self-Regulated Learning Strategies 16
Relationships Between Writing Self-Efficacy and Writing Self-Regulated Strategies 18
Measurements of Writing Self-Efficacy and Writing Self-Regulation 19
Method 21
Participants 21
Instruments 23
Procedure 29
Data Analytical Procedure 29
Results 30
Status Quo of Self-Efficacy and Self-Regulation 30
Relationships 32
Unique Contribution of Self-Efficacy and Self-Regulation to Writing 35
Discussion 39
EFL Students' Writing Self-Efficacy and SRL Strategy Use 39
Correlation between Self-Efficacy, SRL Strategies, and Writing 42

Predictive Effects of Writing Self-Efficacy and Writing SRL Strategies 43
Limitations and Future Directions 44
Significance and Implications 45
References 48
APPENDIX A: QUESTIONNAIRE OF ENGLISH WRITING SELF-EFFICACY (QEWSE) 57
APPENDIX B: QUESTIONNAIRE OF ENGLISH WRITING SELF-REGULATED LEARNING STRATEGIES (QEWSRLS) 58
CHAPTER 2 [ARTICLE II]: RELATIONSHIP BETWEEN WRITING SELF-EFFICACY AND SECOND LANGUAGE WRITING ACHIEVEMENT IN ENGLISH: A META-REGRESSION ANALYSIS 59
Literature Review 61
Social Cognitive Theory and Writing Self-Efficacy 61
Relationship between Self-Efficacy and English Writing Achievement for L1 and L2 Writers 62
Moderating Effect of L1/L2 Writing on the Effect Sizes 64
Sample Size 65
Gender 65
Grade Level 66
Statistical Procedures 66
Publication Type 67
Methodology 68
Literature Search 68
Literature Screening 69
Data Coding 70
Publication Bias 74
Missing Values 74
Data Analysis 75
Results 77
Discussion 82
Limitations and Future Research 86
Conclusion and Implications 86
References 89
APPENDIX A: CODEBOOKS 108
APPENDIX B: PREDICTORS OF EFFECT SIZE MAGNITUDE AFTER REMOVING STUDIES WITH BOTH L1 AND L2 WRITERS 111
CHAPTER 3 [ARTICLE III]: COMPARING THREE APPROACHES TO HANDLING DEPENDENCE: A CASE STUDY OF A META-ANALYSIS OF WRITING SELF-EFFICACY AND WRITING PROFICIENCY 112

The Three Methods 115
Averaging Within-Study Effect Sizes 115
Three-Level Meta-Analysis 117
Robust Variance Estimation 118
Literature in Comparing Different Methods 119
Methodology 121
Data 121
Effect Size Calculation 121
Subsets of Data Generation 122
Data Analysis 124
Results 125
Condition of All Effect Sizes 125
Condition of 61-2 126
Condition of 42-2 126
Condition of 26-2 127
Condition of 42-4 127
Condition of 26-4 127
Condition of 26-6 128
Discussion 136
Similarities between the Three Methods 136
Differences between the Three Methods 137
Overall Average Effect Size 137
Standard Error 137
Variance of True Effect Size Estimates 138
Limitations and Future Research 139
Implications and Conclusion 139
References 142
OVERALL CONCLUSION 147
References 150

LIST OF TABLES

Table 1-1 Descriptive Statistics of Participants' Demographic Information by University 22

Table 1-2 Model Fit Indices for the CFA of QEWSE and QEWSRLS 26

Table 1-3 Descriptive Statistics of Writing Self-Efficacy and Writing Self-Regulated Learning Strategies 31

Table 1-4 Pearson Correlation Matrix for Writing Proficiency and Writing Self-Efficacy 33

Table 1-5 Correlation Matrix for Writing Proficiency and Writing SRL Strategies 34

Table 1-6 Means and Standard Deviations of Writing Self-Efficacy, Writing SRL Strategies and Writing Proficiency by Gender and SES 36

Table 1-7 Hierarchical Multiple Linear Regression Analysis Summary for Gender, Writing SRL Strategies, and Writing Self-Efficacy Predicting Writing Scores (N = 288) 37

Table 2-1 Inclusion and Exclusion Criteria 71

Table 2-2 Frequency Distribution of Studies by Publication Type, Research Setting, L1/L2 Writing, and Statistical Procedures 78

Table 2-3 Predictors of Effect Size Magnitude 80

Table 2-4 Model-Based Effect Size Estimates 81

Table 3-1 Frequency of the Number of Effect Sizes per Study 123

Table 3-2 Six Conditions by Number of Primary Studies and Number of Effect Sizes per Study 124

Table 3-3 Effect Size and Heterogeneity Statistics for the Condition of All the Effect Sizes 129

Table 3-4 Effect Size and Heterogeneity Statistics for the Condition of 61-2 130

Table 3-5 Effect Size and Heterogeneity Statistics for the Condition of 42-2 131

Table 3-6 Effect Size and Heterogeneity Statistics for the Condition of 26-2 132

Table 3-7 Effect Size and Heterogeneity Statistics for the Condition of 42-4 133

Table 3-8 Effect Size and Heterogeneity Statistics for the Condition of 26-4 134

Table 3-9 Effect Size and Heterogeneity Statistics for the Condition of 26-6 135

LIST OF FIGURES

Figure 1-1 Structure of QEWSE 27

Figure 1-2 Structure of QEWSRLS 28

Figure 2-1 PRISMA Flow Diagram of Literature Search and Screening 72

LIST OF ABBREVIATIONS

EFL	English as a Foreign Language
ESL	English as a second language
L1	first language
L2	second language
CET-4	College English Test band 4
CET-6	College English Test band 6
TOFEL	Test of English as a Foreign Language
IELTS	International English Language Testing System
NAEP	National Assessment of Educational Progress
ES	effect size
UNM	univariate method
AVM	averaging method
RVE	robust variance estimation
2LM	two-level meta-analysis
3LM	three-level meta-analysis
MLM	multilevel meta-analysis
SRL	self-regulated learning
QEWSE	Questionnaire of English Writing Self-Efficacy
QEWSRLS	Questionnaire of English Writing Self-Regulated Learning Strategies
SES	socioeconomic status

SRSD	Self-Regulated Strategy Development
CFA	confirmatory factor analysis
MANOVA	multivariate analysis of variance
GFI	Goodness of Fit Index
AGFI	Adjusted Goodness of Fit Index
NFI	Normed Fit Index
NNFI	Non-Normed Fit Index
CFI	Comparative Fit Index
SRMR	Standardized Root Mean Residual
RMSEA	Root Mean Square Error of Approximation

INTRODUCTION

As a major component of literacy proficiency, English writing is predictive of students' academic success (Graham, 2006), and also serves as a criterion for college admission, employee recruitment and career promotion (National Commission on Writing, 2004). Writing is not only critical for individuals' academic and professional opportunities but also essential to businesses. For example, two-thirds of job positions require writing skills, and the annual cost of writing remediation and writing training for American firms may reach up to $3.1 billion (National Commission on Writing, 2004). English writing is also important in countries where English is taught as a foreign language (EFL) because it is a key indicator of EFL learners' overall linguistic competence (Archibald, 2016). In China, for example, the three most important national-level standardized tests, the College Entrance Examination, the College English Tests band 4 (CET-4), and College English Test band 6 (CET-6) were all designed with an English writing section as an essential part. In addition, EFLs who apply to universities in English-speaking countries are required to submit their scores on the Test of English as a Foreign Language (TOFEL) or the International English Language Testing System (IELTS) and they need to reach a certain level in the writing section.

Writing is also a difficult task and a complex process. National Center for Education Statistics (2012, p. 4) defined writing as "a complex, multifaceted, and purposeful act of communication that is accomplished in a variety of environments, under various constraints of time, and with a variety of language resources and technological tools". According to the 2011 National Assessment of Educational Progress (NAEP) writing assessment, 20% students in grade 8 and 21% students in grade 12 performed below the basic level (National Center for Education

Statistics, 2012). Writing is not only difficult for native English speakers but also challenging and demanding for the EFL learners (Zhang & Guo, 2012).

The difficulty of writing can be accounted for by the following reasons. First, the writing process is provided with less scaffolding in terms of contextual information compared with other domains of language learning (Bruning & Horn, 2000) because of its displacement from unseen audiences (Bruning et al., 2013; Bruning & Horn, 2000). Second, writing is a dynamic and multidimensional process that involves not only linguistic proficiency, but also involves cognitive, motivational, and affective factors on the part of writers (Anastasiou & Michail, 2013; Bruning et al., 2013). Therefore, the acquisition of vocabulary and grammar alone is deemed insufficient to make a proficient writer, and other attributes such as self-efficacy and self-regulation are also needed (Schunk & Zimmerman, 2007).

Bandura defined self-efficacy as "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performance" (1986, p.391). According to social cognitive theory, people's beliefs play a critical role in human functioning because they can affect their behavior, emotions, and outcomes of events in which they are engaging in (Bandura, 1986, 1997). Meta-analysis studies in self-efficacy confirmed the facilitative and mediational role self-efficacy in various disciplines (e.g., Judge & Bono, 2001; Sadri & Robertson, 1993; Stajkovic & Luthans, 1998; Unrau et al., 2018). Self-efficacy was found to have a significantly positive relationship with work-related performance, behaviors (Sadri & Robertson, 1993; Stajkovic & Luthans, 1998) and job satisfaction (e.g., Judge & Bono, 2001). Self-efficacy was also found to positively influence students' academic outcomes in general (Multon et al., 1991) and their language proficiency in particular (Unrau et al., 2018). Multon et al. (1991) conducted a meta-analysis of 38 studies in English reading and revealed that

academic outcomes were statistically significantly related to self-efficacy beliefs (r = .38) across various educational settings and disciplines.

In the domain of writing, self-efficacy was operationalized 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). A body of research evidence showed that self-efficacy beliefs were related to various desirable outcomes in the process of writing (e.g., Pajares & Valiante, 2001; Woodrow, 2011; Zhang & Guo, 2012). Writing self-efficacy was statistically significantly related to some motivation constructs such as self-concept, achievement goals, value of writing, and self-efficacy for self-regulation (Pajares et al., 2000; Pajares & Valiante, 2001). Efficacious students tend to be more highly motivated (Zhang & Guo, 2012), adopt more effective writing strategies (Ekholm et al., 2015), experience less negative emotions (Woodrow, 2011), and persist longer in face of writing difficulties and frustrations (Pajares, 2002). Writing self-efficacy was not only positively correlated with writing outcomes (Bruning et al., 2013; Pajares, 2007) but also made an independent contribution to the prediction of writing achievement (Graham et al., 2019; Prat-Sala & Redford, 2012).

Statement of the Problem

Given the importance of self-efficacy beliefs to writing achievement, it is significant to investigate the relationship between the two variables. However, a large body of research has been conducted in the English as a first language context (e.g., Bruning et al., 2013; Graham et al., 2017; Jones, 2008; Pajares et al., 2007; Prat-Sala & Redford, 2012; Sanders-Reio et al., 2014; Wright et al., 2019), and our knowledge of the relationship between self-efficacy and proficiency in the EFL context is limited. Moreover, among the few studies conducted in the EFL context, researchers adopted writing self-efficacy measures developed in the English as a first language

context (e.g., Zabih, 2018; Zhang & Guo, 2012). For example, Chae (2013) adopted the Writing Self-Efficacy Scale developed by Pajares and Valiante (1999) in the United States to assess Korean students' writing self-efficacy beliefs. According to Bandura (1997), self-efficacy is not a global construct, and should be domain- and context-specific. Therefore, it is time to conceptualize a writing self-efficacy measure that is aligned with the EFL students and use this measure to examine the relationship of writing self-efficacy and writing achievement in the EFL context.

In order to understand the relationship between writing self-efficacy and writing achievement, results from one study is not conclusive since it may not yield an accurate estimate of the true effect because of sampling errors. A meta-analysis offers a viable alternative to the better estimation of effect sizes by increasing statistical power and having more representative participants (Borenstein et al., 2009). A meta-analysis can estimate not only the overall average effect (e.g., the magnitude of the relationship between writing self-efficacy and writing proficiency), but also the variation of true effect sizes across studies. So far, meta-analyses on self-efficacy have been conducted from various perspectives such as self-efficacy and academic outcomes (e.g., Multon et al., 1991), self-efficacy and work-related performance (e.g., Judge & Bono, 2001; Sadri & Robertson, 1993; Stajkovic & Luthans, 1998), children and adolescents' self-efficacy (e.g., Holden et al., 1990), gender differences in self-efficacy (e.g., Huang, 2013), and effects of interventions on reading self-efficacy (e.g., Unrau et al., 2018). However, no meta-analysis has been conducted on the relationship between self-efficacy and achievement in the writing domain.

To conduct a meta-analysis on the relationship between self-efficacy and students' writing achievement, effect sizes reported in primary studies are to be synthesized in a systematic and

scientific manner. However, a number of primary studies on this topic reported more than one effect size calculated from multiple measures of writing self-efficacy or/and writing achievement (e.g., Bruning et al., 2013). For example, Neugebauer and Howard (2015) reported multiple effect sizes from five measures of writing self-efficacy (i.e., general progress, specific progress, observational comparisons, social feedback, and physiological state). These effect sizes are dependent since they are working with the same sample, using the same research design, and conducted in the same context. Univariate meta-analysis assumes effect sizes to be synthesized are independent of each other, so it is not an appropriate procedure when the dependence issue arises (Moeyaert et al., 2017). Several procedures can be used to deal with the dependence issue, including averaging effect sizes or weighted effect sizes within each study, using multivariate meta-analysis (Kalaian & Raudenbush, 1996), robust variance estimation (RVE; Hedges et al., 2010), and three-level meta-analysis (3LM; Cheung, 2014; Van den Noortgate et al., 2013). Moeyaert et al. (2017) conducted a study with simulated data in comparing the averaging effect sizes, RVE, and multilevel meta-analysis (MLM) in dealing with the dependence arising from multiple outcomes and found both RVE and MLM yielded unbiased estimates of effect sizes. So far, there is inadequate evidence about the performance of the three methods in handling the dependence issue with real data. Although simulated data has the advantage of generating evaluation criteria against which different methods can be compared, a study with real data is also needed because real data "can better emulate the types and nature of dependence that typically exist in studies that education researchers struggle to meta-analyze" (Scammacca et al., 2014, p.336). In addition, previous methodological studies used the effect sizes of standardized mean differences (i.e., Cohen's *d* or Hedges' *g*) extracted from experimental or quasi-experimental studies and their results may not be applied to effect sizes based on correlation.

Therefore, a methodological study is warranted to examine methods in handling dependent effect sizes reported in the primary studies that investigate the relationship between writing self-efficacy and writing achievement. This study provides evidence for choosing appropriate methods to estimate the effect size of their relationship. An accurate estimate of the relationship between writing self-efficacy and writing performance provide applied researchers, practitioners, and program designers with evidence-based guidance for doing research, teaching or program implementation in English writing.

Research Purposes and Research Questions

The primary purpose of this dissertation is to obtain evidence of the relationship between writing self-efficacy and writing proficiency in an empirical study and a meta-analysis, and to compare methods in handling multiple dependent effect sizes within studies in doing the meta-analysis. The three research questions that guide this dissertation are as follows:

1. What is the relationship between writing self-efficacy and writing performance in the EFL context?
2. How does this relationship compare with that as evidenced in a meta-analysis?
3. How do the methods (i.e., the averaging method, RVE, and 3LM) differ in their performance of handling multiple dependent effect sizes in doing the meta-analysis?

Overview of Study: The Three Articles

The First Article

This first article aims to examine how writing self-efficacy and writing self-regulated learning (SRL) strategies are related to writing proficiency among college students in an EFL context. The Questionnaire of English Writing Self-Efficacy (QEWSE) and the Questionnaire of English Writing Self-Regulated Learning Strategies (QEWSRLS) were administered to Chinese

college students. Their writing proficiency was measured by their writing scores on the writing section of the College English Test Band 4 (CET-4). The study revealed that EFL students reported a medium level of self-efficacy and infrequent use of SRL strategies in writing. Students' writing scores were statistically significantly and moderately correlated with both the composite value and the five individual constructs of writing self-efficacy (i.e., Ideation, Organization, Grammar and Spelling, Use of English Writing, and Self-Efficacy for Self-Regulation; *r* ranging from .29 to .48). Writing scores were also significantly correlated with the use of SRL strategies (*r* ranging from .10 to .38). Moreover, regression results revealed both writing self-efficacy and writing SRL strategies explained 23% of students' writing proficiency. Specifically, Review of Records Strategies, Seeking Opportunities Strategies, and Self-Evaluation Strategies significantly predicted writing scores. In addition, two subcategories of self-efficacy, Grammar and the Use of English writing significantly explained the variance in writing scores. Since self-efficacy and self-regulation are crucial to students' writing proficiency, the first article provides classroom teachers with insights about how to incorporate instructional strategies into EFL classrooms to improve students' writing outcomes.

The Second Article

The second article aims (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 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 = .44) being statistically significantly larger that yielded with L1 learners (r = .23), 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.

The Third Article

The third article compares three approaches (i.e., averaging effect sizes, 3LM, and RVE) to handling the dependence issue in conducting meta-analyses. A univariate meta-analysis and two-level meta-analysis (2LM) were also performed as the baseline methods against which the three methods were compared. Data used in this article were from a meta-analytic study (the second article) examining the relationship between writing self-efficacy and writing achievement. Seven conditions were created to examine the differences of the performance of the three methods by the number of primary studies and the number of effect sizes per study. The study found that while the averaging methods, 3LM, and RVE produced similar results in the average effect size and standard error estimates, the averaging method had much smaller variances. In addition, 3LM produced the smallest value of variance when the number of studies

was small. This study informs meta-analysts of appropriate procedures in handling dependent effect sizes.

Significance of the Study

The study contributes to the literature by providing evidence of the relationship between writing self-efficacy and writing achievement from both an empirical study and a meta-analysis. The empirical study provides observed effect sizes in an EFL context, while the meta-analysis estimates the true effect size across different contexts. The overall average effect size estimated from the meta-analysis provides an opportunity to compare a particular observed effect size with the estimated true effect. The discrepancy between the observed effects and the true effect may shed some light on factors that moderate the relationship between writing self-efficacy and writing achievement. Methodologically, this study provides meta-analysts with insights into understanding potential procedures in handling multiple dependent effect sizes. Choosing appropriate methods would aid in the accurate estimation of effect sizes. An accurate estimate of the relationship between writing self-efficacy and writing performance provides applied researchers, practitioners, and administrators with evidence-based guidance for doing research, teaching or policymaking in English writing.

This study is also significant in that it adapted a self-efficacy measure in the EFL context based on the social cognitive theory and cognitive process theory of writing. Multiple instruments were developed and operationalized to measure writing self-efficacy beliefs in the English as a first language context (e.g., *the Self-Efficacy for Writing Scale* in Bruning et al., 2013; *the Writing Self-Efficacy Scale* in Pajares and Valiante, 1999; Prat-Sala & Redford, 2012; *the Writing Self-Efficacy Index* in Sanders-Reio et al., 2014; *the Writing Self-Efficacy Instrument* Shell et al., 1989) but few were constructed in the EFL context (Teng et al., 2018).

Conceptualizing a writing self-efficacy measure that is aligned with linguistic and psychological theories would help corroborate theoretical foundations of self-efficacy and provides insights into understanding this construct form linguistic, cognitive, and psychological perspectives.

This study also has pedagogical implications. It sheds some light on the understanding of the relationships between writing self-efficacy and students' writing outcomes. Since the first study noted a positive correlation between writing self-efficacy and students' writing achievement, educators and schools should take the dual responsibility of raising students' proficiency and their confidence in the domain of writing (Pajares, 2007). Assessment of writing should include both the assessment of students' writing competence and students' self-efficacy beliefs (Pajares, 2003).

Limitations

The limitations of the study are as follows. First, convenience sampling was employed in the first article, which reduces the generalizability of the results. Second, the two universities from which data were collected are not balanced in student population, gender, and major composition. Further studies are recommended to replicate the results with students from more diverse universities to have a better representation of EFL learners. The limitation of the second article arises from the inclusion of only journal articles and dissertations/theses in the meta-analysis. Other grey literature such as conference abstracts/presentations, books and book chapters, unpublished technical reports, or white papers are suggested to be included for further studies for purposes of obtaining more presentative data and further minimizing publication bias. Lastly, the third article only used real data, which limits the opportunity of having the criterion against which to compare the methods of interest in terms of estimation accuracy. Future research is suggested to conduct a simulation study to confirm the results of this study.

CHAPTER 1 [ARTICLE I]: COLLEGE STUDENTS' WRITING SELF-EFFICACY AND WRITING SELF-REGULATED LEARNING STRATEGIES IN LEARNING ENGLISH AS A FOREIGN LANGUAGE

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*, 102221. https://doi.org/10.1016/j.system.2020.102221

Writing embodies language learners' overall linguistic competence and is a challenge for learners in any language (Anastasiou & Michail, 2013). For learners of English as a foreign language (EFL), writing is often viewed as the most difficult among all language skills (Zhang & Guo, 2012). This is because less scaffolding is provided in the writing process compared with other domains of language learning (Bruning & Horn, 2000). The displacement from its readers makes writing a thought-demanding and effortful task (Bruning et al., 2013; Bruning & Horn, 2000). In addition, writing is a complex and multidimensional process that involves many factors such as task environment, motivation, working memory, long-term memory and cognitive process (Hayes, 2000).

Given that writing is a cognitive process (Bruning et al., 2013; Hayes, 2000), it is unsurprising to find positive relationships between writing proficiency, self-efficacy, and self-regulation (Schunk & Zimmerman, 2007). Writing self-efficacy refers to individuals' judgments of how well they can accomplish a writing task based on their assessments of "various composition, grammar, usage, and mechanical skills" (Pajares & Valiante, 2001, p. 369). Efficacious writers usually have higher motivation (Zhang & Guo, 2012), exert more effort, experience less apprehension (Pajares, 2003; Pajare et al., 2000; Woodrow, 2011), and demonstrate greater perseverance in the writing process (Pajares, 2003; Prat-Sala & Redford, 2012). Writing self-regulated learning (SRL) strategies are defined as self-initiated behaviors that

writers utilize to achieve multiple literary goals, such as enhancing their writing products and writing skills (Zimmerman & Risemberg, 1997). Self-regulation was correlated with the improvement of writing competency since writing is "self-planned, self-initiated and self-sustained" (Zimmerman & Risemberg, 1997, p. 73). Although self-efficacy and self-regulation have been widely documented, there is inadequate evidence about how self-efficacy and self-regulation contribute to students' writing proficiency among college students in an EFL writing context (Zhang & Guo, 2012).

