2019

Variable- and Person-Centered Approaches to Examining Construct-Relevant Multidimensionality in Writing Self-Efficacy

Morgan DeBusk-Lane
Virginia Commonwealth University

Variable- and Person-Centered Approaches to Examining Construct-Relevant Multidimensionality in Writing Self-Efficacy

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

by

Morgan L. DeBusk-Lane
Bachelor of Science, Pennsylvania State University, 2016

Director: Sharon Zumbrunn, Ph.D.
Associate Professor, Foundations of Education
School of Education

Virginia Commonwealth University
Richmond, Virginia
May, 2019

Acknowledgement

There are several people to whom I own a great amount of thanks, and it would be a monumental understatement to say I made it this far without their love, support, and friendship along the way. First, my wife, my love, my everything. There are few people in this world that we meet that shape the entire scope of our lived experience. Thank you for your support, time, energy, and love through all of this. I am beyond fortunate.

To my committee: your effort, passion, and belief in me has allowed me to develop, fall in love with statistics, fall in love with more statistics, and find passions along any path I stumbled upon. Mike: your vigor towards statistical methods, especially the cool fancy ones, has driven me to find my own. Christine: the opportunities you've provided and openness to those crazy stats ideas I've found has shown me no bounds. Roger: your perspective and longevity in the field of writing motivation has truly served as an example I wish to follow. Sharon: thank you for guiding me, challenging me, and, more often than not, helping me shimmy on down. Your guidance, candor, and mentorship throughout the last few years is something I am extremely grateful for—thank you!

There is absolutely no way that I'd be where I am without my fellow cohort members: Eric and Ashlee. Our struggles, collaborations, successes, and conversations over the past few years will always give me found memories of graduate school and the experiences we navigated. I have been and will continue to believe I was very lucky to have both of you.

Dedication

This dissertation is dedicated to my two children. It is with this effort that I hope you see my passion to figure things out, I hope you see learning, and I hope you see that both take work, time, and the surrounding support of many. You are both what drives me to be better, succeed, and provide. It is my wish to give you opportunity and that you find in life passion, love, and the drive to always be better.

Table of Contents

List of Tables

List of Figures

Abstract

VARIABLE- AND PERSON-CENTERED APPROACHES TO EXAMINING CONSTRUCT-RELEVANT MULTIDIMENSIONALITY IN WRITING SELF-EFFICACY

By Morgan L. DeBusk-Lane, Ph.D.

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University.

Virginia Commonwealth University, 2018

Director: Sharon Zumbrunn, Ph.D.
Associate Professor, Foundations of Education
School of Education

Writing self-efficacy is a vital component to a students' motivation and will to succeed towards writing. The measurement of writing self-efficacy over the past 40 years, despite its development, continues to largely be represented by Confirmatory Factor Analysis models that are limited due to their restricted item to factor constraints. These constraints, given prior literature and the theoretical understanding of self-efficacy, do not adequately model construct-relevant psychometric multidimensionality as a product of conceptual overlap or a hierarchical or general factor. Given this, the present study's purpose was to examine the adapted Self-efficacy for Writing Scale (SEWS) for the presence of construct-relevant psychometric multidimensionality through a series of measurement model comparisons and person-centered approaches. Using a sample 1,466 8th, 9th, and 10th graders, a bifactor exploratory structural equation model was found to best represent the data and demonstrate that the SEWS exhibits both construct-relevant multidimensionality as a function of conceptual overlap and the presence of a hierarchical theme. Using factor scores derived from this model, latent profile analysis was conducted to further establish validity of the measurement model and examine how students disaggregate into groups based on their response trends of the SEWS. Three profiles emerged greatly differentiated by global writing self-efficacy, with obvious and substantively varying specific factor differences between profiles. Concurrent, divergent, and discriminant validity evidence was established through a series of analyses that assessed predictors and outcomes of the profiles (e.g. demographics, standardized writing assessments, grades). Theoretical and educator implications and avenues for future researcher were discussed.

Chapter 1: Introduction

"Self-belief does not necessarily ensure success, but self-disbelief assuredly spawns failure"

-Bandura, *Self-Efficacy: The Exercise of Control, 1997*

As a foundational component to Albert Bandura's (1997) social cognitive theory, self-efficacy, or "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments" (p. 3), is an integral com ponent to the function of human agency. Therein, self-efficacy describes how self-perceptions of capacity to perform tasks and skills influence one's behavior, affect, persistency, and achievement. Seemingly obvious, the domain of writing, which necessitates various interrelated sub-skills, frames, and procedures--spelling, grammar usage, punctuation, organization, voice, prose--and the ability to orchestrate them in a cohesive manner, is appropriately dependent on one's efficacy. As such, research over the past 40 years have focused great attention to how, why, and to what degree efficacious beliefs influences writing performance and its relationship to other forms of motivation, while little attention has focused on examining and progressing the methodological exploration of instruments to adequately capture the dimensionality of writing self-efficacy. Therefore, the overarching purpose of this study is to extensively examine writing self-efficacy's dimensionality and build validity evidence to further substantiate the adapted Self-efficacy for Writing Scale (SEWS; Ekholm, Zumbrunn, & Conklin, 2015; Zumbrunn, Marrs, & Mewborn, 2016).

Writing Self-Efficacy

Provided its integral position as a psychological mechanism for both effort and persistence (Bandura, 1997; Schunk & DiBenedetto, 2016; Schunk & Usher, 2012), it is no surprise that self-efficacy has been extensively studied as a major component to writing motivation (Pajares, 2003, 2007, Schunk, 2003). Rightfully so, writing self-efficacy has long shown to be predictive and related to writing achievement (Bruning & Horn, 2000; Graham, Harris, Kiuhara, & Fishman, 2017; Pajares & Johnson, 1996; Pajares, Miller, & Johnson, 1999; Pajares, & Valiante, 1997, 2001; Pajares, Johnson, & Usher, 2007; Shell, Colvin, & Bruning, 1995; Shell, Murphy & Bruning, 1989; Zimmerman & Bandura, 1994). Nevertheless, among these works and others, the effort to capture, measure, or otherwise operationalize writing self-efficacy has not been without difficulty, as appropriately aligning it with theory has been problematic (Klassen & Usher, 2010; Pajares, 2003). Self-efficacy researchers have consistently been warned that without adhering to proper item wording, time-vantage, focus, and conceptualization, "the future of self-efficacy research as a theoretically grounded means of understanding human behavior is threatened." (Klassen & Usher, 2010, p. 20). Given this, the field has recently progressed both in its theoretical alignment and the extent to which it focuses on specific process-oriented facets of writing (Bruning, Dempsey, Kauffman, McKim, & Zumbrunn, 2013; Klassen & Usher, 2010). Although, little research has focused on the psychometric properties of measures of writing self-efficacy that can further validate how these measures capture efficacious beliefs of writing and how it is theoretically situated.

Psychometric Properties of Writing Self-Efficacy

Over the past 40 years, researchers have commonly used basic means of item reduction (e.g. exploratory factor analysis), reliability, and confirmatory factor analyses (CFA), to judge the psychometric quality of writing self-efficacy measures (McCarthy, Meier, & Rinderer, 1985; Pajares et al., 1999; Pajares & Valiante, 2001; Shell et al., 1989, 1995). The advent and ease of employing advanced psychometric methods has only recently become prevalent and highly accessible as computers have advanced (Brown, 2015; Kline, 2016). For example, work by Englehard and Behizadeh (2012) used Rasch measurement theory (a type of item response theory; Rasch, 1960) to examine the psychometric quality of the Writing Self-Efficacy Scale (WSES; Pajares et al., 1999). Similarly, although not focused specifically on assessing psychometric quality, works by De Smedt and colleagues (2017, 2018) and Zumbrunn, Broda, Varier, & Conklin (2019) have employed structural equation models to examine writing self-efficacy's relationship to other motivational and cognitive constructs.

Drawing from the increased use of factor analyses across the literature, writing self-efficacy has commonly been depicted as a unidimensional factor (Pajares & Valiante, 2006); however, a growing amount of literature suggests that it is multidimensional as its conceptualization continue to evolve (Bruning et al., 2013; De Smedt, Van Keer, & Merchie, 2016; De Smedt, Merchie, Barendse, Rosseel, Van Keer, & De Naeghel, 2017; De Smedt, Graham, & Van Keer, 2018; MacArthur, Philippakow, & Ianetta, 2015). That is, measures of writing self-efficacy have commonly sought to capture efficacious beliefs of particular tasks and skills inherent to writing (e.g. grammar, punctuation, organization, syntax usage, argument) (see Pajares, 2003). Newer research, however, has consistently added and incorporated items focused on writing self-regulation (e.g. organization, focus, strategy use, planning) and other cognitive components (e.g. ideation, creativity, idea development) involved in the writing process (e.g. Bruning et al., 2013; Graham et al., 2017; MacArthur et al., 2015). Of these, Bruning and colleagues' (2013) Self-Efficacy for Writing Scale (SEWS) focuses on efficacious beliefs of *ideation*, traditional writing *mechanics*, and *self-regulation*, and has been widely used and adapted since publication (e.g., De Smedt et al., 2016, 2017, 2018; Ekholm et al., 2015; Ramos-Villagrasa, Sanchez-Iglesias, Grande-de-Prado, Olivan-Blazquez, Martin-Pena, & Cancer-Lizaga, 2018; Zumbrunn et al., 2016). Therein, *ideation* serves to depict a writer's efficacious beliefs of their ability to produce, create, and use ideas. *Conventions*, similar to many measures often focused on writing's skills and tasks, seeks to capture a writer's beliefs associated with common standards, such as grammar and spelling, that are employed to communicate with writing. Lastly, self-efficacy for writing *self-regulation* depicts a writer's confidence to "direct themselves" (affective response), organize, and navigate through the writing process (Bruning et al., 2013).

Given its wide use, various studies have confirmed the multidimensional factor structure originally portrayed by Bruning and colleagues (2013; De Smedt et al., 2016, 2017, 2018; Yilmaz Soylu et al., 2017). Additionally, studies have adapted or extended the SEWS to new languages and samples (Ekholm et al., 2015; Ramos-Villagrasa et al., 2018; Zumbrunn et al., 2016). Notably, work by Ekholm, Zumbrunn, and Conklin (2015), adapted the SEWS by reducing it to 9 items, yet in doing so confirmed a single factor structure with an undergraduate sample. Extending this work to be more developmentally-appropriate for younger writers, Zumbrunn and colleagues (2016) further adapted the SEWS by adjusting the traditional 0-100 rating scale, to a 0-4 rating scale. Incorporating both adaptations, recent work by Zumbrunn and colleagues (2019), which used a 9-item, 0-4 rating scale, adaptation of the SEWS, found a 3-factor measurement structure invariant across elementary and high-school students. Furthermore, work by DeBusk-Lane, Lester, and Zumbrunn (2018) found a 3-factor measurement structure of the adapted SEWS with middle and high-school students. Although a well-fitting 3-factor structure is seemingly evident across developmental spectrums, this structure has also exhibited clues that suggest other models may more accurately model the data. Together, with the field progressing in how it conceptualizes and operationalizes writing self-efficacy and the ease in which advanced statistical methods can be employed to answer or shed light on critical motivational relationships, it is vital to ensure such measures are psychometrically sound and accurately represent the data.