Studies also provided evidence that the relationships between writing self-efficacy, writing SRL, and writing outcomes were moderated by students' demographic characteristics such as gender (McGettigan, 2008; Pajares, 2007; Pajares & Valiante, 2001) and socioeconomic status (SES; Korat & Schiff, 2005). Writing was considered a stereotypically feminine activity (Pajares, 2002), and studies noted gender differences in writing self-efficacy and writing outcomes favoring females (Andrade et al., 2009; Pajares & Valiante, 2001). SES refers to social standing and economic characteristics of students (Ensminger & Fothergill, 2003). Students' SES was linked to their social and family capital and learning experiences and thus had an impact on their literacy development (Korat & Schiff, 2005). Previous research also suggested that the relationship between writing self-efficacy and writing proficiency varies as a function of SES (Korat & Schiff, 2005). Korat and Schiff (2005) found that improvement was shown in both writing knowledge and writing self-efficacy for students in low SES schools as they grow older while the opposite development trend was displayed in the two in students in high SES schools. However, most previous studies did not control gender and SES when examining the relationship between writing proficiency and self-efficacy and writing strategies (e.g., Chien, 2012; Prat-sala & Redford, 2012). The investigation of the relationships between writing self-efficacy, writing

SRL strategies and writing proficiency in an EFL context with both gender and SES controlled would provide teachers with insights about how to enhance college students' English writing proficiency.

The purposes of the study were threefold: (a) to describe the status quo of EFL learners' writing self-efficacy and writing SRL strategies; (b) to examine the relationships between writing self-efficacy, writing SRL strategies, and writing proficiency of EFL learners; and (c) to examine how much variation in EFL learners' writing proficiency is explained by their writing self-efficacy and writing SRL strategies after controlling for gender and SES.

Theoretical Framework

According to the cognitive process theory of writing (Flower & Hayes, 1981), writing is a hierarchically structured and process-oriented activity that involves the writer's task environment, the writer's long-term memory, and the writing process, which are interactive and dynamic in nature. Twenty years after developing the aforementioned theory, Hayes (2000) proposed a new and revised model of this theory, which is comprised of two components: the task environment and the individual. The former incorporates the social environment (e.g., comments from teachers or social distractions) and the physical environment (e.g., task materials or online writing sources) which may positively or negatively impact the writing process. The latter consists of motivation/affect, working memory and cognitive processes. The task environment would affect individuals' affective and cognitive attributes in writing, which in turn, influence their ongoing writing tasks (i.e., physical environment). The two intertwined and interactive components constitute the whole writing processes. Based on the cognitive process theory, writing is a communicative and generative activity involving "cognitive, affective, social,

and physical conditions" (Hayes, 2000, p. 5). Therefore, cognitive and affective factors such as self-efficacy and self-regulation are vital for writing outcomes.

Social cognitive theory holds that the beliefs people have predict their behaviors and influence the outcomes (Bandura, 1997) so self-efficacy was believed to be a facilitative and mediating mechanism of human agency (Bandura, 1986). In the context of English writing, students with higher levels of self-efficacy are more likely to put more effort, exert more persistence, and thus have higher writing outcomes (Usher & Pajares, 2008). The social cognitive theory posits that self-regulation encompasses the reciprocity of environmental, behavioral and personal processes (Bandura, 1986), based on which Zimmerman and Risemberg (1997) conceptualized a triadic system for writing self-regulation. Environmental processes are "writers' self-regulation of the physical or social setting" (p. 76). Self-regulation of the physical setting includes environmental structuring (e.g., in a library or a cafeteria), and self-regulation of the social setting includes self-selected tools and sources (e.g., use of a dictionary and seeking social assistance). Behavioral processes refer to the self-regulation of writing activities, including self-monitoring, self-consequences, and self-verbalization. Personal (covert) processes refer to the self-regulation of affective and/or cognitive factors (e.g., time planning, goal setting, and self-evaluation). Studies also suggested that gender and SES influenced writing self-regulation (Albard & Lipschultz, 1998; Callan et al., 2017), which provided evidence that the two factors interact with the writing process.

There are quite a lot in common between cognitive process theory and social cognitive theory: (a) both focus on the interplay between the environment and the individual; (b) both emphasize the cognitive process. The cognitive process theory of writing decomposes writing into psychological and cognitive components. Social cognitive theory focuses on the interaction

of personal, environmental, and behavioral factors. The two theories shift from an emphasis on product-oriented writing to process-oriented writing, from an emphasis on writers' linguistic proficiency to their cognitive and affective competence. The exploration of EFL writing within these two theoretical frameworks is an interdisciplinary collaboration between cognitive science, psychology and writing (Chien, 2012), which will contribute to the literature in the second language and EFL writing field.

Literature Review

Writing Self-Efficacy

A large body of research has investigated the relationship between writing self-efficacy and writing proficiency (e.g., Bruning et al., 2013; Prat-Sala & Redford, 2012; Wang et al., 2012). Studies also found the mediational effect of writing self-efficacy on the association between writing proficiency and affect constructs (i.e., anxiety or motivation; Woodrow, 2011; Zhang & Guo, 2012). Self-efficacious student writers were highly motivated (Zhang & Guo, 2012), experienced less anxiety, exerted more effort and ultimately did well in writing (Woodrow, 2011). Writing self-efficacy was also positively associated with other motivational attributes such as value of writing, self-concept, self-efficacy for self-regulation, task-approach goals, and performance-approach goals, but negatively correlated with performance-avoidance goals (Pajares, 2003; Pajares et al., 2000; Pajares & Valiante, 2001).

Attempts were also made to investigate gender differences in writing self-efficacy (e.g., Anastasiou, & Michail, 2013; Pajares, 2003). Pajares and Valiante (2001) noted that females had higher writing self-efficacy scores than males, but these gender differences were nullified if students' feminine orientation beliefs (i.e., stereotypic beliefs about femininity) were controlled.

Same results were reported in another study (Andrade et al., 2009) but the differences favoring girls were rendered nonsignificant during the “in writing” and “after writing” periods.

SES was defined as the social standing and economic characteristics of individuals (Ensminger & Fothergill, 2003) and can be measured by parental income, education, occupation, or home resources (Sirin, 2005). Sirin (2005) did a meta-analysis of 201 studies and found a moderate to strong relationship between SES and students’ academic achievement. The researcher further noted that their relationships were moderated by methodological characteristics and student characteristics. Previous research provided evidence that family’s SES is a factor influencing students’ academic performance in general (Ren & Xin, 2013; Sirin, 2005) and language abilities in particular (Kim et al., 2015; Wen et al., 2016). In the domain of writing, students’ self-efficacy levels varied as a function of SES. Students with higher SES reported significantly higher levels of self-efficacy beliefs than their counterparts with lower SES (Korat & Schiff, 2005). SES was also positively correlated with students’ writing knowledge (Korat & Schiff, 2005). Given the empirical evidence about the relationship between SES and writing self-efficacy, it is reasonable to include this covariate in our analysis when examining the relationship between self-efficacy and writing proficiency.

Writing Self-Regulated Learning Strategies

A large body of research identified various SRL strategies that writers utilize in the course of writing (Graham & Harris, 2000; Zimmerman & Risemberg, 1997). Graham and Harris (2000) identified 16 SRL strategies in writing, including goal setting and planning, seeking information, record keeping, organizing, transforming, self-monitoring, reviewing records, self-evaluating, revising, self-verbalizing, rehearsing, environmental structuring, time planning, self-consequence, seeking social assistance, and self-selecting models. Ching (2002) found that the

SRL strategies that Malaysian students used frequently in the course of English writing were organizing and transforming, self-evaluation, seeking social assistance, and seeking information. The least frequently used SRL strategies noted by Ching included environmental structuring, rehearsing and memorizing, and reviewing records.

Previous research suggested positive relationships between writing SRL strategy use and students' writing proficiency (Asmari, 2013, Chen, 2011; Ching, 2002; Teng & Huang, 2019). With 601 graduate and undergraduate students in the United States, Cho et al. (2010) examined the role of self-monitoring in writing and found that students who showed gains in self-monitoring skills displayed improvement in writing quality through an instruction in a system called scaffolded writing and revision in the discipline. Ching (2002) noted that the employment of the SRL strategies contributed to the improvement of students' knowledge associated with English writing. In addition, high achievers in writing tended to use certain categories of SRL strategies. For example, Chien (2012) revealed that high-achieving students spent more time in making position statements in the planning process, generating longer text in the writing process, and making more revision in the review process than low-achieving students. This conclusion was echoed by another study with EFL Spanish speakers (Roca de Larios et al., 2008).

Research provided evidence that there were differences in SRL strategies as a function of gender (Albard & Lipschultz, 1998; Bidjerano, 2005; Pajares, 2002). Albard and Lipschultz (1998) found that girls reported a significantly higher number of SRL strategies than boys even after controlling for their academic achievement. Specifically, female students displayed more frequent use of organizing and transforming, goal setting and planning, keeping records and monitoring, seeking assistance from peers, and reviewing strategies (Albard & Lipschultz, 1998). Researchers claimed that these differences in SRL strategy use in favor of females may be

attributed to the stereotypical beliefs that girls are more organized and have good self-discipline and time management skills (Bidjerano, 2005).

Studies suggested a positive and consistent relationship between SES and strategy use (Jensen, 2009). Callan et al. (2017) examined the Programme for International Student Assessment (PISA) data from 65 countries and found that students from high SES families had more frequent use of academic strategies such as understanding, summarizing learning, and control strategies. The researchers further noted that students from higher SES schools employed strategies related to achievement more frequently. A major problem with learner strategy research is the inconsistency of the definition of strategies (Rose et al., 2018). Rose et al. urged researchers to move away from language learner strategy research and turn to self-regulation research. Although our knowledge about the relationship between SES and students' use of SRL strategies in the writing domain is limited, we expect its impact on the SRL strategy use in writing and include this covariate when examining the relationship between writing SRL and writing proficiency.

Relationships Between Writing Self-Efficacy and Writing Self-Regulated Strategies

Bandura (1986, 1997) stated that self-efficacy beliefs affect the SRL strategies that students utilize in learning. In the domain of writing, studies also revealed that writing self-efficacy had positive correlation with SRL strategy use (Lavelle & Guarino, 2003; Schunk & Zimmerman, 2007; Zimmerman & Risemberg, 1997). Steward et al. (2015) found that an enhancement of undergraduate students' writing self-efficacy was related to an improvement in students' perceptions of the utilization of metacognitive writing strategies. Working with 115 undergraduate students, Ekholm et al. (2015) contended that writing self-efficacy was positively and significantly correlated with self-regulation aptitude (i.e., perceived self-regulative

behaviors). In addition, Ching (2002) argued that instruction in self-regulation increases ESL students' self-efficacy beliefs. However, not all researchers came to the same conclusion. For example, after receiving Self-Regulated Strategy Development (SRSD) instruction, struggling students did not yield an expected improvement in their self-efficacy beliefs, which suggested no relationship between self-regulation and writing self-efficacy (Graham et al., 2005). Graham et al. (2005) further indicated that the nonsignificant relationships between writing self-efficacy and writing SRL strategies might be on account of the inaccuracies in the assessment of their capabilities by students at lower grades.

Measurements of Writing Self-Efficacy and Writing Self-Regulation

Shell et al. (1989) made an initial effort to develop a writing self-efficacy instrument, which was to examine U.S. undergraduates' self-efficacy in a university. The instrument consists of two subscales, tasks and skills. The task subscale measures how confident students are in performing writing tasks, and the skill subscale measures how confident they are in using writing skills. A number of researchers established the psychometric properties of the instrument (e.g., Bruning et al., 2013; Shell et al., 1995). However, this instrument was conceptualized in the first language context and does not have optimal utilization in the EFL context. There are other measures of writing self-efficacy in the first language context (i.e., *Self-Efficacy in Writing* in Prat-Sala & Redford, 2012; *Writing Self-Efficacy Index* in Sanders-Reio et al., 2014) or for K-12 students (i.e., *Self-efficacy for Writing Scale* in Bruning et al., 2013). Few measures of writing self-efficacy were developed for college students in second language context (Teng et al., 2018). Since self-efficacy is domain- and task-specific (Bandura, 1997), there is a need to develop an instrument to measure self-efficacy in the foreign language context. Although satisfactory psychometric properties of the Second Language Writer Self-Efficacy Scale were reported (Teng

et al., 2018), this instrument failed to address task specificity. For example, an item "I can do an excellent job on the assignments in writing courses" does not specify which writing assignment as we know the difficulty levels of writing assignments vary.

Tragant et al. (2013) developed a 17-item instrument to measure language learning strategies and came up with two factors: (a) skills-based deep processing strategies; and (b) language study strategies. Skills-based deep processing strategies included higher-order skills and the application in reading, listening, and writing contexts. Language study strategies represented learning vocabulary and grammar as isolated activities. Existing self-regulation measures are not specific in the writing domain (e.g., *the Strategy Inventory for Language Learning* in Oxford, 1990; *the Questionnaire of English Self-regulated Learning Strategies* in Wang & Bai, 2017). Therefore, the development and validation of instruments for measuring college students' writing self-efficacy and writing self-regulation in the EFL context are warranted. Recently, Teng and Zhang (2016) developed an instrument to measure self-regulated learning strategies in EFL writing using the self-regulation theory framework, but the factors (i.e., cognitive, metacognition, social behavior, and motivational regulation) do not align well with the personal, behavioral, and environmental processes of social cognitive writing model (Zimmerman & Risemberg, 1997). As a result, a new instrument was developed in this study that aligns with Zimmerman and Risemberg's (1997) model.

Although research has suggested that writing self-efficacy and writing SRL strategies are precursors to students' writing proficiency (i.e., Asmari, 2013; Bruning et al., 2013), the relationships between self-efficacy, self-regulation, and writing outcomes among college students in an EFL context have received modest attention. Furthermore, few researchers have investigated the unique contributions of self-efficacy and self-regulation to students' writing

outcomes (Zhang & Guo, 2012). In addition, most studies did not examine the predictive effects of writing self-efficacy and writing self-regulation on EFL learners' writing proficiency when considering the influences of gender or SES (e.g., Chien, 2012; Zhang & Guo, 2012). Therefore, this study fills this gap by answering these questions:

1. What is the status quo of EFL learners' writing self-efficacy and use of writing SRL strategies?
2. What are the relationships between self-efficacy, SRL strategies, and proficiency in an EFL writing context?
3. How much variance in writing proficiency can be explained by writing self-efficacy and use of writing SRL strategies, respectively when gender and SES are controlled?

Method

Participants

Convenience sampling was employed, and participants consisted of 330 sophomores from seven intact classes enrolled in College English Course at two major universities situated in northwest China. College English Course was offered during the first two years at college in China to improve students' comprehensive English proficiency, with a focus on reading and writing. Sophomore students were selected because they have received sufficient writing instructions and practice. The two universities are both tier-one public universities and the curricula and textbooks of English courses are similar for non-English major students in China.

Of the 330 students, 319 signed the informed consent form and agreed to participate in the present study. The majority of the participants (n = 256; 80%) was from University B and the rest (n = 63; 20%) were from University A. The sample consisted of 208 males (65.2%) and 106 females (33.2%). Five students (1.6%) did not report gender information. We used pairwise

deletion to deal with missing values considering the small proportion of these cases. Their age ranged from 18 to 25 (M = 20.02, SD = 0.78), and their years of English learning ranged from 3 to 16 years (M = 10.97, SD = 1.94). An overwhelming majority of students in University A majored in liberal arts (90.5%) and all students in University B majored in science. The participants were predominantly from middle SES families (43.6 %). There were no statistically significant differences between the two universities in either students' age, $t(314) = 1.189$, $p > .05$, $d = 0.16$, or English language learning experience, $t(307) = 0.209$, $p > .05$, $d = 0.03$. Neither were differences between the two universities in their families' socioeconomic status, $x^2(2, n = 289) = 2.22$, $p > .05$. However, statistically significant differences were noted in gender composition, $x^2(1, n = 313) = 70.42$, $p < .001$, and major, $x^2(1, n = 319) = 282.01$, $p < .001$. This is because University A is a liberal arts institution and University B is a science and technology institution. Demographic information of the participants was shown in Table 1-1.

Table 1-1

Descriptive Statistics of Participants' Demographic Information by University

	Variables	University A (n = 63)	University B (n = 256)	Total (n = 319)
Gender	Male (%)	13 (20.6 %)	195 (76.2%)	208 (65.2%)
	Female (%)	49 (77.8 %)	57 (22.3%)	106 (33.2%)
Age	Age M (SD)	20.13 (0.83)	20.00 (0.77)	20.02 (0.78)
Year	Year of English Learning M (SD)	11.01 (1.94)	10.96 (1.94)	10.97 (1.94)
SES	High (%)	13 (23.2%)	35 (15%)	48 (16.6%)
	Middle (%)	23 (41.1%)	103 (44.2%)	126 (43.6 %)
	Low (%)	20 (35.7%)	95 (40.8%)	115 (39.8%)
Major	Liberal Arts (%)	57 (90.5%)	0 (0%)	57 (17.9%)
	Science (%)	6 (9.5%)	256 (100%)	262 (82.1%)

Instruments

The Questionnaire of English Writing Self-Efficacy (QEWSE) was adapted from the Self-Efficacy for Writing Scale (SEWS; Bruning et al., 2013), which was to measure middle and high school students' writing self-efficacy, as well as the Questionnaire of English Self-Efficacy (QESE; Wang & Bai, 2017), which was to measure EFL students' general English self-efficacy. Some items of the SEWS (Bruning et al., 2013) such as "I can punctuate my sentences correctly" were removed because they were not measuring students' writing self-efficacy at the college level. One adaptation was exemplified by changing the item from "I can begin my paragraph in the right spots" into "I can write a paragraph in a coherent way" to cater for students at the college level. These adaptations were made based on formal expert reviews and the College English Curriculum Requirement (2017). The QESE (Wang & Bai, 2017) consists of self-efficacy in four domains of language learning (i.e., listening, speaking, reading, and writing) and only items on the writing construct were drawn because self-efficacy is domain-specific (Bandura, 1997). QEWSE (see Appendix A) consisted of five subscales with 27 items: Ideation (3 items), Organization (5 items), Grammar and Spelling (4 items), Use of English Writing (8 items), and Self-Efficacy for Self-Regulation (7 items). Students were asked to rate their writing self-efficacy beliefs on a 7-point Likert scale, ranging from 1 (I cannot do it at all) to 7 (I can do it well). Working with middle school students, Bruning et al. (2013) reported the Cronbach's alpha value as .90 for the ideation and .88 for the self-regulation. Wang and Bai (2017) administered QESE to secondary school students twice and reported the Cronbach's alpha value for the writing subscale of QESE as .89 for the first assessment, and .91 for the second assessment. For the current study, the internal consistencies (Cronbach's alpha) for QEWSE were .94 (all items), .71 (Ideation), .82 (Organization), .78 (Grammar and Spelling), .87 (Use of

English Writing), and .78 (Self-Efficacy for Self-Regulation) respectively, which indicated good internal consistencies of participants' responses to the questionnaire.

The Questionnaire of English Writing Self-Regulated Learning Strategies (QEWSRLS) was adapted from the Questionnaire of English Self-Regulated Learning Strategies (Wang & Bai, 2017) to measure participants' use of SRL strategies in writing (see Appendix B). For example, we changed "Check my English homework before turning it in" and "Keep reading when I encounter difficulties in English reading" to "Check my composition before turning it in" and "Keep writing when I encounter difficulties in English writing", respectively. Formal expert reviews were gone through to ensure the content validity and face validity of this instrument. The questionnaire included 3 categories with 26 items: (a) Environmental SRL Strategies (8 items); (b) Behavioral SRL Strategies (8 items); (c) Personal SRL Strategies (10 items) based on Zimmerman and Risemberg's (1997) writing self-regulation model. Environmental SRL strategies were also subcategorized into Seeking Assistance Strategies (Items 3, 12, and 21), Persistence Strategies (Item 4, 13, and 22) and Review of Records Strategies (Items 9 and 18). Behavioral SRL strategies comprised Seeking Opportunities Strategies (Items 5, 14, 23, and 25), Self-Monitoring Strategies (Item 6 and 15), and Self-Consequences Strategies (Items 7 and 16). Personal SRL strategies consisted of Self-Evaluation Strategies (Items 1, 10, and 19), Organization and Transformation Strategies (Items 2, 11, 20, 24, and 26), and Goal Setting and Planning Strategies (Items 8 and 17). Participants were asked to respond to items on a 4-point Likert scale from 1(I never use it) to 4 (I often use it). Internal consistency (Cronbach's alpha) and test-retest reliability with secondary school students were reported as .92 and .79, respectively (Wang & Bai, 2017). The Cronbach's alpha values from the current study were .88 for all items, .72 for environmental SRL strategies, .65 for behavioral SRL strategies, and .78 for

personal SRL strategies. Both questionnaires were translated into Chinese using blind back-translation (Brislin, 1970) to ensure the same interpretation of the items in both contexts.

Confirmatory Factor Analysis (CFA) was run using LISREL 8.80 to assess the goodness-of-fit of the five-factor model of QEWSE and three-factor model of QEWSRLS to the data, respectively. Results suggested that the five-factor model of QEWSW fit adequately (NNFI = 0.956, CFI = 0.960, RMSEA = .083, and SRMR = .062) based on the fit indices criteria (Byrne, 2006). CFA was also conducted to evaluate the fit of the three-factor model of QEWSRLS to the data. Results indicated that this model fit as well (NNFI = 0.901, CFI = 0.911, RMSEA = .074, and SRMR = .074). The model fit statistics were presented in Table 1-2 (see Figure 1-1 and Figure 1-2 for the measurement structure of QEWSE and QEWSRLS).

The measures of gender and SES were included in demographic information part of the QEWSE and assessed by students' self-reported information. Gender was coded as "1" for males and "0" for females. SES was measured by the annual family incomes of the students. Students were asked to choose among three options: below $10,000, $10,000-$30,000, and above $30,000. This categorization is based on the National Bureau of Statistics of China and Wen's (2018) report. The three categories were coded as "1" for low SES families, "2" for middle SES families, and "3" for high SES families in the data analysis phase.

Table 1-2

Model Fit Indices for the CFA of QEWSE and QEWSRLS

x^2	*df*	ratio	GFI	AGFI	NFI	NNFI	CFI	SRMR	RMSEA	90%LL	90%UL
1014.46	314	3.23	0.792	0.750	0.944	0.956	0.960	0.062	0.083	0.078	0.089
798.13	293	2.72	0.812	0.775	0.867	0.901	0.911	0.074	0.074	0.068	0.080

Note. GFI = Goodness of Fit; AGFI = Adjusted Goodness of Fit; NFI = Normed Fit Index; NNFI = Non-Normed Fit Index; CFI = Comparative Fit Index; SRMR = Standardized Root Mean Residual; RMSEA = Root Mean Square Error of Approximation; LL = Lower Limit; UL = Upper Limit.

Figure 1-1

Structure of QEWSE

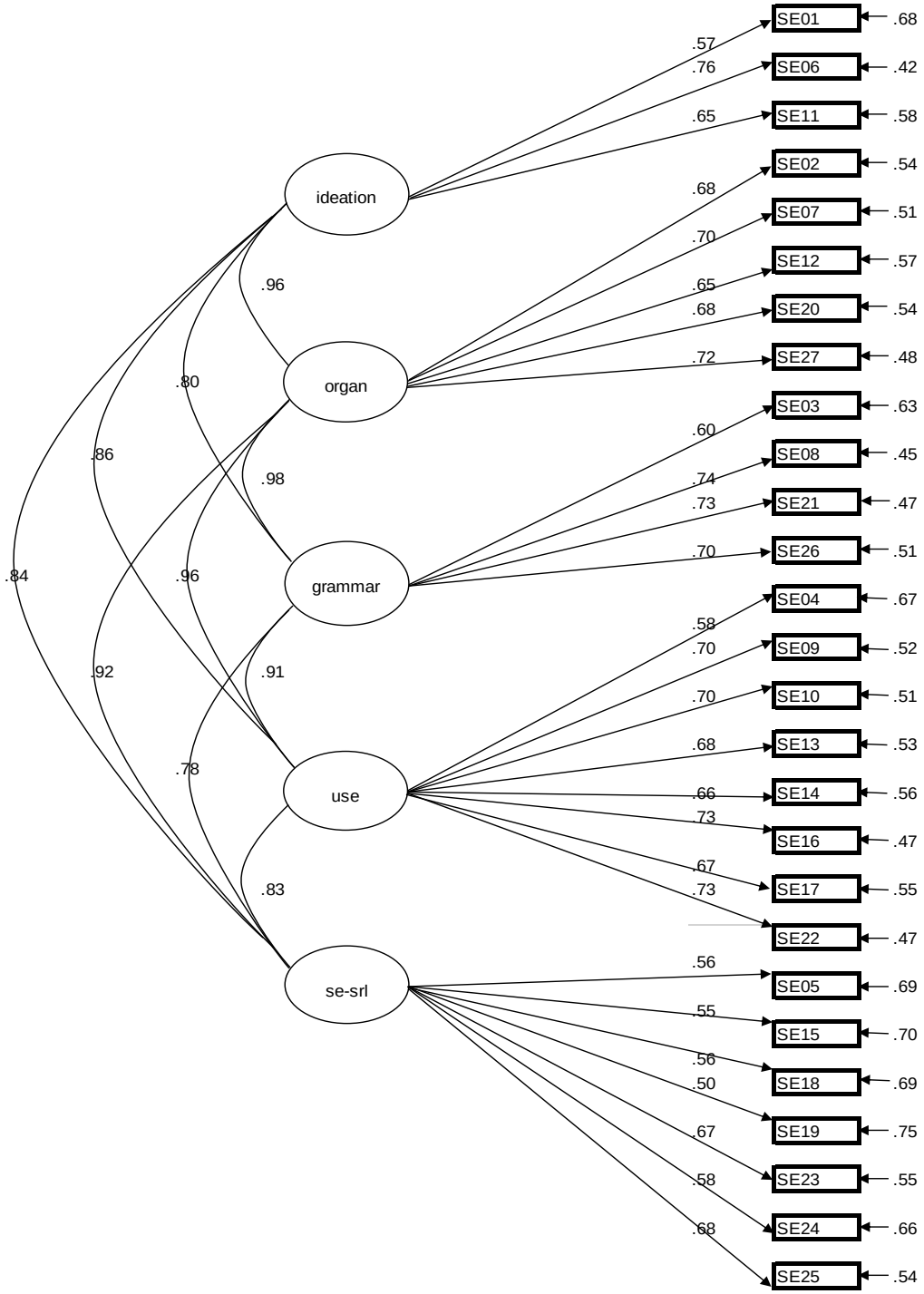
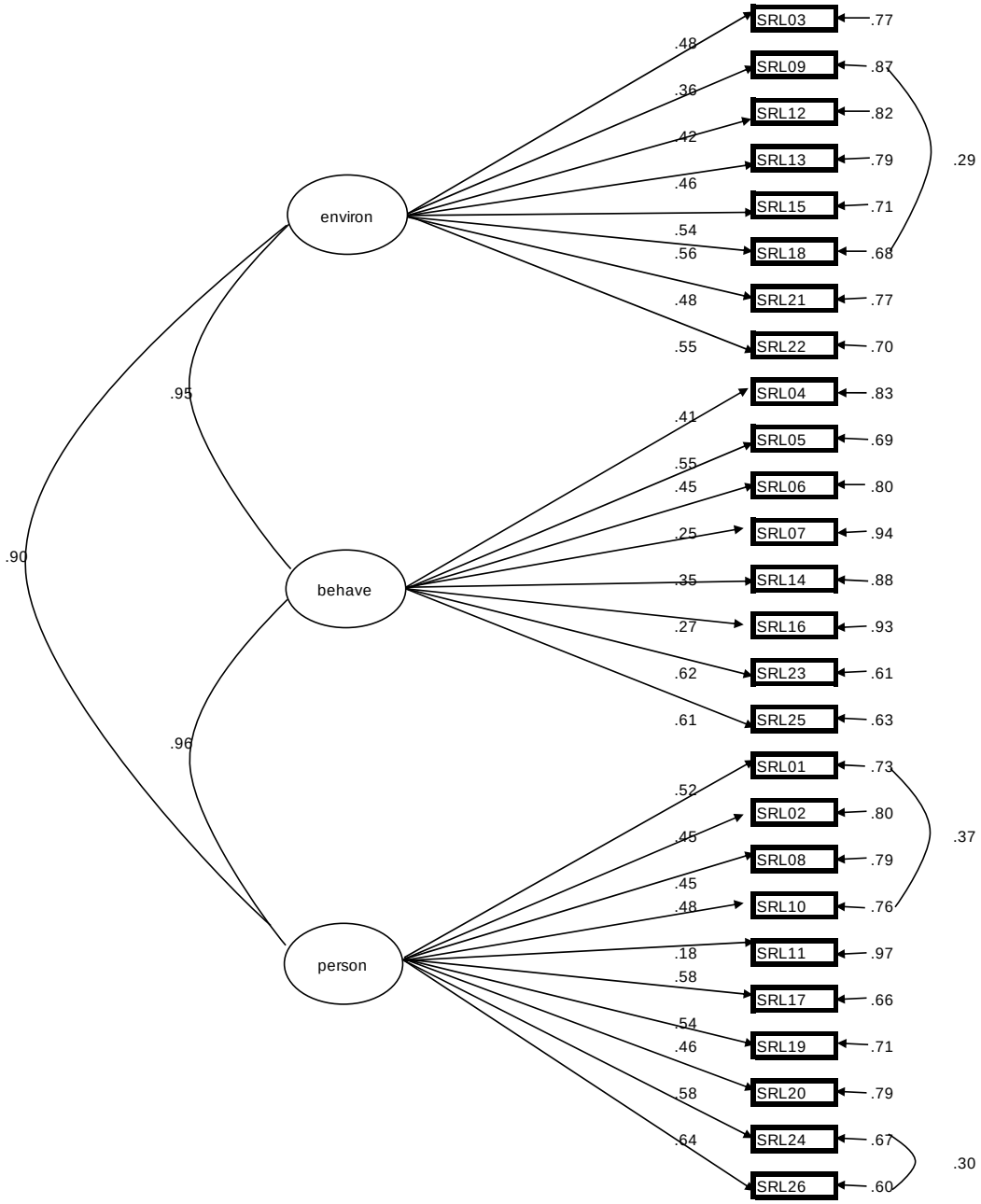


Figure 1-2

Structure of QEWSRLS



Students' writing proficiency was measured by their writing scores on the College English Test Band Four (CET-4), a national standardized test. Colleges students in China have to get a minimum passing score on CET-4 to fulfill the graduation requirement. In addition, it is one of the required documents to show their English proficiency for future job applications. CET-4 consisted of four parts: writing, listening comprehension, reading comprehension, and translation. Scores on the writing and translation section, with the full point of 213 (20% of the total), was used in the current study. The writing section asks students to write an essay of no less than 120 words within 30 minutes (National College English Testing Committee, 2006). The inter-rater and intra-rater reliability were ensured by the rigorous process of selection of and training of raters, the calibration of raters, and the monitoring of the rating process. The validity and reliability of the scores from CET-4 was established by previous researchers (Yang & Weir, 1998).

Procedure

The Institutional Review Board (IRB) of the researchers' university approved the study and all participants returned signed informed consent forms upon entering the study, which specified the purpose of the study, the participants' roles in data collection, and the confidential and voluntary nature of the research. Students who did not sign the consent form were excluded from the current study. The two questionnaires were administered in two class sessions of the College English Course. It took approximately 15 minutes to complete each questionnaire. Students' scores on the CET-4 were provided by their College English Course teachers.

Data Analytical Procedure

Descriptive statistics such as means and standard deviations of the participants' writing self-efficacy, writing SRL strategies and students' writing proficiency were reported and

interpreted. Pearson correlation coefficients were used to represent the relationships between writing self-efficacy beliefs, writing SRL strategy use and writing proficiency. Hierarchical linear regressions were adopted to measure the extent to which writing self-efficacy and writing SRL strategies can explain the variance of participants' writing proficiency after the control of covariates (i.e., in addition to the contribution of the covariates). In order to see if gender, SES, and major areas of study need to be controlled for the regression analyses, multivariate analysis of variance (MANOVA) was used to examine the mean differences between participants.

Results

Status Quo of Self-Efficacy and Self-Regulation

The means and standard deviations of students' writing self-efficacy beliefs, writing self-regulated learning strategies were shown in Table 1-3. Table 1-3 suggested that students' overall writing self-efficacy was at the medium level (M = 4.36, SD = 0.80), between 4 (Maybe I can do it) and 5 (I basically can do it). Students were relatively more efficacious in Organization (M = 4.53, SD = 0.90), but less efficacious in Use of English Writing (M = 4.13, SD = 0.92).

Students' overall SRL strategy score was below 2 (I sometimes use it), indicating that they did not have a frequent use of SRL strategies in writing. Their most frequently used SRL strategies were Organization and Transformation Strategies (M = 1.88, SD = 0.55) and Persistence Strategies (M = 1.81, SD = 0.68). Their least frequently used strategies were Review of Records Strategies (M = 1.16, SD = 0.75) and Goal Setting and Planning Strategies (M = 1.21, SD = 0.69).

Table 1-3

Descriptive Statistics of Writing Self-Efficacy and Writing Self-Regulated Learning Strategies

		n	*M*	*SD*
Writing self-efficacy	Overall	318	4.36	0.80
	Ideation	318	4.52	0.92
	Organization	318	4.53	0.90
	Grammar	318	4.35	0.97
	Use	318	4.13	0.92
	Self-efficacy for SR	318	4.42	0.86
Writing SRL Strategies	Overall	316	1.57	0.43
Environmental	SAS	316	1.57	0.60
	PS	316	1.81	0.68
	RRS	316	1.16	0.75
	Subtotal	316	1.51	0.51
Behavioral	SOS	316	1.47	0.55
	SMS	316	1.50	0.64
	SCS	316	1.35	0.72
	Subtotal	316	1.44	0.46
Personal	SES	316	1.65	0.76
	OTS	316	1.88	0.55
	GPS	316	1.21	0.69
	Subtotal	316	1.58	0.52

Note. SR = Self-Regulation; SAS = seeking assistance strategies; PA = persistence strategies; RRS = review of records strategies; SOS = seeking opportunity strategies; SMS = self-monitoring strategies; SCS = self-consequences strategies; SES = self-evaluation strategies; OTS = organization and transformation strategies; GPS = goal setting and planning strategies.

Relationships

Pearson correlation coefficients (see Table 1-4) noted a significantly positive relationship between writing self-efficacy and writing proficiency, $r = .47$, $p < .001$. All the subcategories of writing self-efficacy were also significantly correlated with writing outcomes, with Grammar having the highest correlation coefficient, $r = .48$, $p < .001$, and Ideation having the lowest coefficient, $r = .29$, $p < .001$.

Table 1-5 suggested writing performance had a moderately positive correlation with writing SRL strategy, $r = .37$, $p < .001$. Except for Self-consequences Strategies, which displayed no significant correlation, $r = .10$, $p > .05$, other subcategories of SRL strategies were all statistically significantly correlated with students' writing proficiency (r ranged from .17 to .38), with Seeking Opportunity Strategies having the highest correlation coefficient, $r = .38$, $p <.001$.

Table 1-4

Pearson Correlation Matrix for Writing Proficiency and Writing Self-Efficacy

	Proficiency	WSE	Ideation	Organization	Grammar	Use	SESR
Proficiency	--	.47***	.29***	.44***	.48***	.45***	.37***
WSE		--	.79***	.92***	.85***	.92***	.85***
Ideation			--	.72***	.59***	.67***	.61***
Organization				--	.78***	.80***	.72***
Grammar					--	.75***	.60***
Use						--	.67***
SESR							--

Note. WSE = Writing Self-Efficacy; SESR = Self-Efficacy for Self-Regulation.

***p<.001, two-tailed.

Table 1-5

Correlation Matrix for Writing Proficiency and Writing SRL Strategies

	Proficiency	WSRLS	SAS	PS	RRS	SOS	SMS	SCS	SES	OTS	GPS
Proficiency	--	.37***	.17**	.24***	.30***	.38***	.23***	.10	.31***	.25***	.19**
WSRLS		--	.69***	.67***	.60***	.78***	.61***	.44***	.68***	.76***	.66***
SAS			--	.43***	.36***	.48***	.39***	.21***	.36***	.48***	.39***
PS				--	.31**	.50***	.37***	.15**	.34***	.45***	.34***
RRS					--	.43***	.30***	.30***	.34***	.30***	.44***
SOS						--	.40***	.23***	.44***	.55***	.45***
SMS							--	.24***	.38***	.35***	.40***
SCS								--	.20***	.18**	.40***
SES									--	.46***	.38***
OTS										--	.37***
GPS											--

Note. WSRL= Writing Self-Regulated Learning Strategies; SAS = seeking assistance strategies; PA = persistence strategies; RRS = review of records strategies; SOS = seeking opportunity strategies; SMS = self-monitoring strategies; SCS = self-consequences strategies; SES = self-evaluation strategies; OTS = organization and transformation strategies; GPS = goal-setting and planning strategies. **p < .01. ***p < .001, two-tailed.

Unique Contribution of Self-Efficacy and Self-Regulation to Writing

In order to see if gender and SES need to be controlled for the regression analyses, a 2×3 MANOVA was conducted using SES and gender as two main factors and writing self-efficacy, SRL strategy and writing proficiency as the dependent variables. The omnibus Wilks's lambda (Λ) was not significant for the main effect of gender, $\Lambda = .991$, $F(4, 260) = 0.824$, $p > .05$. However, the overall Wilks's lambda was statistically significant for the main effect of SES, indicating that the linear combination of the dependent variable differed, on average, across students from various SES background, $\Lambda = .945$, $F(4, 260) = 2.478$, $p < .05$, partial $\eta^2 = .028$. The interaction between gender and SES was not statistically significant, $\Lambda = 0.967$, $F(4, 260) = 1.467$, $p > .05$.

Follow-up univariate *F* statistics suggested no significant gender difference in writing self-efficacy, $F(1, 260) = 1.959$, $p > .05$, writing self-regulated learning strategies, $F(1, 260) = 1.818$, $p > .05$, or writing proficiency, $F(1, 260) = 0.088$, $p > .05$. However, statistically significant differences in writing self-efficacy were noted, $F(2, 260) = 4.464$, $p < .05$, partial $\eta^2 = .033$, between students from low SES families ($M = 4.16$, $SD = 0.77$) and students from middle SES families ($M = 4.51$, $SD = 0.82$). No statistically significant differences existed across students from families with various SES levels in either writing proficiency, $F(2, 260) = 0.642$, $p > .05$, or writing self-regulated learning strategies, $F(2, 260) = 0.022$, $p > .05$. Means and standard deviations of writing self-efficacy, writing SRL strategies and writing proficiency by gender and SES were presented in Table 1-6.

Table 1-6

Means and Standard Deviations of Writing Self-Efficacy, Writing SRL Strategies and Writing Proficiency by Gender and SES

		SE		SRL		Proficiency	
		M	*SD*	*M*	*SD*	*M*	*SD*
Gender	Male	4.28	0.78	1.66	0.42	130.08	24.05
	Female	4.52	0.86	1.51	0.43	134.65	25.10
SES	High	4.48	0.82	1.63	0.39	133.55	25.03
	Middle	4.51	0.82	1.56	0.43	132.96	24.81
	Low	4.16	0.77	1.54	0.45	129.46	23.91

Note. SES = socioeconomic status.

To investigate how well writing self-efficacy and writing SRL strategies predict writing proficiency, two hierarchical linear regression analyses were conducted, with the first model using the composite values of the two variables and the second using the individual constructs. Since MANOVA results indicated statistically significant differences for the main effect of SES, SES was entered in the first step as the covariate, resulting in regression models treating English writing scores (CET-4) as the dependent variable and SES, writing self-efficacy, writing SRL strategies as the independent variables. Gender was not entered into the regression because MANOVA results failed to show gender differences in writing proficiency. The data were screened for missingness and violation of assumptions prior to analysis. The assumptions of linearity, normality, independence, homoscedasticity, and multicollinearity were satisfactorily met.

Table 1-7

Hierarchical Multiple Linear Regression Analysis Summary for Gender, Writing SRL Strategies, and Writing Self-Efficacy Predicting Writing Scores (N = 288)

	Model 1			Model 2		
Variable	*B*	*SE(B)*	β	*B*	*SE(B)*	β
Step 1						
SES_High	2.54	4.26	.04	2.54	4.26	.04
SES_Middle	1.97	3.04	.04	1.97	3.04	.04
R^2	.002 (.002)			.002 (.002)		
Step 2						
SES_High	0.71	3.97	.01	2.46	3.95	.04
SES_Middle	1.43	2.83	.03	0.86	2.80	.02
Writing SRL Strategies	20.56	3.03	.37***			
SAS				-3.65	2.66	-.09
PS				2.01	2.35	.06
RRS				5.38	2.03	.17**
SOS				12.15	3.23	.28***
SMS				2.51	2.34	.07
SCS				-0.86	2.01	-.03
SES				4.40	2.08	.14*
OTS				-0.59	3.10	-.01
GPS				-1.98	2.42	-.06
R^2 (ΔR^2)	.14*** (.14***)			.20 (.20***)		
Step 3						
SES_High	-1.30	3.77	-.02	0.53	3.81	.008
SES_Middle	-1.67	2.73	-.03	-1.43	2.70	-.03
Writing SRL Strategies	6.59	3.71	.12			
Writing Self-Efficacy	11.94	2.01	.40***			
SAS				-2.36	2.54	-.06
PS				1.58	2.28	.04
RRS				4.85	1.93	.15*
SOS				5.30	3.29	.12
SMS				1.72	2.27	.05
SCS				-1.39	1.91	-.04
SES				1.60	2.04	.05
OTS				-3.73	3.03	-.09
GPS				-0.87	2.30	-.03
Ideation				-3.00	2.07	-.12
Organization				2.78	2.99	.11
Grammar				6.57	2.18	.27**
Use				3.65	2.48	.14*
SESR				0.34	2.26	.01
R^2 (ΔR^2)	.23*** (.09***)			.30 (.10***)		

Note. SES = socioeconomic status; SRL= Self-Regulated Learning; SAS = seeking assistance strategies; PA = persistence strategies; RRS = review of records strategies; SOS = seeking opportunity strategies; SMS = self-monitoring strategies; SCS = self-consequences strategies; SES = self-evaluation strategies; OTS = organization and transformation strategies; GPS = goal setting and planning strategies; SESR = Self-Efficacy for Self-regulation.
*p<.05. **p<.01. ***p < .001, two-tailed.

The results of the first regression model indicated that SES did not significantly predict writing scores, $F(2, 291) = 0.29$, $p > .05$, $R^2 = .002$. However, when the construct of writing SRL strategies was entered at Step 2, it significantly improved the prediction, $\Delta R^2 = .14$, $F(3, 290) = 15.61$, $p < .001$. The unstandardized partial slope ($B = 20.56$) and standardized partial slope ($\beta = .37$) for writing SRL strategies were statistically significantly different from zero ($t = 6.79$, $df = 290$, $p < .001$); with one unit increase in the writing SRL strategies, the writing score was expected to increase by 20.56. Furthermore, when Self-Efficacy was added at Step 3, the prediction also improved significantly, $\Delta R^2 = .09$, $p < .001$. The unstandardized partial slope ($B = 11.94$) and the standardized partial slope ($\beta = .40$) were statistically significantly different from zero ($t = 5.94$, $df = 293$, $p < .001$); with one unit increase in writing self-efficacy, the writing score was expected to increase by approximately 11.94 for students from families with the same SES status and who used the same amount of writing SRL strategies. This suggested that writing self-efficacy explains a significant proportion of the variance for writing scores beyond what was accounted for by SES and writing SRL strategies. The model included the entire group of variables significantly predicted writing scores, $F(4, 289) = 21.91$, $p < .001$, $R^2 = .23$ (adjusted $R^2 = .23$), indicating that 23% of the variance in writing scores was predicted by writing self-efficacy, writing SRL strategies, and SES. According to Cohen' (1988) benchmark, effect size can be interpreted as small when $R^2 = 0.01$, medium when $R^2 = 0.09$, and large when $R^2 = 0.25$. This is a medium to large effect according to Cohen (1988). The regression weights, presented in Table 1-7, suggested that writing self-efficacy and writing SRL strategies

contributed significantly to predict writing scores when SES was controlled. The comparison of squared semi-partial correlations suggested that writing self-efficacy contributed most to predicting writing scores.

In the second regression model, SES was entered at Step 1, followed by the nine categories of writing SRL strategies at Step 2, and finally the five constructs of writing self-efficacy at Step 3. The entry of SES did not predict writing score, $F(2, 291) = 0.29$, $p > .05$. However, the entry of nine categories of SRL strategies at Step 2 significantly increased the explained variance ($\Delta R^2 = .20$), with Review of Records Strategies ($\beta = .17$, $p < .01$), Seeking Opportunities Strategies ($\beta = .28$, $p < .001$), and Self-Evaluation Strategies ($\beta = .14$, $p < .05$) statistically significantly predicting students' writing scores. The entry of the five constructs on the final step also statistically improved the predication ($\Delta R^2 = .10$), with Grammar ($\beta = .27$, $p < .01$) and Use of English Writing ($\beta = .14$, $p < .05$) significantly predicting writing scores after the control of SES and the SRL strategies. The model included the entire group of variables significantly predicted writing scores, $F(16, 277) = 7.41$, $p < .001$, $R^2 = .30$, indicating that 30% of the variance in writing scores was explained by SES, nine categories of writing SRL strategies, and the five constructs of writing self-efficacy. This is a large effect according to Cohen (1988).

Discussion

EFL Students' Writing Self-Efficacy and SRL Strategy Use

The first purpose of the study was to investigate the status quo of writing self-efficacy beliefs and writing SRL strategy use among college students. The study found that EFL students reported moderate levels of self-efficacy in writing, which was largely consistent with previous studies (Kim et al., 2015; Wang et al., 2012; Zhang & Guo, 2012). The participants expressed

relatively higher levels of self-efficacy for Organization and lower levels of self-efficacy for Use of English Writing, which suggested that students felt more efficacious in paragraph construction and idea development, but less efficacious in doing practical writings such as sending e-mails to their friends or writing diaries in English. This finding is not surprising, as students in the EFL setting do not have many opportunities to practice writing (Ai, 2015). According to Bandura's (1997) theory of four sources of self-efficacy (i.e., mastery experience, vicarious experience, verbal persuasion, and emotional and physiological states), lack of mastery experience leads to low self-efficacy beliefs. The result could also be explained by the dominant teaching approach in English writing classrooms in China, which is product-oriented and examination-driven with a focus on linguistic skills and grammar (Ai, 2015). Writing tests for college students, including writing sections on CET-4 were designed to measure students' writing abilities on diction, syntax and discoursal development rather than practical use and application of writing in either academic or practical contexts. Future researchers are encouraged to replicate this study in other contexts where the pedagogy is different from that in China. Furthermore, the city in which data was collected is not an international city, and the two universities do not have any international students from English speaking countries. There were only three native speakers of English in University A and two native English speakers in University B. As a result, students in the EFL context do not often communicate with native speakers of English through either emails or social networks, which may account for their lack of confidence in their capabilities in the use of English writing (Chen et al., 2019).