Advancing Psychometric Quality

Given the prevalence of a 3-factor model, two trends suggest and put into question the predominant ways in which the SEWS has traditionally been modeled. First, because the measure was originally constructed to capture efficacious beliefs of writing collectively through multiple dimensions, it is likely that it does, in fact, represent both global and specific constructs. That is, it is both theoretically expected and logically plausible to expect subscales within a measure with similar or related domain specific facets to exhibit some amount of a global or hierarchical factor for which reflect participants' overall sense of writing self-efficacy (Reise, Bonifay, & Haviland, 2013). Theoretically, Bandura (1997) explained in detail how multidimensional measures of self-efficacy can exhibit these trends. In that, he explained self-efficacy factors may share similar subskills, incorporate skills that are developed together, enact similar self-regulatory mechanisms, use similar approaches to problem solving, and query constructs that similarly draw from past experiences that have bolstered one's belief in their ability (Bandura, 1997). Recent studies have brought into question whether the adapted SEWS is best modeled by three distinct factors (DeBusk-Lane et al., 2018; Zumbrunn et al., 2019) or a single factor alone (Ekholm et al., 2015; Zumbrunn et al., 2016). Furthermore, across both the original 16- and the adapted 9-item measures, moderate latent factor correlations, large first factor eigenvalues, and moderate correlations among the specific factors to other unidimensional writing self-efficacy measures suggest the presence of a hierarchical or global factor (DeBusk-Lane et al., 2018; MacArthur et al., 2016; Ramos-Villagrasa et al., 2018; Reise et al., 2013; Zumbrunn et al., 2019). Second, it can be expected that efficacious beliefs derived and exhibited by items that query beliefs associated with "writing even when it is difficult" likely translate and extend to cross-factor items that query beliefs associated with a writer's effort to "think of many words to describe my ideas." This conceptual relationship or overlap further suggests the items may be related to more than one specific factor. Therefore, because the items themselves are imperfect indicators that likely associate with other similar latent constructs, aside from their a priori forced factor relationship, current depictions through CFA may be bias and not accurately depict reality (Asparouhov, Muthen, & Morin, 2015; Morin, Arens, & Marsh, 2016; Morin et al., 2017).

Together, these two hypothesized influences (e.g. global or hierarchical factor and item cross-factor relationships or cross-loadings) are referred to as sources of construct-relevant psychometric multidimensionality (Morin et al., 2016; 2017). That is, in typical CFA models, item factor relationships restrict cross-loadings to zero, forcing true-score variability between factors (of both cross-loading and hierarchical/global factors) to be absorbed by only a-priori factors, negating both the presence of hierarchically ordered and conceptually overlapped constructs, which may result in bias parameter estimates (Asparouhov et al., 2015).

Furthermore, although the original SEWS has been related to various other psychological and motivational constructs (see De Smedt et al., 2017, 2018; Zumbrunn et al., 2019), they are commonly modeled by either composite scores (specific factor item means) or latent factor values derived from plausibly biased CFA latent factor scores. Therefore, to aid in expanding and providing more robust validity evidence, the adapted SEWS herein will also be examined as it relates to both writing apprehension and a separate writing self-efficacy measure, the WSES (Pajares, 2007). In this effort, this study will employ latent profile analysis to disaggregate the final factor structure to provide a more detailed, person-centered, approach. In doing so, this study will explore differences among profiles in relation to these well-established constructs (writing apprehension and writing self-efficacy), providing further validity evidence using advanced statistical methods.

Problem Statement

The purpose of this study is to examine construct-relevant multidimensionality within the adapted SEWS and provide further validity evidence (Ekholm et al., 2016; Zumbrunn et al., 2016). To date, no other study has further examined the adapted SEWS beyond traditional CFA model depictions, which have been shown to be limited and less than accurate among multidimensional measures that purport to capture a particular construct with a multi-dimension multi-factorial approach (Asparouhov et al., 2015; Morin et al., 2016, 2017). Although the field has consistently modeled writing self-efficacy with multidimensional measures, such specific dimensions or factors have commonly been conceptually similar and likely draw on beliefs that underscore each other (e.g. common writing skills, tasks), as Bandura (1997) clearly explained, may purport some level of generality or global level therein. Furthermore, with the growing trend of statistically assessing latent concepts with structural equation modeling, it is important to accurately model the data to ensure relational parameter estimates represent accurate true score and construct-irrelevant variation among, arguably, the most powerful predictor of eventual success--self-efficacy. To this end, the current problem is that the SEWS has traditionally been modeled within a common CFA framework, while clues clearly suggest the presence of a global latent factor and conceptual overlap among the domain specific factors. This problem instigates the following research question.

Research Questions

1. Are the items of the SEWS conceptually related across a priori factors?
2. Does the SEWS exhibit hierarchically-ordered constructs?
3. What specific quantitative profiles of writing self-efficacy emerge?
4. What forms of validity evidence is found for the profiles of the SEWS?
    a. Do the profiles exhibit concurrent validity evidence based on responses to the WSES?
    b. Do the profiles exhibit divergent/discriminant validity evidence based on responses to the Writing Apprehension Scale (WAS-12)?
    c. Do the profiles exhibit predictive validity?