The study also found that students reported infrequent use of SRL strategies in writing, which was partially consistent with Shell and Husman's study (2008), in which students reported infrequent use of strategies such as question-asking strategies and knowledge-building strategies.

It is noteworthy that “Think out a composition in Chinese before writing it in English” was the most frequently used strategy, which suggests that native language transfer exercises a pronounced influence on the writing process among EFL students. Native language transfer was defined as the phenomenon that “learners’ use of prior linguistic information (chiefly form his/her mother tongue) or some physically carryover of native language surface to a second language context” (Dai & He, 2002, p.166), which can be either positive when mother language facilitates language learning process or negative when mother language impedes learning. The use of this transformation strategy may ease the writing process on the part of EFL learners, but it may result in a composition with unidiomatic expressions and grammatical errors. It was found that students reported infrequent use of Goal Setting and Planning Strategies and Review of Records Strategies, adding some support to Asmari’s (2013) study, which found that EFL college students used more “during writing” strategies than “before writing” strategies. According to Roca de Larios et al. (2008), high-achieving students invested more time in planning and revising than on formulating. Therefore, more Goal Setting and Planning Strategies are recommended in writing classes to improve students’ writing proficiency. Additionally, at the item level, the least frequently used strategies students reported were “Send emails to friends in English on my initiative” and “Reward myself when I make a progress in writing”. These students were learning English in an EFL setting and had limited opportunities to use English writing for communication purposes. In addition, self-rewarding was less used by Chinese EFL students. Our speculation is that Chinese students do not receive much positive reinforcement in schooling from either their parents or teachers. Given that reinforcement is more effective than punishment in eliciting desirable behaviors among students (Ormrod et al., 2017), classroom teachers are advised to adopt more positive reinforcement or motivate students to use self-

consequences strategies to encourage their production.

Correlation between Self-Efficacy, SRL Strategies, and Writing

The second research purpose was to examine the relationships between writing self-efficacy, writing self-regulated learning strategy use and students' writing proficiency. The results revealed that writing self-efficacy was positively associated with writing proficiency, which echoed results from previous studies (e.g., Zhang & Guo, 2012). This study also noted that students' writing proficiency had a stronger correlation with self-efficacy for Grammar than with self-efficacy for Ideation. One possible explanation is that in China English compositions are evaluated more on learners' linguistic competence such as lexical, syntactic and grammatical skills than on idea generation and contents (Ai, 2015). Therefore, students who felt confident in grammar tend to have better performance in writing tests.

The study also noted a modest statistically significant correlation between writing proficiency and writing SRL strategy use. This result not only echoed previous studies (Asmari, 2013; Chen, 2011; Chien, 2012) but also contributed to the literature by revealing a specific strong correlation between writing proficiency and Seeking Opportunity Strategies, which suggested that students who exhibit proactive behaviors and take initiative to seek more opportunities to practice writing tend to achieve higher scores in writing tests. However, this study was inconsistent with the previous studies in the strength of correlation. This disparity may be a function of participants' grade level and age as students' self-regulation increases with age and schooling (Graham & Harris, 2000). The differences in the magnitude of the relationships between writing SRL strategies and writing proficiency among students across age and grade levels merit additional investigation.

Predictive Effects of Writing Self-Efficacy and Writing SRL Strategies

The MANOVA results found no gender differences in writing self-efficacy, writing self-regulation, and writing proficiency. These findings, however, contradicted results from some previous studies (e.g., Andrade et al., 2009; Pajares & Valiante, 2001), which found gender differences in writing self-efficacy and writing proficiency favoring girls. Although Pajares and Valiante (2001) noted that gender differences in favor of females were nullified when controlling for feminine orientation beliefs because writing was considered as a stereotypically female activity, this contention needs to be replicated in the EFL writing context. The study found SES was a factor impacting writing self-efficacy but not writing proficiency. This finding was partially consistent with Korat and Schiff's (2005) study, noting that high SES students have significantly higher levels of self-efficacy beliefs than their low SES counterparts. Students from higher SES families may have more resources and more opportunities to practice their English writing. Therefore, their previous successes, known as mastery experience, served as one of the sources of self-efficacy development (Bandura, 1997), which may explain their higher level of self-reported self-efficacy beliefs.

The results of the regression model with the composite variables suggested that writing self-efficacy and writing SRL strategies significantly predicted writing scores after controlling for SES. This confirmed previous research findings that either writing self-efficacy or writing SRL made an independent contribution to the prediction of writing outcomes (Pajares, 2003; Prat-sala & Redford, 2012; Teng & Huang, 2019). The model of individual constructs suggested that Review of Records Strategies, Seeking Opportunities Strategies, and Self-Evaluation Strategies significantly predicted students' writing scores. This finding indicated that students who conduct frequent note reviewing and revising, and seek more opportunities to practice

tended to have higher writing outcomes. This provided support for the study conducted by Roca de Larios et al. (2008), which revealed a positive relationship between the use of evaluation and revision strategies and writing proficiency. Although EFL students adopted frequent use of "in writing" strategies such as composing and organizing (Asmari, 2013), "before writing" and "after writing" strategies such as review and evaluation strategies contribute more to their writing performance. As for self-efficacy, Grammar and Use of English writing significantly predicted writing scores, which suggests that students who are efficacious in their grammar and spelling skills and in doing different genres of writing had higher writing scores. Since students reported the lowest levels of self-efficacy in the Use of English Writing, it is imperative for EFL teachers to include the instruction of various genres of writing such as expository, argumentative, narrative, and descriptive writing and practical writing. In China, writing classes for non-English majors focus more on argumentative writing, and writing purposes, target audiences and features for other genres need to be incorporated in writing instruction and curricula.

Although the present study supported previous studies in some aspects, it is inconsistent with previous research in the magnitude of the relationship of the two variables. In the second article of this dissertation, a meta-analysis on the relationship between self-efficacy and English writing proficiency was conducted and found cultural context is a strong moderator of the relationship, with students from Asian culture (r = .441) having statistically significantly larger effect size than students from Western culture (r = .233).

Limitations and Future Directions

Although great care was taken to research design and processing, this study is limited in many ways. The two universities from which data were collected are homogeneous in terms of students' age, writing proficiency, and learning experience, but heterogeneous in student

population, gender, and major composition. Further studies are recommended to replicate the results with students from more diverse universities to have a better representation of Chinese college students. Second, the five dimensions of QEWSE were highly correlated with each other, which is not surprising since they are all aspects of writing competence but challenges the structural aspect of construct validity of the responses to QEWSE. Future studies should examine the instrument at an item level using Item Response Theory to check item difficulty, item discrimination, item information function, and response patterns.

Significance and Implications

The current study contributes to the cognitive process theory of writing and social cognitive theory by providing further evidence from an EFL writing context in China. The low self-efficacy in English writing was explained by the lack of opportunities to practice, which supports Bandura's (1997) theory of sources of self-efficacy. The positive correlations between writing self-efficacy, writing self-regulation, and English writing proficiency corroborates Hayes' (2000) model of cognitive process theory that motivation/affect is a major component of the writing process. Furthermore, the significant predictive effect of Review of Record Strategies, Seeking Opportunities Strategies, and Self-Evaluation Strategies confirmed the equal importance of the three iterative processes in writing self-regulation suggested by social cognitive theory: environmental, behavioral and personal processes.

This study also has pedagogical implications. Similar to previous studies (e.g., Woodrow, 2011; Zhang & Guo, 2012), this study noted that writing self-efficacy was positively related to proficiency in an EFL context in China. It sheds some light on the understanding of the relationships between writing self-efficacy, writing self-regulation, and students' writing outcomes in an EFL context. EFL teachers in writing are encouraged to nurture students' self-

efficacy through providing more opportunities for writing success (mastery experience), providing opportunities for students to learn from their peers through modeling (vicarious experience), offering immediate and constructive feedback (verbal persuasion), and caring for the physiological and affective states of the students (emotional and physiological states). Moreover, since self-efficacy is a significant predictor of students' writing outcomes, assessment of writing should include both the assessment of students' writing competence and students' self-efficacy (Pajares, 2003). Similarly, it is critical that instructional sessions on writing strategies be incorporated in the classroom settings (Asmari, 2013). De Silva and Graham (2015) noted that strategy instruction benefitted both high and low proficiency learners. Tragant et al. (2013) delineated between deep and surface level strategy clusters and reported that upper-grade students were more inclined to use deep-level strategies and used less surface-level strategies (e.g., memorization of vocabulary and grammar). Some interventions are effective in boosting students' writing self-efficacy and writing self-regulation. For example, Schunk and Zimmerman (2007) contended that students' self-regulatory skills and self-efficacy in writing could be developed and enhanced through modeling interventions. Observing a successful model would not only motivate students in writing but also aid in the internalization and adaptation of writing skills and strategies, which could move students from the observational level through the final self-regulated level. Future studies are recommended to examine how to incorporate strategy instruction into the classroom setting and what SRL strategies are most effective in enhancing students' writing outcomes at various developmental stages through quasi-experimental designs. A Self-Regulated Strategies Development (SRSD) instructional model was also found to statistically significantly improve students' writing performance (Harris et al., 2012).

The study is also significant in providing information about the contribution of each individual factor of writing self-efficacy and writing SRL strategy use to student' writing performance. EFL teachers should encourage students to adopt more review and revision strategies in writing and provide more writing resources and opportunities for them to practice. Additionally, it is imperative for EFL teachers to focus on the instruction of various genres of writing and emphasize the pragmatic aspect of writing in both academic and practical contexts.

References

Ablard, K. E., & Lipschultz, R. E. (1998). Self-regulated learning in high-achieving students: Relations to advanced reasoning, achievement goals, and gender. *Journal of Educational Psychology, 90*, 94-101. http://doi.org/10.1037/0022-0663.90.1.94

Ai, B. (2015). A study of the EFL writing of Chinese learners: A critical narrative. *Changing English, 22*, 294-309. http://doi.org/10.1080/1358684X.2015.1056091

Anastasiou, D, & Michail, D. (2013). Exploring discourse between self-efficacy and writing performance among low-literate adult students. *Learning Disabilities: A Contemporary Journal, 11*, 53-87. Retrieved from http://web.a.ebscohost.com

Andrade, H. L., Wang, X., Du, Y., & Akawi, R. L. (2009). Rubric-Referenced self-assessment and self-efficacy for writing. *The Journal of Educational Research, 102,* 287-302. https://doi.org/10.3200/joer.102.4.287-302

Asmari, A. A. (2013). Investigation of writing strategies, writing apprehension, and writing achievement among Saudi EFL-major students. *International Education Studies, 6*, 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.

Bidjerano, T. (2005, October). *Gender differences in self-regulated learning*. Paper presented at the 36th Annual Meeting of the Northeastern Educational Research Association, Kerhonkson, NY. Retrieved from https://eric.ed.gov/contentdelivery/servlet/ERICServlet?accno=ED490777

Brislin, R. W. (1970). Back-translation for cross-cultural research. *Journal of Cross-Cultural*

Psychology, 1(3), 185-216. http://doi.org/10.1177/135910457000100301

Bruning, R., Dempsey, M., Kauffman, D. F., McKim, C., & Zumbrunn, S. (2013). Examining dimensions of self-efficacy for writing. *Journal of Educational Psychology, 105*, 25-38. http://doi.org/10.1037/a0029692

Bruning, R., & Horn, C. (2000). Developing motivation to write. *Educational Psychologist, 35,* 25-37. http://doi.org/10.1207/S15326985EP3501_4

Byrne, B. M. (2006). *Structural equation modeling with EQS: Basic concepts, applications, and programming* (2nd ed.). Erlbaum.

Callan, G. L., Marchant, G. J., Finch, W. H., & Flegge, L. (2017). Student and school SES, gender, strategy use, and achievement. *Psychology in the Schools, 54*, 1106–1122. https://doi.org/10.1002/pits.22049

Chen, X., Wang, C., & Kim, D.-H. (2019). Self-regulated learning strategy profiles among English as a foreign language learners. *TESOL Quarterly, 54*(1), 234-251. https://doi.org/10.1002/tesq.540

Chen, Y. (2011). Study of the writing strategies used by Chinese non-English majors. *Theory and Practice in Language Studies, 1*(3), 245-251. http://doi.org/10.4304/tpls.1.3.245-251

Chien, S. (2012) Students' use of writing strategies and their English writing achievements in Taiwan. *Asia Pacific Journal of Education, 32*(1), 93-112. http://doi.org/10.1080/02188791.2012.655240

Ching, L. C. (2002). Strategy and self-regulation instruction as contributors to improving students' cognitive model in an ESL program. *English for Specific Purposes, 21*, 261-289. http://doi.org/10.1016/S0889-4906(01)00008-4

Cho, K., Cho, M., & Hacker, D. J. (2010). Self-monitoring support for learning to write. *Interactive Learning Environments, 18*(2), 101-113. https://doi.org/10.1080/10494820802292386

College English Curriculum Requirements. (2017). Ministry of Education of the People's Republic of China.

Dai, W., & He, Z. (2002). 新编简明英语语言学教程. [*A new concise course on linguistics for students of English*]. Shanghai Foreign Language Education Press.

De Silva, R., & Graham, S. (2015). The effects of strategy instruction on writing strategy use for students of different proficiency levels. *System, 53*, 47-59. https://doi.org/10.1016/j.system.2015.06.009

Ekholm, E., Zumbrunn, S., & Conklin, S. (2015). The relation of college student self-efficacy toward writing and writing self-regulation aptitude: Writing feedback perceptions as a mediating variable. *Teaching in Higher Education, 20*, 197-207. http://doi.org/10.1080/13562517.2014.974026

Ensminger, M. E. & Fothergill, K. E. (2003). A Decade of measuring SES: What it tells us and where to go from here. In Bornstein & Bradley (Eds.). *Socioeconomic status, parenting, and child development* (pp. 13–27), Lawrence Erlbaum Associates. http://doi.org/10.2307/30035506

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College Composition and Communication, 32*, 365-387. https://doi.org/10.2307/356600

Graham, S., & Harris, K. R. (2000). The role of self-regulation and transcription skills in writing and writing development. *Educational Psychologist, 35*(1), 3-12. http://doi.org/10.1207/S15326985EP3501_2

Graham, S., Harris, K. R., Mason, L. (2005). Improving the writing performance, knowledge, and self-efficacy of struggling young writers: The effects of self-regulated strategy development. *Contemporary Educational Psychology, 30*, 207-241. https://doi.org/10.1016/j.cedpsych.2004.08.001

Harris, K. R., Lane, K. L., Graham, S., Driscoll, S. A., Sandmel, K., Brindle, M., Schatschneider, C. (2012). Practice-Based professional development for self-regulated strategies development in writing: A randomized controlled study. *Journal of Teaching Education, 63*, 103-119. http://doi.org/10.1177/0022487111429005

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.

Jensen, E. (2009). *Teaching with poverty in mind: What being poor does to kids' brains and what schools can do about it*. ASCD.

Kim, D.-H., Wang, C., Bong, M., & Ahn, H. S (2015). English language learners' self-efficacy profiles and relationship with self-regulated learning strategies. *Learning and Individual Differences, 38,*136-142. https://doi.org/10.1016/j.lindif.2015.01.016

Korat, O., & Schiff, R. (2005). Do children who read more books know "What is good writing" better than children who read less? A comparison between grade levels and SES groups. *Journal of Literacy Research, 37,* 289-324. https://doi.org/10.1207/s15548430jlr3703_2

Lavelle, E., & Guarino, A. J. (2003). A multidimensional approach to understanding college writing processes. *Educational Psychology, 23,* 295-305. http://doi.org/10.1080/0144341032000060138

McGettigan, J. M. (2008). *The relationship between writing achievement, writing self-efficacy, writing apprehension and perceived value of writing by gender for third-grade students in a suburban school district* (Publication No. 10286044) [Doctoral dissertation, Western Connecticut State University). ProQuest Dissertations Publishing.

National College English Testing Committee. (2006). *Syllabus of National College English Test Band 4 and Band 6 (Revised Version)*. Retrieved from http://cet.neea.edu.cn/res/Home/1704/55b02330ac17274664f06d9d3db8249d.pdf

Ormrod, J. E., Anderman, E. M., & Anderman, L. (2017). *Educational psychology: Developing learners* (9th ed.). Pearson.

Oxford, R. (1990). *Language learning strategies: What every teacher should know*. Heinle & Heinle. doi:10.1017/CBO9780511667206

Pajares, F. (2002). Gender and perceived self-efficacy in self-regulated learning. *Theory Into Practice, 41*, 116-125. https://doi.org/10.1207/s15430421tip4102_8

Pajares, F. (2003). Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading & Writing Quarterly, 19*, 139-158. http://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 Counselling and Development, 39*, 239-249. http://doi.org/10.1080/07481756.2007.11909801

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. http://doi.org/10.1006/ceps.1999.1027

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. http://doi.org/10.1006/ceps.2000.1069

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. http://doi.org/10.1080/01443410.2011.621411

Ren, C., & Xin, T. (2013). 家庭社会经济地位对小学生成绩预测效应的追踪研究[Follow-up study on predicting effect of family socioeconomic status on primary school students' academic achievements]. *Education Research, 3,* 79-87.

Roca de Larios, J., Manchón, R., Murphy, L., & Marín, J. (2008). The foreign language writer's strategic behavior in the allocation of time to writing process. *Journal of Second Language Writing, 17*(1), 30-47. https://doi.org/10.1016/j.jslw.2007.08.005

Rose, H., Briggs, J. G., Boggs, J. A., Sergio, L., & Ivanova-Slavianskaia, N. (2018). A systematic review of language learner strategy research in the face of self-regulation. *System, 72,* 151-163. https://doi.org/10.1016/j.system.2017.12.002

Sanders-Reio, J., Alexander, P. A., Reio, Jr., T. G., Newman, I. (2014). Do students' beliefs about writing relate to their writing self-efficacy, apprehension, and performance. *Learning and Instruction, 33,* 1-11. http://doi.org/10.1016/j.learninstruc.2014.02.001

Schunk, D. H., & Zimmerman, B. J. (2007). Influencing children's self-efficacy and self-regulation of reading and writing through modeling. *Reading & Writing Quarterly, 23,* 7-25. http://doi.org/10.1080/10573560600837578

Shell, D. F., Colvin, C., & Bruning, R. H. (1995). Self-efficacy, attribution, and outcome expectancy mechanisms in reading and writing achievement: Grade-level and

achievement-level differences. *Journal of Educational Psychology, 87*, 386-398. http://doi.org/10.1037/0022-0663.87.3.386

Shell, D. F., & Husman, J. (2008). Control, motivation, affect, and strategic self-regulation in the college classroom: A multidimensional phenomenon. *Journal of Educational Psychology, 100*, 443-459. http://doi.org/10.1037/0022-0663.100.2.443

Shell, D. F., Murphy, C. C., & Bruning, R. H. (1989). Self-efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology, 81*, 91-100. http://doi.org.librarylink.uncc.edu/10.1037/0022-0663.81.1.91

Sirin, S. R. (2005). Socioeconomic status and academic achievement: A meta-analysis review of research. *Review of Educational Research, 75*, 417-453. https://doi.org/10.3102/00346543075003417

Steward, G., Steifert, T. A., Rolheiser, C. (2015). Anxiety and self-efficacy's relationship with undergraduate students' perceptions of the use of metacognitive writing strategies. *Canadian Journal for the Scholarship of Teaching and Learning, 6*(1), 1-17. http://doi.org/10.5206/cjsotl-rcacea

Teng, F., & Huang, J. (2019). Predictive effects of writing strategies for self-regulated learning on secondary school learners' EFL writing proficiency. *TESOL Quarterly, 53*, 232-247. https://doi.org/10.1002/tesq.462

Teng, L. S., Sun, P. P., & Xu, L. (2018). Conceptualizing writing self-efficacy in English as a foreign language context: Scale validation through structural equation modeling. *TESOL Quarterly, 52*, 911-942. https://doi.org/10.1002/tesq.432

Teng, L. S., & Zhang, L. J. (2016). A questionnaire-based validation of multidimensional models of self-regulated learning strategies. *The Modern Language Journal, 100,* 674-701. https://doi.org/10.1111/modl.12339

Tragant, E., Thompson, M. S., & Victori, M. (2013). Understanding foreign language learning strategies: A validation study. *System, 41,* 95-108. https://doi.org/10.1016/j.system.2013.01.007

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,* 751-796. https://doi.org/10.3102/0034654308321456

Wang, C., & Bai, B. (2017). Validating the instruments to measure ESL/EFL learners' self-efficacy beliefs and self-regulated learning strategies. *TESOL Quarterly, 51,* 931-947. https://doi.org/10.1002/tesq.355

Wang, C., Hu, J., Zhang, G., Chang, Y., & Xu, Y. (2012). Chinese college students' self-regulated learning strategies and self-efficacy beliefs in learning English as a foreign language. *Journal of Research in Education, 22*(2), 103-135.

Wen, H., Liang, K., & Liu, X. (2016). 家庭环境对中学生阅读能力的影响: 阅读投入、阅读兴趣的中介作用 [The influence of family environment on middle school students' reading ability: The mediating role of reading engagement and interest]. *Psychological Journal, 48,* 248-257.

Wen, L. (2018). Income and living standards across China. *Federal Reserve Bank of St. Louis.* Retrieved from https://www.stlouisfed.org/on-the-economy/2018/january/income-living-standards-china

Woodrow, L. (2011). College English writing affect: Self-efficacy and anxiety. *System, 39,* 510-522. https://doi.org/10.1016/j.system.2011.10.017

Yang, H., & Weir, C. J. (1998). *The CET validation study*. Shanghai Foreign Language Education Press.

Zhang, Y., & Guo, H. (2012). A study of English writing and domain-specific motivation and self-efficacy of Chinese EFL learners. *Pan-Pacific Association of Applied Linguistics, 16*(2), 101-121. Retrieved from http://web.a.ebscohost.com

Zimmerman, B., & Risemberg, R. (1997). Becoming a self-regulated writer: A social cognitive perspective. *Contemporary Educational Psychology, 22*(1), 73-101. https://doi.org/10.1006/ceps.1997.0919

APPENDIX A: QUESTIONNAIRE OF ENGLISH WRITING SELF-EFFICACY (QEWSE)

Notes: Please read the following questions carefully and make an accurate evaluation of your current command of English no matter whether you are doing it or not. These questions are designed to measure your judgment of your capabilities, so there are no right or wrong answers.

Please use the following scales to answer these questions accordingly. Please choose the number accurately representing your capabilities.						
1	2	3	4	5	6	7
I cannot do it at all	I cannot do it.	Maybe I cannot do it.	Maybe I can do it.	I basically can do it.	I can do it.	I can do it well.