Brief Overview of Methodology

The present study is a substantive-methodological synergy (Marsh & Hau, 2007), which employs new, evolving, or advanced methodological approaches to substantively import research questions or topics. Substantively, this thesis examines the dimensionality of writing self-efficacy from the adapted SEWS, which depicts writing self-efficacy as "three classes of activities… consistently involved in the writing act: self-regulation, ideation, and conventions" (Bruning et al., 2013, p. 25; Ekholm et al., 2015; Zumbrunn et al., 2016, 2019). This effort may clarify and better align the depiction of writing self-efficacy to follow theoretical assumptions, provide both theoretical and instrumental validity evidence, and further advance the field of self-efficacy at large by examining advanced models that can also help inform theory and the construct's relation to other motivational variables. Methodologically, this thesis tests the aforementioned hypotheses and assumptions with a number of competing models (e.g.

hierarchical CFA, exploratory structural equation modeling (ESEM), bi-factor CFA, and bifactor ESEM models) and provides validity evidence through person-centered methods by assessing profiles themselves, predictors, and outcomes from a latent profile analysis (LPA).

Definition of Terms

Writing Self-efficacy. One's belief or self-perceptions in their capability to write (Bandura, 1997, 2006, 2012; Pajares, 2003).

Multidimensionality. The presence or intention of a measurement instrument to exhibit or display more than one concept, factor, or statistical grouping of items (Anderson, Kahn, & Tindal, 2017; Reise et al., 2013).

Substantive-Methodological Synergies. Approaches to research that employ advanced methodologies in novel ways on substantively important and viable problems or topics (Marsh & Hau, 2007).

Construct-relevant Psychometric Multidimensionality. In classical test theory, score variance is naturally comprised of three sources: random measurement error, construct-irrelevant sources of true score variance, and construct true score variance. In this case, construct-relevant true score variance refers to variation that represents participants' actual variability of the target construct of measure, as opposed to influences that are 'irrelevant' to the target construct (e.g. another measured construct or their own cognitive ability), and random error (e.g. room temperature, an annoying sound, or a recent argument). Therefore, construct-relevant psychometric multidimensionality specifically targets construct-relevant sources of true-score variability among multidimensional measures, of which are consistently comprised of at least two other sources of true-score variation.

True-score variation. Absent error or construct-irrelevant variability, true-score variation refers to variability found among finite test occasions. Comparatively, a true-score represents one's score averaged among an infinite amount of test occasions (Lord, Novick, & Birnbaum, 1968; Mellenbergh, 1996).

Chapter 2: Review of Literature

The purpose of this study is to examine construct-relevant multidimensionality within the adapted SEWS and provide further validity evidence (Ekholm et al., 2015; Zumbrunn et al., 2016). Collectively, this chapter will set this stage by reviewing relevant and important literature instrumental in arranging and orchestrating a substantive-methodological synergy. To start, I will review self-efficacy's positionality within Social Cognitive Theory and detail how theory contends it be measured. Next, I will review important studies throughout the history of writing self-efficacy that have developed, examined, and used multidimensional measures. Then, I will introduce and describe the methodological utility of a substantive-methodological synergy, that focuses on examining substantively important topics with applications of emerging and advanced statistical methodologies. Finally, I will conclude the chapter by discussing how both the theoretical and methodological considerations inform the current study's conceptual framework and development of hypotheses.

Theoretical Framework

As a basis to understanding self-efficacy's positionality in theory, Albert Bandura's work on a social cognitive interactional model of human functioning, which squarely contends self-referent and representative thoughts as central to that which motivates human behavior, is paramount. In 1977, Bandura postulated that one's expectations of personal efficacy was instrumental to influencing behavior, persistence, and effort. Fundamental to this causal relationship was the hypothesis that self-referent thoughts or efficacy beliefs mediate the relationship (Bandura, 1982). Bandura (1982) further situated this cognitive influence as a central mechanism in regulating the exercise of control, often referred to as agency. Furthermore, from within the conceptualization of human agency, self-efficacy beliefs are but one facet involved in a “broad network of socio-structural influences” that interact bidirectionally with one’s environment (Bandura, 2001, p. 1). Together, the interaction between one’s environment, behavior, and their own personal and cognitive factors is the basis for what Bandura referred to as triadic reciprocal determinism (1997, see Figure 1). Therein, personal agency operates within the triadic reciprocal model with personal attribu. tes of self-efficacy beliefs, behavior, and the environment bidirectionally towards the perpetual motive of self-control (Bandura, 1997).

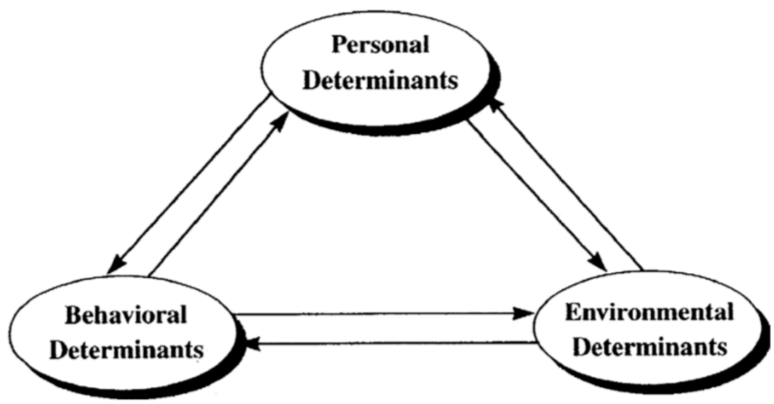


Figure 1. Bandura’s (2012) “Schematization of Triadic Reciprocal Determination”

Self-efficacy

As the central tenant to human agency, personal efficacy is the foundational personal influencing capability inherent to agentic influence (Bandura, 1997). That is, self-efficacy, or “judgments of how well one can execute courses of action required to deal with prospective situations,” are not isolated predictions of future ability (Bandura, 1982, p. 122). Self-efficacy beliefs are dynamic cognitive guidance mechanisms that are derived from various experiences, observations, situations, and relations that help regulate motivation, influence choices, and perpetuate behavior (Bandura, 1997). Consequently, because what underscores efficacy beliefs is laden in personal attributes, experiences, and views, it is no surprise that efficacy beliefs vary by person, developmental trajectories, situation, topic, and skill level across numerous domains and competencies (Bandura, 1997). To this point, efficacy beliefs are unique to both individuals and that which they are contended. Classical pianists may have very little perceived efficacy in terms of playing a jazz piano composition, however, they will likely vary in their beliefs to play Chopin's Prelude No. 4 with well-established expressions, changes in tempo, or timbre. As such, efficacy beliefs are not a broad collective system of beliefs, but a set of beliefs that distinctly focus on unique domains of functioning (Bandura, 1997).

Self-efficacy beliefs operate to orchestrate an individual's behavioral, cognitive, affective, and social dispositions to effectively influence personal choice, motivation, cognition, and one's own environment (Bandura, 1993). Although individuals may harness the needed skills, competencies, and ability to perform a task, they may lack the motivation to self-orchestrate them into effective directed effort and behavior. Moreover, self-efficacy is concerned with what an individual believes they can do with such skills and ability. Those skills, competencies, and abilities that align with positive self-beliefs naturally produce a generative path of further competency and development. In other words, that which people believe they can succeed in is likely to reinforce their own aspirations, goals, behavior, and motivation accordingly. Herein lies the perpetuated relationship between self-efficacy beliefs and performance. This causal relationship, between beliefs and performance, is noticeably bidirectional. Even with the needed skills and ability, a lack of self-belief or doubt can cripple an individual's use of them (Bandura & Jourden, 1991). Consequently, even with inferior skills, ability, or competence, a fortified sense of efficacy can endure and enable success through perseverance (Bandura, 1992; White, 1982). The fostering of self-efficacy beliefs is not only derived from the relationship between beliefs and skill performance but is also a product of influences and interpretations of other sources. Although Bandura (1997) postulated that individuals rely upon their interpretation of four main sources of information (mastery experiences, vicarious experiences, social persuasions, and physiological and affective states) when judging what they believe they can do, recent theoretical developments question how fully they encompass the breadth of what is interpreted to establish efficacious beliefs (Graham, 2018; Usher & Weidner, 2018).

Beliefs derived from past experiences of success or failure (*mastery experience*), the interaction with others (*social persuasion*), watching others (*vicarious experience*), and who how one feels when thinking about a particular activity or tasks (*physiological response*), have traditionally been viewed as the predominant sources for which people shape efficacious beliefs (Usher & Pajares, 2008). Furthermore, recent work by Graham (2018) and Usher and Weidner (2018) suggest people also incorporate and acknowledge a cultural lens that influences their efficacious beliefs. In doing so, work by Kitayama and Uskul (2011) and Chang and colleagues (2011), demonstrate that in some collectivist cultures, self-efficacy may be equally or more predictive of performance, which may be attributed to the view of self between cultures and identities and the extent to which some cultures tend to focus their efforts (towards others, feelings or more inwardly focused self-evaluation). Work by Usher and Weidner (2018) also described that such influences do not only exist globally, but may also include other demographic influences such as gender, societal stratifications, and contextual hierarchical positioning in situationally dependent frames. In sum, sources are critical to what underscores efficacious beliefs, marking their obvious individuality and experiential premise, while also acknowledging it is not only past positive experiences that support self-efficacy.

Measurement of Self-Efficacy

Clearly the distinction of self-efficacy necessitates its need to be uniquely operationalized. To adequately collect and subsequently interpret accurate appraisals of self-efficacy, measures should be operationalized consistent and aligned with theory's conceptualization (McMillan, 2016; APA, AERA, NCME, JCSEPT, 2014). As Klassen and Usher (2010) state, "Valid measurement is a keystone of scientific inquiry; problems with measurement hamper progress in knowledge building and in practical application of research findings" (p. 15).

This section will seek to describe the components of an appropriate self-efficacy measure earlier outlined from the theoretical placement of self-efficacy. That is, efficacy scales should be structured such to measure an individual's judgement of their future capability across that which encompasses the domain in question (*specificity/generality*), through varying levels of demand (*level*), from varying degrees of confidence (*strength*), conceptually aligned (*alignment*), and connected to the criterial task at hand (*correspondence*) (Bandura, 1997; 2006; Klassen & Usher, 2010; Marsh et al., 2018; Pajares, 1996).

Specificity and Generalizability

Judgements of self-efficacy have long been prescribed to be strongly domain or skill dependent (Bandura, 1986, 1997, 2006, 2018; Bong & Skaalvik, 2003; Klassen & Usher, 2010; Marsh et al., 2018; Pajares & Usher, 2008; Pajares, 1996; 2006; Usher, 2015). This characteristic is often connected with and described throughout the literature as needing to be overtly aligned and in the same conceptual scope as a given performance or achievement assessment (Bandura, 1997, 2006; Marsh et al., 2018; Pajares, 1996). This is often prescribed to optimize the predictive nature between a particularly specified self-efficacy belief and the same specifically defined domain performance. In its most aligned form, this has been described as *test-related* self-efficacy (Marsh et al., 2018). As mentioned earlier, the field of self-efficacy has yet to fully examine the full hierarchical nature of self-efficacy and the inherent relationships posited to exist along a spectrum of efficacy beliefs held between global (or as Marsh described as *generalized*) and *test-related* measures of self-efficacy. Further, in some cases there is no scientific need or reasoning to include performance or achievement in a study and therefore the precise alignment is not needed. However, even without a criterial task to align to, that which the self-efficacy measure encompasses should be conceptually and meaningfully congruent with the topic at hand. For example, if the topic of a study is primarily concerned with a child's ability to perform mathematical fractions, a global mathematical self-efficacy measure that likely includes many skills and abilities not cumulatively involved in fractions may be less predictive and informative. Conversely, a math self-efficacy measure that focuses on the skills and abilities related to understanding and applying fractions may be less predictive and applicable when a study's focus and performance measures is surmised to a child's performance on an end of course evaluation or summative assessment. In relation to writing, a high level of *specificity* may focus items to particular types of writing or particular skills and tasks inherent to particular genres, or it may be a collection of similar facets inherent to the writing process. Therefore, dependent upon the need and use, it is up to the researcher to acknowledge and support such *specificity* to conceptual alignment a study's focus to that of its measure. Despite this approach, Bandura (1997) strongly contends greater *specificity* than not.