1. I can think of many ideas for my writing.	1	2	3	4	5	6	7
2. I can organize sentences into a paragraph to express an idea.	1	2	3	4	5	6	7
3. I can correctly spell all the words in the compositions I write.	1	2	3	4	5	6	7
4. I can compose messages in English on the internet through social network (e.g., WeChat and blogs)?	1	2	3	4	5	6	7
5. I can focus on my writing for at least one hour.	1	2	3	4	5	6	7
6. I can put my ideas into writing.	1	2	3	4	5	6	7
7. I can organize different paragraphs into a composition.	1	2	3	4	5	6	7
8. I can correctly use verb tenses in English writing.	1	2	3	4	5	6	7
9. I can make new sentences with given words.	1	2	3	4	5	6	7
10. I can write an expository paragraph in English.	1	2	3	4	5	6	7
11. I can think of appropriate words to describe my ideas.	1	2	3	4	5	6	7
12. I can focus on the main ideas when writing.	1	2	3	4	5	6	7
13. I can write an argumentative paragraph in English.	1	2	3	4	5	6	7
14. I can write email messages in English.	1	2	3	4	5	6	7
15. I can finish writing assignments in time.	1	2	3	4	5	6	7
16. I can write a descriptive paragraph in English.	1	2	3	4	5	6	7
17. I can write diaries in English.	1	2	3	4	5	6	7
18. I can plan what I want to say before I start writing.	1	2	3	4	5	6	7
19. I can avoid distractions while I write.	1	2	3	4	5	6	7
20. I can write a paragraph in a cohesive way.	1	2	3	4	5	6	7
21. I can write a sentence with proper grammatical structures.	1	2	3	4	5	6	7
22. I can write a narrative paragraph in English.	1	2	3	4	5	6	7
23. I can revise my writing to make it better.	1	2	3	4	5	6	7
24. I can control my frustration when I write.	1	2	3	4	5	6	7
25. I can keep writing even when it’s difficult.	1	2	3	4	5	6	7
26. I can fix my grammar errors.	1	2	3	4	5	6	7
27. I can write a paragraph in a coherent way.	1	2	3	4	5	6	7

APPENDIX B: QUESTIONNAIRE OF ENGLISH WRITING SELF-REGULATED LEARNING STRATEGIES (QEWSRLS)

Notes: Please read the following questions carefully and make an accurate evaluation of how often you use the following self-regulated learning strategies in the English writing context. These questions are designed to measure your judgment of frequency, so there are no right or wrong answers.

Please use the following scales to answer these questions accordingly. Please choose the number accurately representing your capabilities.			
0	1	2	3
I never use it.	I seldom use it.	I sometimes use it.	I often use it.

1. Check my English composition before turning them in.	0	1	2	3
2. Write an outline before writing English compositions.	0	1	2	3
3. Consult teachers when I encounter difficulties in my English writing.	0	1	2	3
4. Keep writing when I encounter difficulties in English writing.	0	1	2	3
5. Use sentence patterns just learned to make new sentences for practice in writing.	0	1	2	3
6. Write down the mistakes I often make in the process of writing.	0	1	2	3
7. Reward myself when I make a progress in writing.	0	1	2	3
8. Set a goal to improve my writing.	0	1	2	3
9. Review English texts I have learned before writing.	0	1	2	3
10. Proofread my English composition after I complete writing.	0	1	2	3
11. Think out a composition in Chinese before writing it in English.	0	1	2	3
12. Ask classmates when I have questions in my English writing.	0	1	2	3
13. When a friend wants to play with me, but I have not finished my writing yet, I do not play until I finish my writing.	0	1	2	3
14. Send emails to friends in English on my initiative.	0	1	2	3
15. Take notes in English writing classes.	0	1	2	3
16. Have a break when I am tired during writing.	0	1	2	3
17. Make a plan in the process of English writing.	0	1	2	3
18. Review my notes of English class before writing.	0	1	2	3
19. When I finish my English composition, I have a rest and then read it again to check whether it should be revised.	0	1	2	3
20. Make sure to write a topic sentence in each paragraph in writing.	0	1	2	3
21. Search related documents when I have difficulties in English writing.	0	1	2	3
22. Find a quiet place to write when the environment is disturbing.	0	1	2	3
23. Try to use various English expressions to express the same meaning in writing.	0	1	2	3
24. Make sure that the content of each paragraph supports its topic sentence in English writing.	0	1	2	3
25. Use words just learned to make new sentences on my initiative in writing.	0	1	2	3
26. Pay attention to the English language structure during writing.	0	1	2	3

CHAPTER 2 [ARTICLE II]: RELATIONSHIP BETWEEN WRITING SELF-EFFICACY AND SECOND LANGUAGE WRITING ACHIEVEMENT IN ENGLISH: A META-REGRESSION ANALYSIS

Sun, T., Wang, C., Lambert, R. G., & Liu, L. (In press). Relationship between writing self-efficacy and second language writing achievement in English: A meta-regression analysis. *Journal of Second Language Writing*.

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 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.

Literature Review

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 outcomes 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.

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 (Goldberg et al., 2003), formative assessment (Graham et al., 2015), and written corrective feedback (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.

Moderating Effect of L1/L2 Writing on the Effect Sizes

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 even found a negative (2004) 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 estimate the average effect size by conducting a systematic review and 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 (e.g., sample size, gender, and grade level) and study characteristics (e.g., statistical procedures,

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.

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 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.

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.

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 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.

Statistical Procedures

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).

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 sizes.

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 methods, 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:

1. What is the magnitude of the relationship between writing self-efficacy and writing proficiency for first and second language writers in English?
2. How does writing in English as a first or a second language moderate the relationship between writing self-efficacy and writing proficiency after controlling for study characteristics or characteristics related to effect sizes?

Methodology

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/theses after removing duplicates. Then the 1447 documents were imported into the computer software MAXQDA for screening.

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, this 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 and 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 reported 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 et al., 2000; Pajares et al., 2001; Pajares & Valiante, 2001).

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 2-1. The PRISMA flow diagram of literature search and screening processes are presented in Figure 2-1 (Moher et al., 2009).

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.

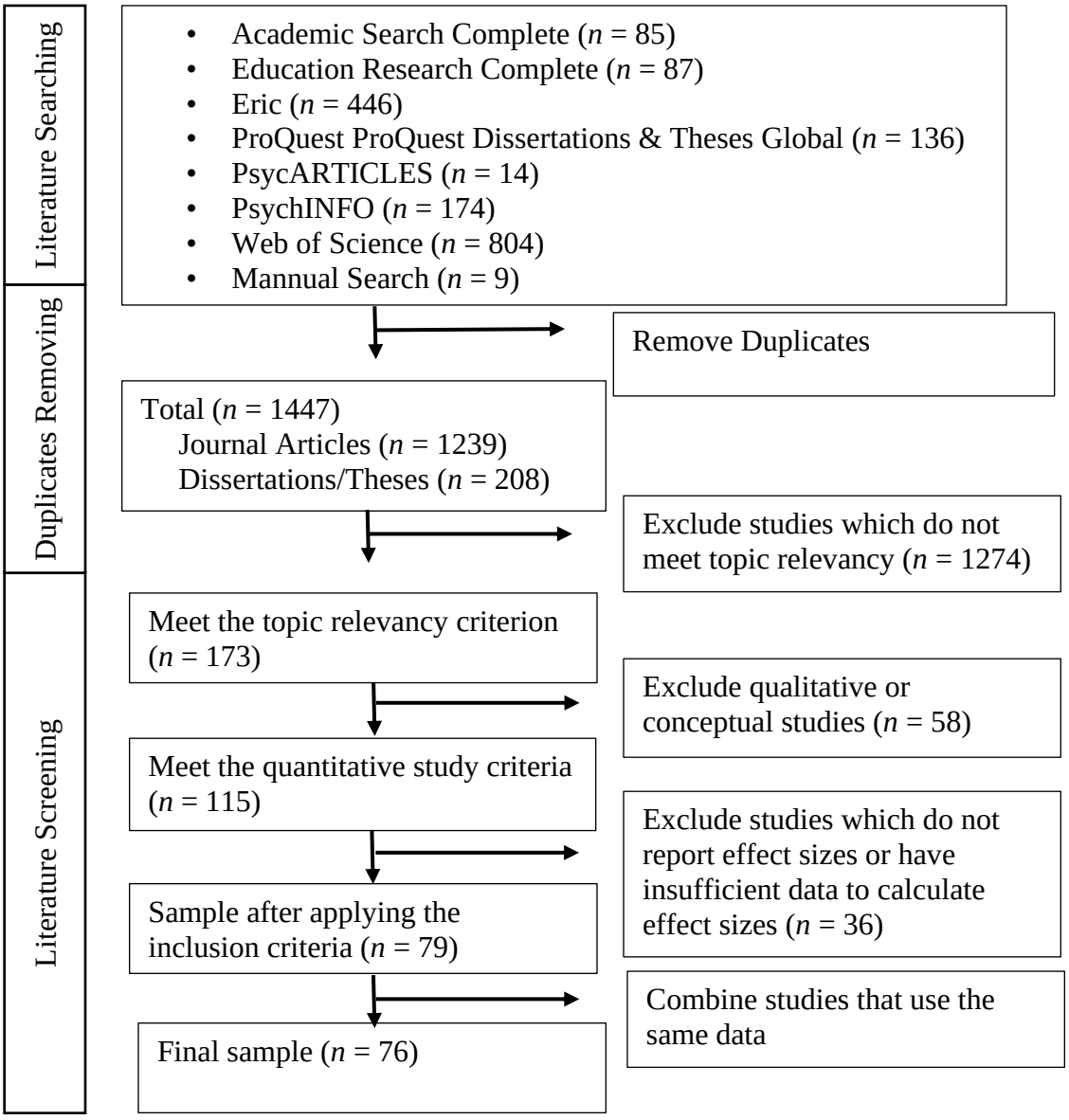
Table 2-1

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.

Figure 2-1

PRISMA Flow Diagram of Literature Search and Screening



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 for 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 school, 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.

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 5*k*+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.

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).

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.

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

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

$$r = \frac{e^{2z}-1}{e^{2z}+1} \quad (2)$$

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). Multilevel 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.

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:

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

$$\beta_{qj} = \gamma_{q0};\ q = 1, 2, 3, 4$$

Results

This meta-regression synthesized 565 effect sizes nested within 76 primary studies. Descriptive statistics of the independent variables are presented in Table 2-2. 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.

Table 2-2

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.

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 model.

The estimates of the predictors in the conditional model are presented in Table 2-3. Note that the values in this Table 2-3 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, $t(58.86) = -0.83$, $p > .05$.

Table 2-3

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.

The effect sizes under certain conditions were also estimated. Table 2-4 presents both the z-transformed estimates, and r estimates that were calculated using Equation (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 2-4

Model-Based Effect Size Estimates

Effect Size Estimates	z 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.

Discussion

The purposes of the present study were to (a) 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) examine the extent to which writing in English as an L1 or L2 moderates 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 (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 being found 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 students writing achievement for L2 students, it is imperative that English language teachers in the English as a foreign/second language (EFL/ESL) context employ effective strategies to nurture L2 writers' self-efficacy.

The present study found that sample size positively and significantly impacted effect size magnitude, which contradicted the observation that effect sizes and sample sizes were negatively correlated (Slavin & Smith, 2009). Methodologists contended that studies with small sample sizes tended to report large 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 effects 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.

This study found statistically significant differences in effect sizes as a function of statistical method, with Pearson correlation yielding larger effect sizes than regression. This result was expected because *r* is the bivariate correlation coefficient whereas β is the partial regression coefficient while controlling for other variables. The partial regression coefficient adjusted for the correlation between writing self-efficacy and other predicting variables and for the correlation between writing achievement and these variables in a regression model (Kline, 2016), which may explain the attenuation in effect sizes. The statistically significantly smaller effect sizes found in regression models after controlling for affective variables and achievement variables may conveyed two messages. First, writing self-efficacy tended to covary with affective or achievement variables (Pajares et al., 2000). Second, there were significant correlations between writing achievement and other affective and achievement variables (Zhang & Guo, 2012). Results also suggested 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 study (2013), 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 sizes, 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.

Limitations 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 are suggested to be included for further studies for purposes of obtaining more presentative data and further minimizing publication bias. Second, the two variables (i.e., 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 methodological 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 ethnicity.

Conclusion and Implications

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. Results provide evidence that statistically significantly larger effect sizes of the relationship between writing self-efficacy and English writing achievement was found with 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) believed that the development and nurture of self-efficacy come from four sources, 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 creating more opportunities for students to achieve success in writing, provide opportunities for students to learn from their peers through group work and modeling, offer encourage 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 L2 English 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 controlling for different variables. 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, their covariances needs 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.

References

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

*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 University, Addis Ababa, Ethiopia.

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.

*Anderson, C. B., Lee, H. Y., Byars-Winston, A., Baldwin, C. D., Cameron, C., & Chang, S. (2016). Assessment of scientific sommunication self-efficacy, interest, and outcome expectations for career development in academic medicine. *Journal of Career Assessment, 24*(1), 182–196. https://doi.org/10.1177/1069072714565780

Archibald, A. (2016). Writing in a second language. Center for Language Linguistics and Area Studies. Retrieved from 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.

*Bissell, K., & Collins, S. J. (2001). Early predictors of ability in a news writing course. *Journalism & Mass Communication Educator*, *56*(2), 69–80. https://doi.org/10.1177/107769580105600206

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. http://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

*Callinan, C. J., van der Zee, E., & Wilson, G. (2018). Developing essay writing skills: An evaluation of the modelling behavior method and the influence of student self-efficacy. *Journal of Further and Higher Education, 42*(5), 608–622. https://doi.org/10.1080/0309877X.2017.1302564

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. (2011). *Contributions of prior knowledge, motivation, and strategies to Korean college students' L2 writing development* (Publication No. 3495368) [Doctoral dissertation, University of Maryland]. ProQuest Dissertations Publishing.

*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. *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. http://doi.org/10.1016/j.asw.2019.03.002

*Chen, M. C., & Lin, H.-J. (2009). Self-efficacy, foreign language anxiety as predictors of academic performance among professional program students in a general English proficiency writing test. *Perceptual and Motor Skills*, *109*(2), 420–430. https://doi.org/10.2466/pms.109.2.420-430

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.

*Corkett, J., Hatt, B., & Benevides, T. (2011). Student and teacher self-efficacy and the connection to reading and writing. *Canadian Journal of Education, 34*(1), 65-98.

*Dempsey, M. S. (2013). *Self-efficacy for metalinguistic control and its relationship to writing quality* (Publication No. 3601000) [Doctoral dissertation, The University of Nebraska - Lincoln]. ProQuest Dissertations Publishing.

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

*Dowd, J. E., Thompson, R. J., Schiff, L., Haas, K., Hohmann, C., Roy, C., Meck, W., Bruno, J., & Reynolds, J. A. (2019). Student learning dispositions: Multidimensional profiles highlight important differences among undergraduate STEM honors thesis writers. *CBE—Life Sciences Education*, *18*(2), 1-12. https://doi.org/10.1187/cbe.18-07-0141

*Duijnhouwer, H, Prins, F. J., & Stokking, K. M. (2010). Progress feedback effects on students' writing mastery goal, self-efficacy beliefs, and performance. *Educational Research and Evaluation, 16*(1), 53-74. http://doi.org/10.1080/13803611003711393

*Eggleston, B. N. (2017). *Relationship between writing self-efficacy and writing fluency in a performance feedback intervention* (Publication No. 10271772) [Master's thesis, Syracuse University]. ProQuest Dissertations Publishing.

*Fatemi, A. H., & Vahidnia, F. (2013). An Investigation into Iranian EFL learners' level of

writing self-efficacy. *Theory and Practice in Language Studies, 3*(9), 1698–1704. https://doi.org/10.4304/tpls.3.9.1698-1704

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

*Goodman, S. B., & Cirka, C. C. (2009). Efficacy and anxiety: An examination of writing attitudes in a first-year seminar. *Journal on Excellence in College Teaching, 20*(3), 5-28.

Graham, S. (2006). Writing. In P. Alexander & P. Winne (Eds.), *Handbook of educational psychology* (pp. 457–478). Erlbaum.

*Graham, S., Daley, S. G., Aitken, A. A., Harris, K. R., & Robinson, K. H. (2018). Do writing motivational beliefs predict middle school students' writing performance? *Journal of Research in Reading, 41*(4), 642–656. https://doi.org/10.1111/1467-9817.12245

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., Kiuhara, S. A., Harris, K. R., & Fishman, E. J. (2017). The relationship among strategic writing behavior, writing motivation, and writing performance with young, developing writers. *The Elementary School Journal, 118*(1), 82-104. http://doi.org/10.1086/693009

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. http://doi.org/10.1037/a0029185

*Graham, S., Wijekumar, K., Harris, K., Lei, P., Fishman, E., Ray, A. B., & Houston, J. (2019). Writing skills, knowledge, motivation, and strategic behavior predict students' persuasive writing performance in the context of robust writing instruction. *The Elementary School Journal, 119*(3), 487-510. 487-510. http://doi.org/10.1086/701720

*Greene, G. L. (1999). *Writing self-efficacy, gender, aptitude, and writing achievement among freshman university students* (Publication No. 9935551) [Doctoral dissertation, The University of Alabama]. ProQuest Dissertations Publishing.

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

*Hashemnejad, F., Zoghi, M., & Amini, D. (2014). The relationship between self-efficacy and writing performance across genders. *Theory and Practice in Language Studies, 4*(5), 1045–1052. https://doi.org/10.4304/tpls.4.5.1045-1052

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.

*Heffernan, J. E. (2014). *From Ideation to Paper: Cognition, Transcription, and Composition* (Publication No. 3629961) [Doctoral dissertation, Saint Mary's College of California]. ProQuest Dissertations Publishing.

Huang, C. (2013). Gender differences in academic self-efficacy: A meta-analysis. *European Journal of Psychology of Education, 28*, 1-35. http://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.

*Jones, E. (2008). Predicting performance in first-semester college basic writers: Revisiting the role of self-beliefs. *Contemporary Educational Psychology, 33*, 109-238. http://doi.org/10.1016/j.cedpsych.2006.11.001

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

*Karaglani, A. H. (2001). *Examining the relationship between writing self-efficacy, writing performance and general achievement for third graders* (Publication No. 30655804) [Doctoral dissertation, University of Southern California]. ProQuest Dissertations Publishing.

*Klassen, R. M. (2007). Using predictions to learn about the self-efficacy of early adolescents with and without learning disabilities. *Contemporary Educational Psychology, 32,* 173-187. http://doi.org/10.1016/j.cedpsych.2006.10.001

Kline, R. B. (2016). *Principles and practice of structural equation modeling*. Guilford.

*La Femina, J. V. (2000). *Teaching students to regulate their emotions and writing performance through relaxation/thought replacement and learning strategy use* (Publication No. 9969705) [Doctoral dissertation, City University of New York]. ProQuest Dissertations Publishing.

*Lambert, T. I. (2015). *A structural equation model of the influence of personal, behavioral, and environmental factors on the writing performance of first-year students at a selected Michigan community college* (Publication No. 10162843) [Doctoral dissertation, Andrews University]. ProQuest Dissertations Publishing.

*Lane, A. M. (2003). Self-efficacy and dissertation performance among sport students. *The Journal of Hospitality Leisure Sport and Tourism*, *2*(2), 59–66. https://doi.org/10.3794/johlste.22.46

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. http://doi.org/10.1037/a0031391

Lipsey, M. W., & Wilson, D. B. (1993). The efficacy of psychological, educational, and behavioral treatment: Confirmation from meta-analysis. *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

*McGettigan, J. M. (2008). *The relationship between writing achievement, writing self-efficacy, writing apprehension and perceived value of writing by gender for third-grade students in a suburban school district* (Accession or Order No. 10286044) [Doctoral dissertation, Western Connecticut State University]. ProQuest Dissertations Publishing.

*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. http://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). Institute of Education Sciences, U.S. Department of Education, Washington, D.C. Retrieved from https://nces.ed.gov/nationsreportcard/pdf/main2011/2012470.pdf

National Commission on Writing. (2004, September). *Writing: A ticket to work...or a ticket out: A survey of business leaders.* Retrieved May 3, 2005, from http://www.writingcommission.org/report/html

*Neugebauer, S. R., & Howard, E. R. (2015). Exploring associations among writing self-perceptions, writing abilities, and native language of English-Spanish two-way immersion students. *Bilingual Research Journal*, *38*(3), 313–335. https://doi.org/10.1080/15235882.2015.1093039

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 & Writing Quarterly, 19*(2), 139-158. http://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, http://doi.org/10.1080/07481756.2007.11909801

*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. http://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. http://doi.org/10.1080/07481756.2001.12069012

*Pajares, F., & Johnson, M. J. (1994). Confidence and competence in writing: The role of self-efficacy, outcome expectancy, and apprehension. *Research in Teaching of English, 28*(3), 313-331.

*Pajares, F., & Johnson, M. J. (1996). Self-efficacy beliefs and the writing performance of entering high school students. *Psychology in the Schools, 33*, 163-175. http://doi.org/10.1002/(SICI)1520-6807(199604)33:2%3C163::AID-PITS10%3E3.0.CO;2-C

*Pajares, F., Johnson, M. J., & Miller, M. D. (1999). Gender Differences in Writing Self-Beliefs of Elementary School Students. *Journal of Educational Psychology, 91(*1), 50-61. http://doi.org/10.1037/0022-0663.91.1.50

*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. (1997). Influence of self-efficacy on elementary students' writing. *The Journal of Educational Research, 90*(6), 353-360. http://doi.org/10.1080/00220671.1997.10544593

*Pajares, F., & Valiante, G. (1999). Grade level and gender differences in the writing self-beliefs of middle school students. *Contemporary Educational Psychology, 24*(4), 390–405. https://doi.org/10.1006/ceps.1998.0995

*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. http://doi.org/10.1006/ceps.2000.1069

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 (Series Ed.) and P. Boscolo & S. Hidi (Volume Eds.), *Studies in writing, Vol. 19: Writing and motivation* (pp. 141-159). Elsevier.

*Pelopida, A. (2016). *The relationship of grade 12 high school students' perceptions of writing self-efficacy and academic writing outcomes in a suburban high school* (Publication No. 10031621) [Doctoral dissertation, Johnson & Wales University]. ProQuest Dissertations Publishing.

*Perin, D., Lauterbach, M., Raufman, J., & Kalamkarian, H. S. (2017). Text-based writing of low-skilled postsecondary students: relation to comprehension, self-efficacy and teacher judgments. *Reading and Writing, 30*, 887-915. http://doi.org/10.1007/s11145-016-9706-0

Peterson, R. A., & Brown, S. (2005). On the use of Beta coefficients in meta-analysis. *Journal of Applied Psychology*, *90*(1), 175-181. https://doi.org/10.1037/0021-9010.90.1.175

*Phillips, W. C. (2007). *Writing behaviors and attitude: A survey of writing process behaviors, self-efficacy, writing achievement, and writing preferences of sixth, seventh, and eighth grade students* (Publication No. 3281330) [Doctoral dissertation, George Fox University]. ProQuest Dissertations Publishing.

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. http://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. http://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. http://doi.org/10.1007/s11218-018-9434-1

*Rankin, J. L., Bruning, R. H., & Timme, V. L. (1993). Is writing affected by spelling performance and beliefs about spelling? *Applied Cognitive Psychology, 7,* 155-169. http://doi.org/10.1002/acp.2350070207

*Raoofi, S., & Maroofi, Y. (2017). Relationships among motivation (self-efficacy and task value), strategy use and performance in L2 writing. *Southern African Linguistics and Applied Language Studies*, *35*(3), 299–310. https://doi.org/10.2989/16073614.2017.1391706Ì

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. *The Journal of Experimental Education*. Advance online publication. https://doi.org/10.1080/00220973.2020.1808944

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, Ltd.