Level

Similar to *specificity*, employing items with an appropriate amount of task demands, relative to the domain topic at hand, is vital to capturing accurate representations and variability associated with efficacy beliefs. The *level* at which efficacy is measured must be both developmentally appropriate and include an adequate description sufficient enough to assess differences in perceived capability. In other words, the range of demands presented must adequately cover the domain. That is, items that include appropriate *level* may include situational conditions, such as a performance standard, score, or metric against which individuals may judge their perceived capabilities (Bandura, 1997). For example, if a set of items orchestrated to capture mathematical fraction self-efficacy only inquires upon an individual's perceived ability to 'adequately interpret, add, and subtract fractions', it may diminish respondents' variability about which they truly perceive themselves capable and bias results, as "adequately" can be individually interpreted to mean many things. Furthermore, if the domain of fractions actually covers a far greater span of content, say, including multiplication and division, more elementary concepts may result in ceiling effects (Bandura, 1997; Marsh et al., 2018). Therefore, item *level* effects should include developmentally appropriate criteria and criteria with varying levels of task demands that can be correctly interpreted within the respondents' own context. Measures created to measure an elementary student's fraction self-efficacy is likely not appropriate for high-school students.

Strength

Clearly the most argued and contentious characteristic of self-efficacy measurements, *strength* commonly refers to the response scale used to capture an individual's strength of their beliefs in their ability to complete or execute a particular *level* of task demand (Bandura, 1997). The "standard methodology" suggests items be "phrased in terms of *can do* rather than *will do.*

Can is a judgment of capability; *will* is a statement of intention" (Bandura, 1997, p. 43, 2006). To measure *can* responses in the standard methodology, *strength* is recorded on a 100-point response scale that ranges from 0 to 100 in 10-unit increments from 0 ("Cannot do"); through 50 ("Moderately certain can do"); to complete assurance ("Highly certain can do") (Bandura, 1997; 2006). Dependent on the cognitive developmental condition of respondents, other simpler response formats may be used (Bandura, 2006). Although outside the scope of this review, other response formats have been used with varying levels of success across various domains (Klassen & Usher, 2010). Notably, Pajares, Hartley, and Valiante (2001) compared the same items with two different response scale formats, Bandura's and another shortened 1 ("no confidence at all") to 6 ("completely confident"), and found no significant measurement model differences, yet did identify greater GPA prediction from the 0-100 scale format. Also, work by Smith and colleagues (2003) specifically examined response scales and found a 4-point scale was adequate among Grades 4-5 using Rasch techniques. Furthermore, similar research has also suggested smaller response scales provide adequate psychometric properties (Reeve, Kitche, Sudweeks, Bell, and Bradshaw, 2011). Nevertheless, the extent to which these measures include appropriate and adequate *level*, *strength*, and *specificity* is unknown.

Conceptual Congruence

Many reviews, articles, and commentaries have expressed concern with self-efficacy research's jingle-jangle fallacies that have continued over the years and led to a widespread lack of discriminant validity (Bandura, 1997, 2006; Bong, 1996; Klassen & Usher, 2010; Pajares, 1996). This has resulted in various definitions and alternative definitions being used throughout self-efficacy research. This is most notably evident in Klassen and Usher's (2010) finding that 51% of the articles they reviewed between 2000 and 2009 were *not* congruent with theory, whereby they conceptualized such by asking, “Do the items ask for an evaluation of confidence to carry out a task, and not competence, intention, skill level, social comparison, self-concept, self-esteem, or outcome expectancies?” (p. 18). Simply; to accurately and appropriately theoretically advance the study of self-efficacy, theoretical alignment must be evident in the path between how a researcher conceptualizes self-efficacy and how they go about measuring it. Without *congruence* or *alignment*, there is no logical means to ensure what is being found or what is measured contributes to what has historically been known as self-efficacy. Therefore, extending upon Klassen and Usher’s (2010) focus, suggests not only the items themselves should be congruent, but the manner in which self-efficacy is discussed and theoretically positioned within an article is paramount. If the conceptualization of self-efficacy is incorrect or misguided, even having correctly worded and focused items *congruent* does little to clarify findings and theoretical contributions.

Correspondence

Lastly, the extent to which a measure aligns with and focuses on relevant and related performance criteria, the greater the predictive validity (Bandura, 1997, 2006; Bandura & Schunk, 1981; Bong, 2002; Pajares et al., 1999; Pajares & Miller, 1995). In reference to Pajares and Miller (1995), Marsh and colleagues (2018) referred to this as the “specificity matching principle,” whereby items are commonly constructed with the same or very similar phrases as the actual test items. Self-efficacy measures as these are referred to as *test-related self-efficacy* (Marsh et al., 2018). To be clear and differentiate from *specificity*, which is more focused on domain, skill, or task details, *correspondence* is concerned with a measure’s alignment with some type of criterial task or performance measure. Criterial alignment, or *correspondence*, has proven to be predictive even after controlling for preexisting differences (Huang, 2013;

Valentine, DuBois, & Cooper, 2004). In that, Valentine and colleagues (2004) found that when self-beliefs where aligned to their academic domain, there is a small positive influence on academic achievement while also controlling for prior achievement across 60 independent samples in 55 longitudinal studies.