Ruan, Z (2014) Metacognitive awareness of EFL student writers in a Chinese ELT context, *Language Awareness, 23,* 76-91. http://doi.org/10.1080/09658416.2013.863901

*Saban, A. I., & Erkan, D. Y. (2011). Writing performance relative to writing apprehension, self-efficacy in writing, and attitudes towards writing: A correlational study in the Turkish tertiary-level EFL context. *The Asian EFL Journal Quarterly, 13*(1), 163-191.

*Sanders-Reio, J. (2010). *Investigation of the relations between domain-specific beliefs about writing, writing self-efficacy, writing apprehension, and writing performance in undergraduates* (Publication No. 3409651) [Doctoral dissertation, University of Maryland]. ProQuest Dissertations Publishing.

*Sanders-Reio, J., Alexander, P. A., Reio, Jr., T. G., Newman, I. (2014). Do students' beliefs about writing relate to their writing self-efficacy, apprehension, and performance. *Learning and Instruction, 33,* 1-11. http://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. http://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. http://doi.org/10.1080/00461520.1991.9653133

*Schunk, D. H., & Swartz, C. W. (1993). Goals and progress feedback: Effects on self-efficacy and writing achievement. *Contemporary Educational Psychology, 18*, 337-354. http://doi.org/10.1006/ceps.1993.1024

*Shaw, E. J. (2007). *The reading and writing self-efficacy beliefs of students with discrepant reading and writing performance* (Publication No. 3302121) [Doctoral dissertation, Fordham University]. ProQuest Dissertations Publishing.

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

*Shell, D. F., Murphy, C. C., & Bruning, R. H. (1989). Self-efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology, 81*(1), 91-100. http://doi.org/10.1037/0022-0663.81.1.91

*Sieben, N. (2013). *Writing hope, self-regulation, and self-efficacy as predictors of writing ability in first-year college student* (Publication No. 3594312) [Doctoral dissertation, Hofstra University]. ProQuest Dissertations Publishing.

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

*Soylu, M. Y., Zeleny, M. G., Zhao, R., Bruning, R. H., Dempsey, M. S., & Kauffman, D. F. (2017). Secondary students' writing achievement goals: Assessing the mediating effects of mastery and performance goals on writing self-efficacy, affect, and writing achievement. *Frontiers in Psychology, 8*, 1-11. https://doi.org/10.3389/fpsyg.2017.01406

*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*, 102221. https://doi.org/10.1016/j.system.2020.102221

*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

*Thomas, S. C. (2013). *Self-efficacy and preparation of scholarly writing: Online doctoral coursework to comprehensive examination—A mixed method study* (Publication No. 3558239) [Doctoral dissertation, Capella University]. ProQuest Dissertations Publishing.

*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. http://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.

*Vanhille, J., Gregory, B., & Corser, G. (2017). The Effects of mood on writing apprehension,

writing self-Efficacy, and writing performance. *Psi Chi Journal of Psychological Research, 22*(3), 220–230. https://doi.org/10.24839/2325-7342.JN22.3.220

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. http://doi.org/10.1080/01443410.2013.836157

*Webb, A. F., Vandiver, B. J., & Jeung, S. (2016). Does completing an enriched writing course improve writing self-efficacy of talented secondary students? *Gifted Child Quarterly, 60*(1), 47–62. https://doi.org/10.1177/0016986215605359

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

*Wijekumar, K., Graham, S., Harris, K. R., Lei, P.-W., Barkel, A., Aitken, A., Ray, A., & Houston, J. (2019). The roles of writing knowledge, motivation, strategic behaviors, and skills in predicting elementary students' persuasive writing from source material. *Reading and Writing, 32*(6), 1431–1457. https://doi.org/10.1007/s11145-018-9836-7

*Williams, H. M. (2012). *Third grade students' writing attitudes, self-efficacy beliefs, and achievement* (Publication No. 1529350) [Master's thesis, University of Maryland]. ProQuest Dissertations Publishing.

*Williams, J. D., & Takaku, S. (2011). Help seeking, self-efficacy, and writing performance among college students. *Journal of Writing Research, 3*(1), 1-18. https://doi.org/10.17239/jowr-2011.03.01.1

*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

*Woolford-Singh, A. W. (2004). *An examination of relationships of reading and writing self-efficacy beliefs, standardized placement test scores, and diverse community college students' perceptions of those relationship* (Publication No. 3148946) [Doctoral dissertation, Old Dominion University). ProQuest Dissertations Publishing.

*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. http://doi.org/10.1016/j.asw.2018.12.004

*Yamini, M., & Mandanizadeh, F. B. (2011). Relationship between EFL learners' self-efficacy and imposter phenomenon and their effects on learners' writing ability. *Advances in Language and Literacy Studies, 2*(1), 70-80. http://doi.org/10.7575/aiac.alls.v.2n.1p.70

*Yang, L., & Chen, R. (2012). The effects of positive self-evaluation on senior high school students' English writing performance and writing self-efficacy. *Chinese Journal of Applied Linguistics*, *35*(3). https://doi.org/10.1515/cjal-2012-0021

*Zabihi, R. (2018). The role of cognitive and affective factors in measure of L2 writing. *Written Communication*, 35(1), 32-57. http://doi.org/10.1177/0741088317735836

*Zellner, A. L. (2017). *The influence of peer feedback on writing achievement and individual writing self-efficacy* (Publication No. 10688847) [Doctoral dissertation, Michigan State

University]. ProQuest Dissertations Publishing.

*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.

*Zimmerman, B. J., & Bandura, A. (1994). Impact of Self-Regulatory Influences on Writing Course Attainment. *American Educational Research Journal, 31*(4), 845-862. http://doi.org/10.3102/00028312031004845

* Zimmerman, B. J., & Kitsantas, A. (1999). Acquiring writing revision skill: Shifting from process to outcome self-regulatory goals. *Journal of Educational Psychology, 91*(2), 241-250. http://doi.org/10.1037/0022-0663.91.2.241

*Zimmerman, B. J., & Kitsantas, A. (2002). Acquiring writing revision and self-regulatory skill through observation and emulation. *Journal of Educational Psychology, 94*(4), 660-668. http://doi.org/10.1037/0022-0663.94.4.660

APPENDIX A: CODEBOOKS

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)
    0. Pearson correlation
    1. Regression
6. Demographic (demo)
    0. Not included
    1. Included
7. Affective (cogni)
    0. Not included
    1. Included

8. Achieving (achi)
    0. Not included
    1. Included

Section D: Effect Size Statistics

1. Type of Effect Sizes (ES_Type)
    0. Pearson correlation coefficient *r*
    1. Standardized regression coefficient β
    2. 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)
    0. Journal article
    1. Dissertation/thesis
4. Publication Year (year)
5. Grade (grade)
    1. Elementary school
    2. Middle school
    3. High school
    4. College/University
6. L1 Writing

    0. L2 writing
    1. 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.

CHAPTER 3 [ARTICLE III]: COMPARING THREE APPROACHES TO HANDLING DEPENDENCE: A CASE STUDY OF A META-ANALYSIS OF WRITING SELF-EFFICACY AND WRITING PROFICIENCY

Ting Sun *Richard Lambert* *Chuang Wang*

In social and behavioral sciences, it is common to see inconsistencies in statistical results yielded by similar studies. These discrepancies may come from either random sampling errors or heterogeneity of results across studies. A meta-analysis provides an opportunity to make a scientific and systematic synthesis of these quantitative results, which would help in the estimation of overall average effect sizes in one domain (Moeyaert et al., 2017). According to Borenstein et al. (2009), effect sizes are indices measuring the effectiveness of a treatment (e.g., odds ratio), the effectiveness of an intervention (e.g., standardized mean differences), or the relationship between two variables (e.g., correlation coefficient). A meta-analysis cannot only estimate overall effect sizes but also make projections about the variation of these effect sizes across studies. This would give us information and implications about what may capture the true variation among the effect sizes.

Given the appealing features of meta-analyses, growing attention has been directed to the validity of meta-analysis results (Ahn et al., 2012). One of the challenges to validity arises when synthesizing primary studies that report multiple dependent effect sizes. This is a common issue in conducting meta-analytic studies in education, psychology, or social sciences (Scammacca et al., 2014). Cheung and Chan (2004) found that 31 out of 49 meta-analyses published in the *Journal of Applied Psychology* from 1991 to 2001 encountered dependent within-study effect sizes. Ahn et al. (2012) reviewed meta-analyses in education published from 2000 to 2010 and found approximately two-thirds of them were confronted with the dependence issue.

Tanner-Smith and Tipton (2014) classified the dependence in meta-analyses into two types: hierarchical dependence and correlated dependence. Hierarchical dependence occurs when one study reports effect sizes from multiple samples or the same research teams report effect sizes from multiple studies. Correlated dependence in a meta-analysis may result from multiple time points, multiple measures, or multiple comparisons within a study (Borenstein et al., 2009; Lambert, 1995). For example, Hashemnejad et al. (2014) reported effect sizes on the relationship between self-efficacy and writing outcomes at three different points in time at a one-week interval. Perin et al. (2017) reported eight effect sizes calculated from multiple writing outcome measures (e.g., analytic quality, holistic quality). Multiple comparisons within a study can be exemplified by Shintani and Aubrey's (2016) study, which reported effect sizes calculated from two treatment groups (i.e., synchronous corrective feedback and asynchronous corrective feedback) and one control group (i.e., no corrective feedback). These effect sizes are correlated since they are yielded with the same designs, conducted by the same researchers, measured by the same instruments, or calculated with the same samples (Moeyaert et al., 2017).

Traditional meta-analysis procedure (i.e., univariate meta-analysis) is not appropriate in synthesizing studies reporting dependent effects because the univariate meta-analysis assumes the independence of effect sizes (Borenstein et al., 2009). Since effect sizes within studies are correlated with each other, ignoring their dependence would have the risk of underestimating the standard error of the overall effect size estimates and committing a Type I error (Borenstein et al., 2009; Scammacca et al., 2014). In addition, the estimates of the overall average effect size would be biased towards the studies having more effects because more weights are given to them in the analysis (Scammacca et al., 2014; Van den Noortgate et al., 2013). Therefore, it is significant to evaluate approaches resolving the dependence issue in conducting meta-analyses.

Several methods were suggested by literature to resolve the dependence issues. Multivariate meta-analysis was suggested as the most accurate procedure in handling the dependence (Kalaian & Raudenbush,1996). However, this method requires the correlation of the dependent effect sizes to be known to model the covariance matrix. Since this information is rarely reported by primary studies, implementing the multivariate meta-analysis is not feasible in practice. Performing separate meta-analyses, also known as shifting-unit-of-analysis, is believed to be another alternative to handling the dependence issue (Cooper, 1998; Cooper, 2010). Instead of conducting one meta-analysis, this procedure accounts for the dependence by performing separate meta-analyses by multiple measures, multiple time-points, or multiple comparisons (Borenstein et al., 2009). However, performing separate meta-analyses is not feasible in implementation unless measures or comparisons are consistent across primary studies. Choosing one effect size per study and then conducting a univariate meta-analysis can be another approach in handling the dependence issue (Chambers, 2004). The selection of the effect size can be based on one of the following rationales: (a) select one effect size randomly; (b) select the largest effect size; (c) select the effect size that is aligned with the focus of a meta-analysis; (d) select the effect size calculated by the measure with better validity and reliability statistics (Cheung, 2019; Scammacca et al., 2014). The procedure has appreciable limitations in that it reduces the statistical power by removing other effect sizes in one study (Cheung, 2014). In addition, it restricts the opportunity to examine the effect of within-study moderators on the variances of effect sizes.

The methods that are popular in practice and feasible in implementation include: averaging effect sizes within studies (Borenstein et al., 2009), using robust variance estimation (RVE; Hedges et al., 2010), and conducting a three-level meta-analysis (3LM; Cheung, 2014;

Van den Noortgate et al., 2013). There were studies evaluating the performance of these methods with simulated effect sizes (e.g., Moeyaert et al., 2017). However, our knowledge of the efficacy of the three methods with real data is limited. Moreover, most studies dealt with effect sizes of standardized mean differences (i.e., Cohen's *d* or Hedge's *g*) extracted from experimental or quasi-experimental studies. So far, no study has been conducted to examine different methods in handling dependent effect sizes based on correlation. To bridge the gap in the literature, the current study aims to compare the methods of the averaging method, RVE, and 3LM in terms of dealing with dependent effect sizes of correlation with real data.

The Three Methods

This section provides a review of the three procedures dealing with synthesizing dependent effect sizes, averaging within-study effect sizes, 3LM, and RVE.

Averaging Within-Study Effect Sizes

The most straightforward approach to dealing with dependence is to compute average or weighted average effect sizes within studies before conducting a univariate meta-analysis. This method proceeds on the assumption that observed effect sizes in one study share the common effect in the population, which is reasonable but not the case in the strict sense. Due to its ease in implementation and conceptual understanding, the averaging method was commonly used in meta-analytic studies. In Ahn et al.'s (2012) review of meta-analyses in education, 18 out of the 28 meta-analyses adopted the averaging method to synthesize dependent effect sizes.

The combined effect and the variance of the combined effect in one study can be expressed as follows. Suppose there are m dependent effect sizes in one study. Notations are borrowed from Borenstein et al. (2009).

$$\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i, \quad (1)$$

$$V_{\bar{Y}} = (\frac{1}{m})^2 (\sum_{i=1}^{m} V_i + \sum_{i \neq j} (r_{ij} \sqrt{V_i} \sqrt{V_j})) \quad (2)$$

where Y_i is the i_{th} dependent effect sizes in one study, $\bar{Y}$ is the average effect size of the study, V_i is the variance for Y_i, $V_{\bar{Y}}$ is the variance of the combined effect, and r_{ij} is the correlation between the two effect sizes, Y_i and Y_j.

According to the two formulas, while it is possible to calculate the mean effect sizes within studies, the variance of combined effects requires the correlations of the dependent effect sizes, which are not always available in primary studies. There are several alternatives recommended by Borenstein et al. (2009) when the correlations between dependent effect sizes are unknown: (a) identifying correlations based on empirical evidence or theoretical rationale; (b) assuming their correlation to be zero and treating them independently; (c) assuming their correlation to be one; (d) using the average of sampling variance of each effect size. Assuming their correlation to be zero would result in spuriously smaller variance, narrower confidence intervals, and a higher probability of committing a Type I error (Borenstein et al., 2009; Moeyaert et al., 2017). Conversely, assuming their correlation to be one has the risk of overestimating the variance of combined effects, obtaining larger confidence intervals, and having a higher probability of committing a Type II error (Borenstein et al., 2009; Scammacca et al., 2014). Most meta-analysts use the average of variances of dependent effect sizes to estimate the variance of the combined effect (Hedges et al., 2010). Therefore, the averaging procedure is reliable only when the dependent effect sizes are highly correlated. However, if dependent effect sizes are not homogenous within studies and the goal of meta-analysts is to detect this heterogeneity, the averaging procedure would obscure this rich information of within-study covariates (Marín-Martínez & Sánchez-Meca, 1999; Scammacca et al., 2014).

Three-Level Meta-Analysis

Multilevel model is an optimal method in dealing with nested data since it can account for dependence and variances in multiple levels (e.g., student level, class level, and school level). This method has an application in meta-analytic data, which has the same hierarchical structure (Raudenbush & Bryk, 2002). Literature suggested 3LM as an alternative to handling dependent effect sizes (Van den Noortgate et al., 2013). When the dependence comes from multiple outcomes, the first level is the participants' level, which models sampling variation for effect sizes for one outcome. The second level is the outcome level, varying within the outcome-specific errors. The third level is the study level, varying within the study-specific errors (Cheung, 2014; Konstantopoulos, 2011). The three-level meta-analysis can be modeled as

$$\text{Level 1: } T_{iok} = \beta_{ok} + \varepsilon_{iok,} \tag{3}$$

$$\text{Level 2: } \beta_{ok} = \theta_k + \phi_{ok,} \tag{4}$$

$$\text{Level 3: } \theta_k = \gamma + \eta_{k,} \tag{5}$$

where T_{iok} is the estimate of effect size *i* for outcome o in study *k*, β_{ok} is the population parameter of effect size for outcome o in study *k*, ε_{iok} is the sampling error in level 1 with the mean of zero and variance of σ^2_{iok} (i.e., sampling variance), θ_k is the parameter for effect size in study *k*, and ϕ_{ok} is residual in outcome level with the mean of zero and variance of $\sigma^2_{\phi_{ok}}$ (i.e., between-outcome variance), γ is the parameter for the average effect size in population, η_k is the study-level residual with the mean of zero and the variance of $\sigma^2_{\eta_k}$ (i.e., between-study variance).

Fernández-Castilla et al. (2020) systematically reviewed 178 meta-analyses published from 2002 to 2018 that used multilevel models and found that the majority of them employed a three-level model. Geeraert et al. (2004) examined the effectiveness of early prevention programs on child abuse and neglect with 40 evaluation studies and performed a 3LM to

synthesize dependent effect sizes from multiple outcomes within studies. The three levels account for the covariance of dependent effect sizes within studies by decomposing the total variance into between-effect variance and between-study variance. The 3LM has the advantage of exploring moderators for each level and the flexibility of the number of levels added to the hierarchical structure (Moeyaert et al., 2017; Van den Noortgate et al., 2013). However, this method is better in handling hierarchical dependence than correlated dependence among effect sizes (Hedges et al., 2010).

Robust Variance Estimation

Accurate estimation of overall effect size necessitates covariances between dependent effect sizes, and this information is not always available in primary studies. Hedges et al. (2010) solved the problem by proposing the RVE method. RVE does not require the underlying covariances to be known to calculate the variance of the average effect size for one study. Instead, it obtains a crude estimate of the covariances using the cross-product of residuals within one study. The robust variance estimator can be expressed as

$$V^R = \left(\sum_{j=1}^{m} X_j' W_j X_j\right)^{-1} \left(\sum_{j=1}^{m} X_j' W_j e_j e_j' W_j X_j\right) \left(\sum_{j=1}^{m} X_j' W_j X_j\right)^{-1}, \quad (6)$$

where e_j is the estimated residual vector for study j, m is the number of studies, W_j is the weight for study j. Hedges et al. (2010) found that when the number of studies goes infinite, $e_j e_j'$ is a good estimator of the true covariance matrix. The most efficient weights are inverse-variance weights, which can be expressed as

$$W_{ij} = \frac{1}{(V_{.j} + \tau^2)(1 + (k_j - 1)\rho)}, \quad (7)$$

where $V_{.j}$ is the average of the variance in study j, τ^2 is the between variance component, and ρ is the correlation between dependent effect sizes. Although ρ is unknown, a sensitivity approach

can be taken to see if the results are affected by the choice of different ρ values (between 0 and 1).

The RVE approach has the following advantages compared to other methods in handling the dependence issue. First, RVE is advantageous to 3LM in that it is applicable to both the hierarchical dependence and correlated dependence (Tanner-Smith & Tipton, 2014). Second, it does not require the assumption of a normal distribution of effect sizes (Moeyaert et al., 2017). Third, RVE results are invariant regardless of the choice of weights or meta-analysis models (Hedges et al., 2010). Although RVE has the advantage of accurately estimating mean effects and variance without knowing correlations between dependent effect sizes within one study, it has stringent assumptions for accurate estimation. The implementation of RVE has a requirement on the minimum number of primary studies in a meta-analysis (Tanner-Smith & Tipton, 2014). It requires ten studies for the estimation of the mean effect size and 40 studies for a meta-regression coefficient (Hedges et al., 2010; Tipton, 2013). It also requires that the difference between the number of primary studies and the number of covariates be more than four (Tanner-Smith et al., 2016).

Literature in Comparing Different Methods

Moeyaert et al. (2017) compared three procedures in handling multiple outcomes: averaging effect sizes, RVE, and multilevel meta-analysis (MLM) with 432,000 simulated effect sizes under 216 conditions. The indices of overall effect estimates, standard error estimates, and variance estimates were used to evaluate the three procedures. The study found that while the three procedures functioned equally well in the estimation of overall fixed effects, RVE yielded unbiased estimation of variance under all conditions, whereas MLM underestimated the variance

slightly when the number of primary studies was small (e.g., 25 studies). In the estimation of unbiased standard errors, both RVE and MLM were recommended.

Scammacca et al. (2014) examined procedures handling dependence resulted from multiple measures and multiple group comparisons with real data in a meta-analysis on the effect of reading interventions on students with learning difficulties. Their procedures included choosing the highest effect size, randomly sampling an effect size, choosing a research-question-guided effect size, conducting meta-analyses separately, treating effect sizes independently, averaging effect sizes, using RVE, and employing a 3LM. They concluded that the averaging method, RVE, and 3LM were the recommended procedures when the overall effect size was aligned with the research purpose of a meta-analytic study. In addition, RVE and 3LM were preferred when the number of primary studies was large.

Most of the methodological studies examining methods dealing with dependent effect sizes used simulated data (e.g., Moeyaert et al., 2017). Although simulated data has the advantage of having a criterion against which different method can be compared in terms of estimation accuracy, real data "can better emulate the types and nature of dependence that typically exist in studies that education researchers struggle to meta-analyze" (Scammacca et al., 2014, p.336). In addition, previous studies used the effect size of standardized mean difference, and their results may not be applied to effect sizes based on correlation. Therefore, a methodological study is warranted to examine the three procedures in synthesizing dependent effect sizes based on correlation with real data.

The present study aims to compare three approaches to dealing with dependent effect sizes. Specifically, two research questions are presented as follows.

1. How do the averaging method, RVE, and 3LM differ in the estimation of overall mean, standard error, and heterogeneity of the true effect?
2. Do these estimates differ by the number of primary studies in a meta-analysis and the number of dependent effect sizes per study?

Methodology

Data

Data were from the meta-regression study examining the relationship between writing self-efficacy and writing achievement (Sun et al., in press). In the meta-analysis, primary studies were located using both electronic searches and manual searches. Seven electronic databases were targeted: Academic Search Complete, Education Research Complete, Eric, ProQuest Dissertations & Theses Global, PsycARTICLES, PsycINFO, and Web of Science. Studies were included if they were: (a) scholarly (peer-reviewed) articles or theses/dissertations; (b) published or written in the time frame of 1977-2019; (c) written in English; (d) examining students' writing self-efficacy; (e) quantitative empirical studies; (f) studies that reported or had computable effect sizes. Please refer to the study (Sun et al., in press) for detailed information for literature searches, screening, and coding. This final sample consisted of 565 effect sizes nested within 76 primary studies.

Effect Size Calculation

The effect size used in the current study is Pearson correlation coefficient, *r*. Studies that report standardized beta coefficient (β) were also included by converting β to *r* metric using the two-parameter least-square equation.

$$r = .98\,\beta + .05\lambda, \tag{8}$$

where λ is an indicator variable that is 1 when β is nonnegative and is 0 when β is negative (Peterson & Brown, 2005). Then *r* was transformed to Fisher's z metric using the formula $z = 0.5 \times \ln(\frac{1+r}{1-r})$ for the normalization of the effect size distribution (Borenstein et al., 2009). The final sample consisted of 76 primary studies and 565 effect sizes, with an average of 7 effect sizes per study. The frequency of the number of effect sizes per study was presented in Table 3-1.