Despite the obvious relationship between self-efficacy items that clearly align with test items, this condition can be dubiously deployed. That is, often the predictive power of a non-aligned measure is put into question despite an obvious lack of *correspondence* between the measure and performance measure. Together, *specificity* and *correspondence* clearly enhance the predictive nature of this relationship, however, it is not always required or needed. Therefore, it is, like *specificity*, dependent upon the need and use of the researcher and their prerogative to justly provide rationale and the needed theoretical justification for the particular depth of *correspondence* and how they further situation and explain predictions thereafter.

Obvious Holes, Challenges, Warnings in Self-Efficacy Measurement

Notable self-efficacy researchers have consistently and vehemently contended that without careful attention to details involved in the measurement of self-efficacy, the field is at risk of weakening the theoretical foundation for which it stands (Bandura, 1997, 2006, 2012, 2018; Henson, 2002; Klassen & Usher, 2010; Pajares, 1996; Usher, 2015). Most recently, Bandura (2018) commented in reflection of his work specifically in this area that “Studies were being published with faulty measures and misconceptions of self-efficacy theory” and that to remedy such he published his 2006 work that outlines how to “conduct conceptual analyses to determine the appropriate types of self-efficacy for a given sphere of functioning, and how to scale the items in terms of gradations of challenge” (p. 134). Despite his clear efforts and others over the years (see Pajares, 2006), the review most recently conducted by Klassen and Usher (2010) clearly outlined post 20th century issues surrounding the measurement of self-efficacy.

Klassen and Usher's (2010) work reviewed 96 articles published in prominent educationally related journals between 2000 and 2009 and assessed self-efficacy's measurement across both *congruence* to theory and *specificity*. Assessed on a "global" level, whereby each measure was inspected as a whole as opposed to by individual item, more "influential" (impact factor) journals tended to publish more *congruent* measures, although they all included some level of *non-congruent* measure (Klassen & Usher, 2010, p. 18). Furthermore, they found that measures that lacked *congruence* also often lacked *specificity*, which is vital to aligning the theoretical surmised judgement of specific capability to appropriately worded and interpreted items. Although they provided no definitive metric or quantitative data on how many lacked *specificity*, it did contribute to 51% of the articles being labeled as *not-congruent* with theory. Apart from *congruence* and *specificity*, Klassen and Usher (2010) also noted that no study had *congruently* incorporated or acknowledged a measure of 'collective-efficacy.' That is, measures that seek to capture a group's shared beliefs about their capabilities (Klassen & Usher, 2010; Klassen, Usher, & Bong, 2010). Although a full description and background of collective efficacy, which depicts an individual's perception of groups beliefs, is outside the scope here, it is important to note that a vast majority of studies in the history of self-efficacy research has focused on individual level self-beliefs. Despite echoing past reviewers (Pajares, 1996) "cautions to researchers about problems in faulty conceptualization and measurement of self- and collective-efficacy that continue to pervade research in the field," they offered four main areas for future research to contend: sources of self-efficacy, collective efficacy, cross-cultural, and self-efficacy for self-regulation (Klassen & Usher, 2010, p. 20).

Although Klassen & Usher’s work addressed and included 11 writing self-efficacy measures, it assessed the entire field of self-efficacy in a collective manner. Nevertheless, the most recent substantive reviews of writing self-efficacy specifically were from Klassen (2002) and Pajares (2003). The following section will briefly review both.

Measurement of Writing Self-Efficacy

To best surmise the status of measurement and cover relevant literature for this thesis, the following section will include two main sections. First, I will briefly review literature prior to 2008 (Klassen, 2002 and Pajares, 2003), as the most recent review of writing self-efficacy literature was completed by Pajares in 2003, with non-empirical works helping to fill this gap (Pajares, 2007; Pajares et al., 2007). Second, I will review pertinent literature from the past 10 years and include a brief review of the major substantive themes throughout. The connection between a study’s substantive focus and its mechanism to operationalize it are inextricably interwoven.

Measurement Themes (prior to 2008)

Considering the expansive work that has been done, Klassen’s (2002) work focused on examining writing self-efficacy among adolescent students from 1990 through 1999 across 16 articles. With a focus on measurement, he found common use of self-report scales, with only six including the scale used, two that did not include it at all, a common trend of scales to align with a criterial task, and a range of how “fine-grained” each measures was (Klassen, 2002). Furthermore, Klassen (2002) assessed each article for both *specificity* and *correspondence*. Although they reported that most received high *specificity* and *correspondence* marks, some tended to be broader in scope simply due to their criterial alignment. That is, when the criteria for which performance was measured was more global, so too was the *specificity* required to capture it (Klassen, 2002). This is logical, as a more global approach, such as asking about skills used across a domain or in reference to the domain as a whole, is not specific in and of itself.