Subsets of Data Generation

Different conditions were created to examine if the performance of the three methods differs by the number of primary studies and the number of effect sizes per study. Three levels of the number of effect sizes per study (*K*) were manipulated ($K \geq 2$, $K \geq 4$, and $K \geq 6$), which yielded three levels of the number of primary studies (*N*) of 61, 42, and 26. The three values can be used to represent large-, medium-, and small-scale meta-analyses, respectively. Based on the three levels of *K* and the three levels of *N*, six datasets were created for analysis. The design of the conditions was presented in Table 3-2, where 61-2 represents the condition of 61 primary studies with 2 effect sizes or more for each study. For the conditions 42-2, 26-2, and 26-4, stratified random sampling was performed. Stratification was based on the number of effect sizes per study. The condition of all effect sizes was also used, so there were seven conditions in total.

Table 3-1

Frequency of the Number of Effect Sizes per Study

# of Effect Sizes	Frequency	Percent
1	15	0.20
2	13	0.17
3	6	0.08
4	12	0.16
5	4	0.05
6	11	0.14
9	1	0.01
10	2	0.03
11	1	0.01
12	2	0.03
13	1	0.01
14	1	0.01
18	2	0.03
27	1	0.01
28	1	0.01
42	1	0.01
56	1	0.01
92	1	0.01

Table 3-2

Six Conditions by Number of Primary Studies and Number of Effect Sizes per Study

	# Studies 61	# Studies 42	# Studies 26
# of ES ≥ 2	Condition 61-2	Condition 42-2	Condition 26-2
# of ES ≥ 4	--	Condition 42-4	Condition 26-4
# of ES ≥ 6	--	--	Condition 26-6

Note. ES = Effect Size.

Data Analysis

A random-effects model was chosen for analysis because the assumption that true effects may vary across studies is more plausible. A univariate meta-analysis (assuming all the effect sizes independent) and the two-level meta-analysis (2LM) were also performed as the baseline methods against which the three methods were compared. Each of the five methods (i.e., the univariate method, 2LM, the averaging method, 3LM, and RVE) was applied to the seven datasets, respectively. In the averaging method, the variance of the combined effect size in each study was calculated by averaging the variances of dependent effect sizes. For RVE, five levels of correlations of within-study dependent effect sizes (i.e., ρ = 0, ρ = .1, ρ = .3, ρ = .5, and ρ = 1) were employed to see if the results were robust to different values of ρ. 2LM was approached using SPSS and the other methods were conducted using computer software R with different packages. The univariate and the averaging method were applied using R package *metafor* (Viechtbauer, 2010). 3LM and RVE were implemented by using R packages *metaSEM* (Cheung, 2019) and *robumeta* (Fisher & Tipton, 2014), respectively.

Evaluation Indices

Evaluation indices include the mean effect size, *Z* and its *p*-value, the standard error of the mean effect size, confidence intervals, and three heterogeneity statistics (Q, T^2, and I^2). Z

and its *p*-values test the statistical significance of the mean effect size (i.e., whether it is statistically significantly different from zero). The statistics of standard error and confidence intervals measure the precision of the effect size estimate. Heterogeneity statistics quantify the extent to which true effect size varies across studies. Q statistic and its *p*-value test the statistical significance of heterogeneity (i.e., whether the true effect sizes vary across studies). Q can be obtained by computing the weighted sum of squares of the deviation of effect size in each study from the overall average effect size. T^2 is the estimate of the magnitude of the variance of the true effect sizes across studies (τ^2). I^2 refers to the proportion of variation due to heterogeneity, which can be calculated by dividing the between-study variance by the total variance. The present study takes Higgins et al.'s (2003) benchmark, and I^2 was interpreted as small when $I^2 = 25\%$, moderate when $I^2 = 50\%$, and high when $I^2 = 75\%$.

Results

Condition of All Effect Sizes

The effect size and heterogeneity statistics for all the effect sizes were presented in Table 3-3. The overall effect size and standard error estimates generated by the averaging method and RVE were very similar. 3LM and 2LM yielded a slightly smaller effect size and standard error estimates, whereas the univariate method produced a much smaller effect (0.2631) and standard error (0.0087) compared with the other four methods. As for the heterogeneity statistic, all the methods had statistically significant Q-values, indicating that the true effect sizes were heterogeneous across studies regardless of methods employed. All the I^2 statistics suggested that the majority of observed variance was due to between-study variation. 3LM yielded the largest estimate of true effect sizes variance (T^2 = 0.0596), and the univariate method yielded the smallest (T^2 = 0.0347). In addition, the discrepancies in the variance estimates between the

averaging method (T^2 = 0.0579), 3LM method, and RVE (T^2 = 0.0518) were negligible. The sensitivity test for RVE indicated that the results did not vary based on the choice of ρ value.

Condition of 61-2

See Table 3-4 for the results for the condition of having 61 primary studies with at least 2 effect sizes in each study. RVE generated the highest value of overall effect size (0.03017) and standard error (0.0254), followed by the averaging method and the 3LM. Overall, the three methods produced very similar values in the two statistics. Similar to the condition of all effect sizes, the univariate method had the smallest estimate of overall average effect size (0.2600) and standard error (0.0085). 3LM (T^2 = 0.0430) and RVE (T^2 = 0.0399) yielded the largest true variance estimates across studies, whereas the averaging method produced the smallest (T^2 = 0.0299). RVE results were robust to the choice of ρ value.

Condition of 42-2

The effect size and heterogeneity statistics for the condition of having 42 primary studies with at least 2 effect sizes per study were shown in Table 3-5. Like the condition of 61-2, the averaging method, 3LM, and RVE produced very similar combined effect size estimates and standard error estimates, with RVE having a slightly larger value than the other two methods. 2LM had a slightly smaller effect estimate compared the three (i.e., the averaging method, 3LM, and RVE). The univariate method was the lowest in the estimates of the combined effect (0.2738) and standard error (0.0108). The univariate method (T^2 = 0.341) resulted in a similar true variance estimate as was obtained by RVE (T^2 = 0.331) or the 3LM (T^2 = 0.389) procedures. The averaging method resulted in the smallest true variance estimate (T^2 = 0.0221).

Condition of 26-2

The results for the condition of having 26 primary studies with at least 2 effect sizes per study were presented in Table 3-6. RVE resulted in the highest values of effect size estimate (0.3175) and standard error (0.0341), followed by the averaging method. The univariate method produced the smallest value in both the overall effect size (0.2924) and standard error estimate (0.0158). All the overall effect sizes were statistically significantly different from zero, provided evidence of the statistical significance of the effect size. The heterogeneity statistics showed that the 3LM resulted in the highest value of the true variance estimate (T^2 = 0.0363), followed by RVE (T^2 = 0.0326). The averaging method produced a much smaller variance estimate (T^2 = 0.0227) compared to the other three methods.

Condition of 42-4

For the condition of having 42 primary studies with at least 4 effect sizes per study (see Table 3-7), the averaging method had the largest effect size estimate (0.2747), whereas the RVE had the largest standard error estimate (0.0258). In general, the averaging method, 3LM, 2LM and RVE estimated somewhat similar values of the summary effect and standard error. The univariate method estimated the smallest values of summary effect and standard error. As for the variance of true effect sizes across studies, the 3LM (T^2 = 0.0343) and RVE (T^2 = 0.0337) produced results with negligible differences. However, the averaging method had a much smaller variance estimate (T^2 = 0.0194). RVE results were also robust to the choice of ρ value.

Condition of 26-4

The results of effect size and heterogeneity estimates for the condition having 26 primary studies with at least 4 effect sizes per study were shown in Table 3-8. All the methods resulted in statistically significant effect sizes (p < .001), with the 3LM having the largest (0.2939) and the

univariate method having the smallest value (0.2743). Like the aforementioned conditions, the averaging method, 3LM, RVE, and 2LM yielded somewhat similar overall effect and standard error estimates. As for the heterogeneity statistics, all the methods generated statistically significant Q-values, indicating that the true effects varied across studies, and more than 60% of the observed variance was between-study variance (I^2 ranging from 62.90% to 93.37%). The averaging method had the smallest variance estimate (0.0264) as compared with the other four methods.

Condition of 26-6

The results for the condition of having 26 primary studies with at least 6 effect sizes per study were presented in Table 3-9. All the Z values were statistically significant ($p < .001$), suggesting that the effect size estimates were statistically significantly different from zero, regardless of the methods employed. There were small differences in the estimates of overall effect size and standard error between the averaging method, 3LM, and RVE, whereas the univariate method generated much smaller values in the two estimates. Q-values all indicated that the null hypothesis that the true effect sizes were homogeneous was rejected, and a substantial observed variation was real. RVE resulted in the largest variance (0.0350), which was similar to the results produced by 3LM (0.0301). However, the averaging method produced a much smaller estimate of the variance of true effect sizes across studies (0.0164). The sensitivity test for RVE indicated that the results did not vary based on the choice of ρ value.

Table 3-3

Effect Size and Heterogeneity Statistics for the Condition of All the Effect Sizes

		Effect size and 95% confidence interval							Heterogeneity	
		k	ES (z)	ES (r)	SE	95% CI	Z	p	p of Q	I^2	T^2
UVM		565	0.2631	0.2572	0.0087	[0.2460, 0.2802]	30.1467	< .001	< .001	89.34%	0.0347
AVM		76	0.3179	0.3076	0.0297	[0.2598, 0.3760]	10.7208	< .001	< .001	92.36%	0.0579
2LM		76	0.3110	0.3013	0.0274	[0.2573, 0.3647]	11.332	< .001	< .001	66.95%	0.0487
3LM		76	0.3128	0.3030	0.0267	[0.2604, 0.3652]	11.701	< .001	< .001	21.70%/71.82%	0.0138/0.0458
RVE	ρ = 0	76	0.3183	0.3080	0.0299	[0.2587, 0.3778]	10.6511	< .001	< .001	91.08%	0.0518
	ρ = .1	76	0.3183	0.3080	0.0299	[0.2587, 0.3778]	10.6511	< .001	< .001	91.08%	0.0518
	ρ = .3	76	0.3183	0.3080	0.0299	[0.2587, 0.3778]	10.6511	< .001	< .001	91.10%	0.0518
	ρ = .5	76	0.3183	0.3080	0.0299	[0.2587, 0.3778]	10.6511	< .001	< .001	91.12%	0.0518
	ρ = 1	76	0.3183	0.3080	0.0299	[0.2587, 0.3778]	10.6511	< .001	< .001	91.16%	0.0518

Note. ρ is the correlation of within-study effect size; ES =effect size; SE =standard error; CI = confidence intervals; UNM = univariate method; AVM = averaging method; 2LM = two-level meta-analysis; 3LM = three-level meta-analysis; RVE = robust variance estimation.

Table 3-4

Effect Size and Heterogeneity Statistics for the Condition of 61-2

		Effect size and 95% confidence interval							Heterogeneity	
		k	ES (*z*)	ES (r)	SE	95% CI	*Z*	*p*	*p* of Q	I^2	T^2
UVM		550	0.2600	0.2543	0.0085	[0.2434, 0.2765]	30.7570	< .001	< .001	88.38%	0.0311
AVM		61	0.3008	0.2920	0.0249	[0.2520, 0.3496]	12.0785	< .001	< .001	87.06%	0.0299
2LM		61	0.2973	0.2888	0.0243	[0.3449, 0.2497]	12.244	< .001	< .001	56.11%	0.0304
3LM		61	0.2983	0.2898	0.0238	[0.2516, 0.3450]	12.5199	< .001	< .001	29.11%/62.19%	0.0137/0.0293
RVE	ρ = 0	61	0.3017	0.2929	0.0254	[0.2508, 0.3527]	11.8605	< .001	< .001	89.37%	0.0399
	ρ = .1	61	0.3017	0.2929	0.0254	[0.2508, 0.3527]	11.8605	< .001	< .001	89.38%	0.0399
	ρ = .3	61	0.3017	0.2929	0.0254	[0.2508, 0.3527]	11.8605	< .001	< .001	89.41%	0.0399
	ρ = .5	61	0.3017	0.2929	0.0254	[0.2508, 0.3527]	11.8605	< .001	< .001	89.44%	0.0400
	ρ = 1	61	0.3017	0.2929	0.0254	[0.2508, 0.3527]	11.8605	< .001	< .001	89.51%	0.0400

Note. ρ is the correlation of within-study effect size; ES =effect size; SE =standard error; CI = confidence intervals; UNM = univariate method; AVM = averaging method; 2LM = two-level meta-analysis; 3LM = three-level meta-analysis; RVE = robust variance estimation.

Table 3-5

Effect Size and Heterogeneity Statistics for the Condition of 42-2

		Effect size and 95% confidence interval							Heterogeneity	
		k	ES (*z*)	ES (*r*)	SE	95% CI	*Z*	*p*	*p* of Q	I^2	T^2
UVM		359	0.2738	0.2672	0.0108	[0.2527, 0.2950]	25.4016	< .001	< .001	90.28%	0.0341
AVM		42	0.3096	0.3001	0.0266	[0.2575, 0.3618]	11.6383	< .001	< .001	82.86%	0.0221
2LM		42	0.3059	0.2967	0.0269	[0.3586, 0.2532]	11.366	< .001	< .001	52.37%	0.2520
3LM		42	0.3064	0.2972	0.0263	[0.2549, 0.3578]	11.6658	< .001	< .001	36.40%/54.95%	0.0155/0.0234
RVE	ρ = 0	42	0.3109	0.3013	0.0271	[0.2562, 0.3657]	11.4836	< .001	< .001	86.92%	0.0331
	ρ = .1	42	0.3109	0.3013	0.0271	[0.2562, 0.3657]	11.4836	< .001	< .001	86.94%	0.0331
	ρ = .3	42	0.3109	0.3013	0.0271	[0.2562, 0.3657]	11.4836	< .001	< .001	86.99%	0.0331
	ρ = .5	42	0.3109	0.3013	0.0271	[0.2562, 0.3657]	11.4836	< .001	< .001	87.04%	0.0331
	ρ = 1	42	0.3109	0.3013	0.0271	[0.2562, 0.3657]	11.4836	< .001	< .001	87.17%	0.0331

Note. ρ is the correlation of within-study effect size; ES =effect size; SE =standard error; CI = confidence intervals; UNM = univariate method; AVM = averaging method: 2LM = two-level meta-analysis: 3LM = three-level meta-analysis: RVE = robust variance estimation.

Table 3-6

Effect Size and Heterogeneity Statistics for the Condition of 26-2

		Effect size and 95% confidence interval							Heterogeneity	
		k	ES (z)	ES (r)	SE	95% CI	Z	p	p of Q	I^2	T^2
UVM		160	0.2924	0.2843	0.0158	[0.2615, 0.3234]	18.5056	< .001	< .001	91.44%	0.0334
AVM		26	0.3154	0.3053	0.0336	[0.2495, 0.3813]	9.3845	< .001	< .001	84.72%	0.0227
2LM		26	0.3126	0.3028	0.0335	[0.3783, 0.2469]	9.328	< .001	< .001	60.05%	0.0253
3LM		26	0.3124	0.3026	0.0331	[0.2475, 0.3773]	9.4401	< .001	< .001	32.18%/59.88%	0.0127/0.0236
RVE	ρ = 0	26	0.3175	0.3072	0.0341	[0.2472, 0.3878]	9.3203	< .001	< .001	87.90%	0.0326
	ρ = .1	26	0.3175	0.3072	0.0341	[0.2472, 0.3878]	9.3203	< .001	< .001	87.93%	0.0327
	ρ = .3	26	0.3175	0.3072	0.0341	[0.2472, 0.3878]	9.3203	< .001	< .001	88.01%	0.0327
	ρ = .5	26	0.3175	0.3072	0.0341	[0.2472, 0.3878]	9.3203	< .001	< .001	88.09%	0.0327
	ρ = 1	26	0.3175	0.3072	0.0341	[0.2472, 0.3878]	9.3203	< .001	< .001	88.27%	0.0328

Note. ρ is the correlation of within-study effect size; ES =effect size; SE =standard error; CI = confidence intervals; UNM = univariate method; AVM = averaging method; 2LM = two-level meta-analysis; 3LM = three-level meta-analysis; RVE = robust variance estimation.

Table 3-7

Effect Size and Heterogeneity Statistics for the Condition of 42-4

		Effect size and 95% confidence interval							Heterogeneity	
		k	ES (z)	ES (r)	SE	95%CI	Z	p	p of Q	I^2	T^2
UVM		506	0.2516	0.2464	0.0083	[0.2352, 0.2679]	30.2345	< .001	< .001	86.99%	0.0271
AVM		42	0.2747	0.2680	0.0253	[0.2252, 0.3242]	10.8762	< .001	< .001	82.06%	0.0194
2LM		42	0.2723	0.2658	0.0251	[0.3215, 0.2231]	10.868	< .001	< .001	51.51%	0.0231
3LM		42	0.2738	0.2672	0.0243	[0.2260, 0.3216]	11.2294	< .001	< .001	33.09%/56.36%	0.0127/0.0216
RVE	ρ = 0	42	0.2744	0.2677	0.0258	[0.2223, 0.3265]	10.6	< .001	< .001	87.90%	0.0337
	ρ = .1	42	0.2744	0.2677	0.0258	[0.2223, 0.3265]	10.6	< .001	< .001	87.93%	0.0337
	ρ = .3	42	0.2744	0.2677	0.0258	[0.2223, 0.3265]	10.6	< .001	< .001	87.98%	0.0338
	ρ = .5	42	0.2744	0.2677	0.0258	[0.2223, 0.3265]	10.6	< .001	< .001	88.03%	0.0338
	ρ = 1	42	0.2744	0.2677	0.0258	[0.2223, 0.3265]	10.6	< .001	< .001	88.15%	0.0338

Note. ρ is the correlation of within-study effect size; ES =effect size; SE =standard error; CI = confidence intervals; UNM = univariate method; AVM = averaging method; 2LM = two-level meta-analysis; 3LM = three-level meta-analysis; RVE = robust variance estimation.

Table 3-8

Effect Size and Heterogeneity Statistics for the Condition of 26-4

		Effect size and 95% confidence interval							Heterogeneity	
		k	ES (z)	ES (r)	SE	95%CI	Z	p	p of Q	I^2	T^2
UVM		180	0.2743	0.2676	0.0156	[0.2438, 0.3048]	17.6114	< .001	<.001	92.45%	0.0374
AVM		26	0.2934	0.2853	0.0359	[0.2229, 0.3638]	8.1656	< .001	<.001	87.02%	0.0264
2LM		26	0.2908	0.2829	0.0372	[0.3637, 0.2189]	7.826	< .001	< .001	62.90%	0.0326
3LM		26	0.2918	0.2838	0.0357	[0.2218, 0.3618]	8.1719	< .001	<.001	29.70%/63.67%	0.0137/0.0293
RVE	ρ = 0	26	0.2939	0.2857	0.0364	[0.2189, 0.3689]	8.0799	< .001	< .001	89.87%	0.0380
	ρ = .1	26	0.2939	0.2857	0.0364	[0.2189, 0.3689]	8.0799	< .001	< .001	89.90%	0.0381
	ρ = .3	26	0.2939	0.2857	0.0364	[0.2189, 0.3689]	8.0799	< .001	< .001	89.97%	0.0381
	ρ = .5	26	0.2939	0.2857	0.0364	[0.2189, 0.3689]	8.0799	< .001	< .001	90.38%	0.0381
	ρ = 1	26	0.2939	0.2857	0.0364	[0.2189, 0.3689]	8.0799	< .001	< .001	90.21%	0.0382

Note. ρ is the correlation of within-study effect size; ES =effect size; SE =standard error; CI = confidence intervals; UNM = univariate method; AVM = averaging method; 2LM = two-level meta-analysis; 3LM = three-level meta-analysis; RVE = robust variance estimation.

Table 3-9

Effect Size and Heterogeneity Statistics for the Condition of 26-6

		Effect size and 95% confidence interval							Heterogeneity	
		k	ES (z)	ES (r)	SE	95%CI	Z	p	p of Q	I^2	T^2
UVM		438	0.2484	0.2434	0.0088	[0.2313, 0.2656]	28.3838	< .001	< .001	86.97%	0.0258
AVM		26	0.2799	0.2728	0.0297	[0.2217, 0.3382]	9.4228	< .001	< .001	81.83%	0.0164
2LM		26	0.2688	0.2625	0.0277	[0.3230, 0.2145]	9.688	< .001	< .001	44.40%	0.0177
3LM		26	0.2734	0.2668	0.0271	[0.2203, 0.3265]	10.0945	< .001	< .001	39.23%/40.39%	0.0133/0.0168
RVE	ρ = 0	26	0.2769	0.2700	0.0300	[0.2150, 0.3387]	9.2349	< .001	< .001	89.72%	0.0350
	ρ = .1	26	0.2769	0.2700	0.0300	[0.2150, 0.3387]	9.2349	< .001	< .001	89.46%	0.0350
	ρ = .3	26	0.2769	0.2700	0.0300	[0.2150, 0.3387]	9.2349	< .001	< .001	89.50%	0.0351
	ρ = .5	26	0.2769	0.2700	0.0300	[0.2150, 0.3387]	9.2349	< .001	< .001	89.60%	0.0351
	ρ = 1	26	0.2769	0.2700	0.0300	[0.2150, 0.3387]	9.2349	< .001	< .001	89.79%	0.0352

Note. ρ is the correlation of within-study effect size; ES =effect size; SE =standard error; CI = confidence intervals; UNM = univariate method; AVM = averaging method; 2LM = two-level meta-analysis; 3LM = three-level meta-analysis; RVE = robust variance estimation.

Discussion

Similarities between the Three Methods

There are three similarities between the five methods across all the conditions. First, all the methods resulted in statistically significant *Z*-values, provided evidence of the statistical significance of the overall effect size. This finding suggested a statistically significant relationship between writing self-efficacy and writing achievement regardless of methods employed. This positive and medium correlation was echoed by a large body of studies (e.g., Sun & Wang, 2020). Second, all the methods produced significant Q-values across different conditions, providing evidence that the true effect sizes were heterogeneous across studies. This confirmed the legitimacy of the use of a random-effects model. The significant Q-values only suggested the untenability of the null hypothesis that all the studies share the common true effect size rather than the absolute amount of the true variation (Borenstein et al., 2009). Therefore, this finding can be interpreted as either a large amount of dispersion of the true effect sizes across studies or a little amount of dispersion with a precise estimation of effect sizes. When the effect size is Fisher's z, the precision of estimation depends on the inverse of the sample size ($\frac{1}{N-3}$) (Borenstein et al., 2009). The current study has a mean sample size of 245.03 (*SD* = 247.77). The large sample size may account for the significant Q-values estimated. Third, except for 2LM, all the other four methods produced large I^2 statistics (ranging from 79.62% to 92.45%). According to Higgins et al.'s (2003) benchmark, it can be interpreted as a large value (if larger than 75%), suggesting that a substantial proportion of the observed variation reflected real heterogeneity or the variation of true effect sizes across studies. The large between-study variance suggested the necessity of conducting moderator analyses in the future to examine what factors at the study level may account for the heterogeneity.

Differences between the Three Methods

Overall Average Effect Size

RVE resulted in the highest values of overall average effect size estimates, followed by the averaging method. This pattern was consistent across the seven conditions except for the condition of 42-4 and 26-6, in which the averaging method had slightly larger effect size estimates than those estimated by RVE. Overall, there were negligible differences in the overall effect size estimates between the averaging method, 3LM, and RVE. 2LM had slightly smaller effect size estimates than the three. However, the univariate method resulted in much smaller effects as compared with the other three methods across the seven conditions. This result was consistent with Moeyaert et al.'s (2017) study, which concluded that the averaging method, RVE, and MLM all produced unbiased effect size estimates. The researchers further noted that this was not impacted by the number of primary studies or the number of effect sizes per study. This finding was also partially consistent with Scammacca's (2014) study, which found that RVE and the averaging produced similar effect size estimates. Moeyaert et al. (2017) and Scammacca's (2014) studies examined the effect sizes of standardized mean differences. The current study extends their findings to the effect size of correlation (Fisher' z).

Standard Error

There were little differences between the averaging method, 3LM, 2LM and RVE in terms of standard error estimates, although RVE resulted in slightly larger values while the averaging method resulted in slightly smaller values. The univariate method produced much smaller results of standard errors. The performance of the five methods was consistent across the seven conditions. Moeyaert et al. (2017) found that RVE, MLM, and the averaging method produced unbiased standard errors in most conditions except for the condition of the number of

effect sizes of 4 and the correlation between within-study effect sizes of 0, in which the averaging method overestimated standard errors by 36%. The discrepancy between the current study and Moeyaert et al.'s (2017) study may be due to the use of real data in the current study, where within-study effect sizes are correlated to some extent. The small value of estimated standard error yielded by the univariate method is expected. If we treat dependent effect sizes independently and presume each one is contributing information, the variance of the overall effect size will be underestimated. This would result in a narrow confidence interval for the combined effect size and increase the likelihood of committing a Type I error.

Variance of True Effect Size Estimates

3LM yielded the largest variance estimates, followed by the RVE, whereas the averaging method estimated a much smaller variance. The value produced by the 3LM was similar to the value produced by RVE. This pattern was consistent across the conditions except for the condition of all the effect sizes and the condition of 26-6. This finding was partially consistent with Moeyaert et al.'s (2017) study, which noted that the averaging method extremely underestimated the variance estimate. Given that the averaging method was implemented by averaging the within-study dependent effect sizes first, the variance of these within-study effect sizes was removed in this process. On the contrary, the RVE estimated the total variances, and 3LM estimated the two variances (i.e., between-study variance and between-outcome variance) separately. The current study also found that the variance estimated by 3LM was slightly smaller than RVE for the condition of 26-6. This was also similar to Moeyaert et al.'s (2017) finding that 3LM slightly underestimated variances for the condition of the number of studies of 25.

Limitations and Future Research

The limitation of the present study is the use of real data only, which limited the opportunity of having a criterion against which to evaluate the four methods of interest in terms of estimation accuracy. Future studies are recommended to use simulated data to investigate which method results in unbiased estimates of effect size and heterogeneity statistics. Second, the current study did not examine the effect of sources of dependence (e.g., multiple outcomes, multiple time-points) on the efficacy of the five methods. Dependent effect sizes from different outcome measures differ from those being measured at multiple time-points in data structure and correlation matrices, which may interact with the performance of these methods. Future studies are suggested to compare the methods in dealing with a certain type of dependence.

Implications and Conclusion

We have found statistically significant effect sizes between writing self-efficacy and writing achievement, which implies the statistically significant relationship between writing self-efficacy and writing outcomes. This finding corroborates the social cognitive theory that self-efficacy beliefs play a facilitative and mediating role in human functioning. We found that the true effect sizes were heterogeneous across studies, and a substantial amount of variance was between-study variance regardless of methods employed. This finding suggested the necessity of conducting moderator analyses to explore what factors may account for the heterogeneity.

The study has implications for meta-analysts. The current study found the univariate method produced much smaller estimates of the standard error when synthesizing dependent effect sizes, which could result in spuriously statistically significant effect sizes. Therefore, meta-analysts need to report whether the effect sizes are independent and avoid using the univariate method if the dependence issue occurs. However, this information is not always

available and the dependence issue is routinely ignored by meta-analysts. In Ahn's (2012) review of meta-analyses in education, 27% of meta-analyses did not provide information about whether the effect sizes are independent. The current study also provides meta-analysts with insights into choosing appropriate methods in addressing the dependent issue. While the three methods, the averaging method, 3LM, and RVE had similar results of overall effect size and standard error estimates, they performed differently in estimating variances. The averaging method had much smaller variances and the 3LM yielded the smallest value of variance for the condition of 26-6. A falsely small variance would conceal potential factors that may explain the heterogeneity. Therefore, meta-analysts are not recommended to adopt the averaging method and the 3LM is not preferred when the number of studies is small.

The study also has implications for primary researchers. First, primary researchers can assist meta-analysts by providing effect size information in addition to the information of statistical significance. The recommendation of reporting effect size information can be found in *Educational and Psychological Measurement* (Thompson, 1994) and a series of versions (4^{th}, 5^{th}, 6^{th}, 7^{th}) of *American Psychological Association Publication Manual* (American Psychological Association, 1994; 2001; 2010; 2019). Lack of the effect size information or sufficient statistics to compute effect sizes would invalidate the eligibility of a primary study in a meta-analysis and decrease the statistical power of the meta-analysis. Second, when a primary study has multiple outcome measures or multiple comparisons, primary researchers need to provide detailed information on these measures or treatments. Meta-analysts will benefit by knowing how these measures are conceptualized and operationalized, and how these treatments are implemented. The extent to which these measures or treatments are different helps meta-analysts to choose appropriate approaches to addressing the dependent issue. Primary researchers are also

recommended to provide correlations between multiple measures, which is necessary for modeling the dependency.

References

Ahn, S., Ames, A. J., & Myers, N. D. (2012). A review of meta-analyses in education: Methodological strengths and weaknesses. *Review of Educational Research, 82,* 436–476. http://doi.org/10.3102/0034654312458162

American Psychological Association. (1994). *Publication manual of the American Psychological Association* (4th ed.). Author.

American Psychological Association. (2001). *Publication manual of the American Psychological Association* (5th ed.). Author.

American Psychological Association. (2010). *Publication manual of the American Psychological Association* (6th ed.). Author.

American Psychological Association. (2019). *Publication manual of the American Psychological Association* (7th ed.). Author.

Borenstein, M., Hedges, L. V., Higgins, J. P., & Rothstein, H. R. (2009). *Introduction to Meta-Analysis*. John Wiley & Sons.

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, http://doi.org/10.3200/JOER.98.1.35-45

Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta- analyses: A structural equation modeling approach. *Psychological Methods, 19*(2), 211-229. http://doi.org/10.1037/a0032968

Cheung, M. W. L. (2019). A guide to conducting a meta-analysis with non-independent effect sizes. *Neuropsychology Review, 29,* 387-396. http://doi.org/10.1007/s11065-019-09415-6

Cheung, S. F., & Chan, D. (2004). Dependent effect sizes in meta-analysis: Incorporating the degree of interdependence. *Journal of Applied Psychology, 89*(5), 780-791. https://doi.org/10.1037/0021-9010.89.5.780

Cooper, H. M. (1998). *Synthesizing research: A guide for literature reviews* (3rd ed.). Sage.

Cooper, H. M. (2010). *Research synthesis and meta-analysis: A step-by-step approach* (4th ed.). Sage.

Fernández-Castilla, B., Jamshidi, L. Declercq, L., Beretvas, S. N., Onghena, P., & Van den Noortgate, W. (2020). The application of meta-analytic (multi-level) models with multiple random effects: A systematic review. *Behavior Research Methods, 52*, 2031-2052. https://doi.org/10.3758/s13428-020-01373-9

Fisher, Z., & Tipton, E. (2014). *Robumeta: Robust variance meta-regression*. Available at: http://cran.rproject.org/web/packages/robumeta/index.html

Geeraert, L., VandenNoortgate, W., Grietens, H., & Onghena, P. (2004). The effects of early prevention programs for families with young children at risk for physical child abuse and neglect: A meta-analysis. *Child Maltreatment, 9*, 277-291. https://doi.org/10.1177/1077559504264265

Hashemnejad, F., Zoghi, M., & Amini, D. (2014). The relationship between self-efficacy and writing performance across genders. *Theory and Practice in Language Studies*, *4*(5), 1045–1052. https://doi.org/10.4304/tpls.4.5.1045-1052

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust variance estimation in meta-regression with dependent effect size estimates. *Research Synthesis Methods, 1*, 39–65. http://doi.org/10.1002/jrsm.5

Higgins, J. P., Thompson, S. G., Deeks, J. J., & Altman, D. G. (2003). Measuring inconsistency in meta-analyses. *BMJ,* 327, 557-560. http://doi.org/10.1136/bmj.327.7414.557

Kalaian, H. A., & Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods, 1*, 227-235. http://doi.org/10.1037/1082-989X.1.3.227

Konstantopoulos, S. (2011). Fixed effects and variance components estimation in three-level meta-analysis. *Research Synthesis Methods, 2*, 61–76. http://doi.org/10.1002/jrsm.35

Lambert, R. G. (1995). *The robustness of the standard errors of summarized, corrected validity coefficients to non-independence and non-normality of primary data* (Unpublished doctoral dissertation). Georgia State University, Atlanta, Georgia.

Marín-Martínez, F., & Sánchez-Meca, J. (1999). Averaging dependent effect sizes in meta-analysis: A cautionary note about procedures. *Spanish Journal of Psychology, 2*, 32–38. http://doi.org/10.1017/S1138741600005436

Moeyaert, M., Ugille, M., Beretvas, S. N., Ferron, J., Bunuan, R., & den Noortgate, W. V. (2017). Methods for dealing with multiple outcomes in meta-analysis: a comparison between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *International Journal of Social Research Methodology, 20*(6), 559-572. https://doi.org/10.1080/13645579.2016.1252189

Perin, D., Lauterbach, M., Raufman, J., & Kalamkarian, H. S. (2017). Text-based writing of low-skilled postsecondary students: relation to comprehension, self-efficacy and teacher judgments. *Reading and Writing, 30*, 887-915. http://doi.org/10.1007/s11145-016-9706-0

Peterson, R. A., & Brown, S. P. (2005). On the use of beta coefficients in meta-analysis. *Journal of Applied Psychology, 90*(1), 175-181.

Raudenbush, S. W. & Bryk, A. S. (2002). *Hierarchical Linear Models: Applications and Data Analysis Methods*. Sage.

Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-analysis with complex research designs: Dealing with dependence from multiple measures and multiple group comparisons. *Review of Educational Research, 84*(3), 328-364. http://doi.org/10.3102/0034654313500826

Shintani, N., & Aubrey, S. (2016). The effectiveness of synchronous and asynchronous written corrective feedback on grammatical accuracy in a computer-mediated environment. *The Modern Language Journal, 100*(1), 296-319. http://doi.org/10.1111/modl.12317

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,* 102221. https://doi.org/10.1016/j.system.2020.102221

Sun, T., Wang, C., Lambert, R, & Liu, L. (In press). Relationship between Writing Self-Efficacy and Writing Achievement: A Meta-Regression Analysis. *Journal of Second Language Writing*.

Tanner-Smith, E. & Tipton, E. (2014). Robust variance estimation with dependent effect sizes: Practical considerations and a software tutorial in Stata and SPSS. *Research Synthesis Methods, 5,* 13-30. https://doi.org/10.1002/jrsm.1091

Tanner-Smith, E., Tipton, E., & Polanin, J. R. (2016). Handling complex meta- analytic data structures using robust variance estimates: A tutorial in R. *Journal of Developmental and Life-Course Criminology, 2*(1), 85–112. http://doi.org/10.1007/s40865-016-0026-5

Thompson, B. (1994). Guidelines for authors. *Educational and Psychological Measurement, 54*, 837-847.

Tipton, E. (2013). Robust variance estimation in meta-regression with binary dependent effects. *Research Synthesis Methods, 4*(2), 169-187. https://doi.org/10.1002/jrsm.1070

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods, 45*, 576–594. http://doi.org/10.3758/s13428-012-0261-6

Viechtbauer, W. (2010). Conducting meta-analyses in R with the metaphor package. *Journal of Statistical Software, 36*(3), 1-48. https://doi.org/10.18637/jss.v036.i03

OVERALL CONCLUSION

This dissertation aims to (a) provide the observed effect size of the relationship between writing self-efficacy and writing achievement in the EFL context as evidenced in an empirical study; (b) to estimate the true effect size of their relationship in a meta-analysis, and; (c) to compare three methods in handling multiple dependent effect sizes within studies in doing the meta-analysis.

The first article revealed a medium to large effect size (r = .47) of the relationship between writing self-efficacy and writing proficiency with students in the EFL context. There were statistically significant and positive relationships between writing self-efficacy, SRL strategies, and writing proficiency. Writing self-efficacy and writing SRL strategies made statistically significant contribution to the prediction of writing proficiency after controlling for students' demographic information of gender and SES. Specifically, two subscales of writing self-efficacy (i.e., Grammar and Use of English Writing) and three subscales of SRL strategies (i.e., Review of Records Strategies, Seeking Opportunities Strategies, and Self-Evaluation Strategies) statistically significantly predicted writing scores.

The second article estimated a medium effect size (r = .29) with 565 effect sizes extracted from a pool of 76 primary studies, indicating approximately 9% of the variability in writing achievement was associated with variability in students' self-efficacy. Writing in L1/L2 and statistical procedures were found to moderate effect size estimates. The effect size estimated with L2 writers was statistically significantly larger than that yielded with L1 writers. Additionally, regression tended to yield a statistically significantly smaller effect size estimate than correlation. This attenuation was particularly evident when controlling for affective or achievement variables in regression models.

The third article compared three methods in handling dependent effect sizes within studies in conducting a meta-analysis on the relationship between writing self-efficacy and writing achievement. Results revealed that the averaging methods, 3LM, and RVE yielded similar estimates in the overall average effect size and standard error. However, the averaging method had much smaller variance of true effect size estimates across studies. In addition, 3LM resulted in the smallest variance when the number of studies was small. The performance of the three methods was consistent as a function of the number of effect sizes per study and the number of primary studies.

This dissertation contributed to the extant literature by reinforcing the statistically significant relationship between writing self-efficacy and writing achievement. It corroborates the social cognitive theory that beliefs people have predict their behaviors and influence their outcomes (Bandura, 1997). It also confirms the cognitive theory of writing that cognitive and affective attributes such as self-efficacy are important components of writing processes and influence writing outcomes by interacting with other factors in the writing process (Hayes, 2000).

This dissertation had pedagogical implications for classroom teachers. First, since a medium to large effect size was observed with a sample of EFL students, appropriate strategies need to be incorporated into curriculum and instruction to reinforce self-efficacy beliefs in writing among English learners the EFL context. Teachers need to focus on the instruction of various genres of writing and the pragmatic aspect of writing because results suggested that Grammar and Use of English Writing statistically significantly predicted writing scores. Second, since the meta-analytic study found a statistically significantly larger effect size with L2 writers compared with L1 writers in English, it is more imperative for L2 English teachers or program

designers to employ appropriate strategies or implement effective interventions to boost L2 students' writing self-efficacy.

This dissertation also had implications for primary researchers and research methodologists. First, it has been found that effect size estimates tend to attenuate in regression models after controlling for affective variables (e.g., anxiety or motivation) and achievement variables (e.g., GPA). This finding implied that these variables covary with writing self-efficacy or writing achievement. Therefore, researchers are suggested to consider these covariates in examining the correlation between writing self-efficacy and writing achievement. Second, when a primary study has multiple outcome measures or multiple comparisons, primary researchers need to provide detailed information on these measures or treatments, which would help meta-analysts choose appropriate approaches to addressing the dependent issue. Primary researchers are also recommended to provide correlations between multiple measures for modeling the dependency. Third, meta-analysts are suggested to employ RVE or 3LM to handling dependent effect sizes, and RVE is recommended when the number of primary studies is small.

References

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. Center for Language Linguistics and Area Studies. Retrieved from https://www.llas.ac.uk/resources/gpg/2175.html

Bandura, A. (1986). *Social foundations of thought & action: A social cognitive theory*. Pearson Education.

Bandura A. (1997). *Self-Efficacy: The Exercise of Control*. Freeman.

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. http://doi.org/10.1037/a0029692

Bruning, R., & Horn, C. (2000). Developing motivation to write. *Educational Psychologist, 35*, 25-37. http://doi.org/10.1207/S15326985EP3501_4

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

Cheung, M. W. L. (2014). Modeling dependent effect sizes with three-level meta- analyses: A structural equation modeling approach. *Psychological Methods, 19*(2), 211-229. http://doi.org/10.1037/a0032968

Ekholm, E., Zumbrunn, S., & Conklin, S. (2015). The relation of college student self-efficacy toward writing and writing self-regulation aptitude: Writing feedback perceptions as a mediating variable. *Teaching in Higher Education, 20*(2), 197-207. http://doi.org/10.1080/13562517.2014.974026

Graham, S. (2006). Writing. In P. A. Alexander & P. H. Winne (Eds.), *Handbook of educational psychology* (pp. 457-478). American Psychological Association.

Graham, S., Kiuhara, S. A., Harris, K. R., & Fishman, E. J. (2017). The relationship among strategic writing behavior, writing motivation, and writing performance with young, developing writers. *The Elementary School Journal, 118*(1), 82-104. http://doi.org/10.1086/693009

Graham, S., Wijekumar, K., Harris, K., Lei, P., Fishman, E., Ray, A. B., & Houston, J. (2019). Writing skills, knowledge, motivation, and strategic behavior predict students' persuasive writing performance in the context of robust writing instruction. *The Elementary School Journal, 119*(3), 487-510. http://doi.org/10.1086/701720

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.

Hedges, L. V., Tipton, E., & Johnson, M. C. (2010). Robust Variance Estimation in Meta-regression with Dependent Effect Size Estimates. *Research Synthesis Methods, 1*(2): 164–165. http://doi.org/10.1002/jrsm.17

Holden, G., Moncher, M. S., Schinke, S. P., & Barker, K. M. (1990). Self-efficacy of children and adolescents: A mete-analysis. *Psychological Reports, 66,* 1044-1046.

Huang, C. (2013). Gender differences in academic self-efficacy: A meta-analysis. European *Journal of Psychology of Education, 28*, 1-35. http://doi.org/10.1007/s10212-011-0097-y

Jones, E. (2008). Predicting performance in first-semester college basic writers: Revisiting the role of self-beliefs. *Contemporary Educational Psychology, 33*, 109-238. http://doi.org/10.1016/j.cedpsych.2006.11.001

Judge, T. A., & Bono, J. E. (2001). Relationship of core self-evaluations traits—self-esteem, generalized self-efficacy, locus of control, and emotional stability-with job satisfaction and job performance: A meta-analysis. *Journal of Applied Psychology, 86*, 80-92. http://doi.org/10.1037/0021-9010.86.1.80

Kalaian, H. A., & Raudenbush, S. W. (1996). A multivariate mixed linear model for meta-analysis. *Psychological Methods, 1*, 227-235. http://doi.org/10.1037/1082-989X.1.3.227

Moeyaert, M., Ugille, M., Beretvas, S. N., Ferron, J., Bunuan, R., Van den Noortgare, W. (2017). Methods for dealing with multiple outcomes in meta-analysis: A comparison Between averaging effect sizes, robust variance estimation and multilevel meta-analysis. *International Journal of Social Research Methodology, 20*(6), 559–572. http://doi.org/10.1080/13645579.2016.1252189

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. http://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). Institute of Education Sciences, U.S. Department of Education,

Washington, D.C. Retrieved from https://nces.ed.gov/nationsreportcard/pdf/main2011/2012470.pdf

National Commission on Writing. (2004, September). *Writing: A ticket to work...or a ticket out: A survey of business leaders.* Retrieved May 3, 2005, from http://www.writingcommission.org/report/html

Neugebauer, S. R., & Howard, E. R. (2015). Exploring associations among writing self-perceptions, writing abilities, and native language of English-Spanish two-way immersion students. *Bilingual Research Journal, 38*(3), 313-335. http://doi.org/10.1080/15235882.2015.1093039

Pajares, F. (2002). Gender and perceived self-efficacy in self-regulated learning. *Theory Into Practice, 41*(2), 116-125. http://doi.org/10.1207/s15430421tip4102_8

Pajares, F. (2003). Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading & Writing Quarterly, 19*(2), 139-158. http://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, http://doi.org/10.1080/07481756.2007.11909801

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. http://doi.org/10.1006/ceps.1999.1027

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. (1999). Grade level and gender differences in the writing self-beliefs of middle school students. *Contemporary Educational Psychology, 24*(4), 390-405. https://doi.org/10.1006/ceps.1998.0995

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. http://doi.org/10.1006/ceps.2000.1069

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. http://doi.org/10.1080/01443410.2011.621411

Sadri, G., & Robertson, I. T. (1993). Self-efficacy and work-related behavior: A review and meta-analysis. *Applied Psychology: An International Review, 42*(2), 139-152. https://doi.org/10.1111/j.1464-0597.1993.tb00728.x

Sanders-Reio, J., Alexander, P. A., Reio, Jr., T. G., Newman, I. (2014). Do students' beliefs about writing relate to their writing self-efficacy, apprehension, and performance. *Learning and Instruction, 33*, 1-11. http://doi.org/10.1016/j.learninstruc.2014

Scammacca, N., Roberts, G., & Stuebing, K. K. (2014). Meta-analysis with complex research designs: Dealing with dependence from multiple measures and multiple group comparisons. *Review of Educational Research, 84*(3), 328-364. http://doi.org/10.3102/0034654313500826

Schunk, D. H., & Zimmerman, B. J. (2007). Influencing children's self-efficacy and self-regulation of reading and writing through modeling. *Reading & Writing Quarterly, 23*(1), 7-25. http://doi.org/10.1080/10573560600837578

Shell, D. F., Murphy, C. C., & Bruning, R. H. (1989). Self-efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology, 81*(1), 91-100. http://doi.org/10.1037/0022-0663.81.1.91

Stajkovic, A., & Luthans, F. (1998). Self-efficacy and work-related performance: A meta-analysis. *Psychological Bulletin, 124*(2), 240-261. http://doi.org/10.1037/0033-2909.124.2.240

Teng, L. S., Sun, P. P., & Xu, L. (2018). Conceptualizing writing self-efficacy in English as a foreign language context: Scale validation through structural equation modeling. *TESOL Quarterly, 52*(4), 911-942. http://doi.org/10.1002/tesq.432

Unrau, N. J., Rueda, R., Son, E., Polanin, J. R., Lundeen, R. J., & Muraszewski. (2018). Can reading self-efficacy be modified? A meta-analysis of the impact of interventions on reading self-efficacy. *Review of Educational Research, 88*(2), 167-204. http://doi.org/10.3102/0034654317743199

Van den Noortgate, W., López-López, J. A., Marín-Martínez, F., & Sánchez-Meca, J. (2013). Three-level meta-analysis of dependent effect sizes. *Behavior Research Methods, 45*, 576–594. http://doi.org/10.3758/s13428-012-0261-6

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. http://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. http://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. *Pan-Pacific Association of Applied Linguistics, 16*(2), 101-121. Retrieved from http://web.a.ebscohost.com