Nevertheless, despite more systematic review work in this field, the field is fortunate to have a well-established and regarded basis from which to establish itself. The work by Bandura (1997, 2006) to establish and explicate the role of measurement in self-efficacy is often regarded as a basis to compare measurement items and scales. In Pajares' (2003) synthesis of research of writing self-efficacy, he provided three popular ways in which efficacious beliefs of writing have been measured. First, measurement has focused on students' confidence to execute particular *skills* native to writing (Pajares & Johnson, 1994, 1996; Shell et al., 1989, 1995). For example, items that focus on the confidence to write with certain grammar, verbiage, syntax, or punctuation, such as "correctly punctuate a one-page passage" or "organize sentences into a paragraph so as to clearly express a theme" are found in Pajares and Johnson's (1996, p. 166) work adapted from Shell, Murphy, and Bruning (1989). Second, researchers often focus on writing *tasks* and the confidence in which they have to complete them. Examples include writing a particular length paper, type of genre message, or a certain type of story (Pajares, 2003; Pajares & Johnson, 1994; Shell et al., 1989). As explained in Pajares (2003), the predictive value lies squarely in how aligned such *skills* and *tasks* are to the performance measure. In Pajares and Johnson's (1996) work on undergraduate students whereby they assess *skills* and *tasks'* predictive value, they found that efficacy of *skills* predicted students' skill in composing essays, but *task* self-efficacy did not. This implies, though not explicitly discussed in either Pajares' (2003) review or Pajares and Johnson's (1996) paper, that an undergraduate's skill in composing essays is determined by these particular skills. A more logical and likely appropriate, causal chain of reasoning would be to surmise that such skills, arguably rudimentary, permit a higher level writing that is actually part of collegiate grading criteria. Either way, because this level of detail is not explained, such logical determinants are left to chance and our best guess. The level of *congruence* is not a new complaint among self-efficacy scholars (Bandura, 1997, 2006, 2018; Klassen & Usher, 2010; Pajares, 2003; Usher, 2015). Lastly, scholars have consistently used scales that query students' confidence to earn an A, B, C, or D in their class (Pajares, 1999, 2003; Pajares, Britner & Valiante, 2000). These judgements of confidence are then compared predictively to their actual grade earned in the class. Despite the obvious troubles with comparing subjective grades to self-beliefs, Pajares (2003) reported sound reliability metrics (.86 to .89) in existing literature (Pajares, 1999; Pajares et al., 2000). This relationship may be due in large part to students' individual knowledge of the contextual and environmental conditions that surround their grade, offering a clearer and less subjective alignment of their confidence to what they anticipate their teacher awarding them. Who else is to know better than them as to their foreseen ability, effort, and relationship with the teacher?

Pajares, to whom has likely established the most robust and extensive research program to date, has extended the field primarily with assessing and examining gender differences in efficacy beliefs of writing (Pajares, 2007; Pajares et al., 1999; Pajares & Valiante, 1997, 1999, 2001; Pajares et al., 2007). Nevertheless, his teams have consistently used the *Writing Self-Efficacy Scale* (WSES; Pajares, 2007; Pajares & Valiante, 1999; or a form of it) in an effort to extend Shell's work by making the items more transferable and applicable across developmental ranges (Pajares, 2007). Although he commonly speaks of it as a unidimensional scale, an exploratory factor analysis suggested two factors arranged to represent *skills* and more advanced composition *skills* such as "ending paragraphs with proper conclusions" and "get ideas across in a clear manner by staying focused without getting off topic" (Pajares, 2007, p. 244).

Nevertheless, the initial path of writing self-efficacy measures is not too complicated, yet it does suggest a reliance upon *skill* and *tasks* that may, or may not, fully encompass the breadth of what self-beliefs could tell us about the writing process, cognitive process inherent to writing, or the amount of self-regulation that has been examined to occur throughout the writing process. Therefore, in sum, the status of measurement of writing self-efficacy in or around the early 2000s was still evolving to encompass more, be more directed towards Bandura's guidance, and capture useful information for practitioners and researchers alike to make more informed decisions for student success and towards better understanding the cognitive processes inherent to writing's complexity.

Empirical Studies Investigating Writing Self-Efficacy (after 2007)

Since prior reviews by Pajares (2007) and Klassen (2002), to my knowledge no other systematic or critical review has been conducted focusing on writing self-efficacy. Despite this, ample research has continued that has expanded the purview and scope of writing self-efficacy in the literature. To best capture what research has been conducted over the past 10 years and evaluate the fidelity of measurement, I conducted a systematic literature review that specifically focused on not only measurement uses, advancements, and refinements, but how well each study adhered to Bandura's (1997, 2006) and Klassen and Usher's (2010) work that assessed *specificity*, *level*, *strength*, *congruence,* and *correspondence*. Although a full, in-depth, and fully inclusive review is not warranted herein, relevant components will be presented and discussed below. When applicable, appendix descriptive figures and tables will be referenced.

In all, 60 empirical studies were discovered that measured writing self-efficacy in peer-reviewed journals from the electronic databases of Academic Search Complete, Educational Resources Information Center (ERIC), PychInfo, and Web of Science over the past 10 years. For a more thorough vantage of how many articles were found, excluded, and kept, see Appendix A's PRISMA diagram that depicts the full screening process.

Referencing the descriptive statistics in Appendix B, the 60 articles were comprised of 17 studies that employed an experimental methodology, 16 longitudinal, 28 from the United States, and 7 studies that included students identified with a learning disability. Sample sizes varied, however, only 2 studies included more than 1000 participants. Of these studies, undergraduate samples made up a majority, while elementary samples tended to be more prevalent in the K-12 domain.

Using thematic coding, each study was examined for common domains of study, themes, and topics. Out of 60 studies, 41 involved the relationship between writing self-efficacy and writing performance or achievement in some manner. Not surprising, considering the abundance of gender specific self-beliefs studies (De Smedt et al., 2017; Garcia & Fidalgo, 2008; Graham et al., 2008; Pajares et al., 1999, 2007; Pajares & Valiante, 1999, 2001), 21 studies included some amount of either statistical differentiation or substantive assessment by gender. Also, not surprising, given the preponderance of writing strategy literature in the field of writing research (Graham, McKeown, Kiuhara, & Harris, 2012; Harris & Graham, 2016; MacArthur, Philippakos, & Ianetta, 2015; Santangelo, Harris, & Graham, 2016), 15 studies were focused on some type of strategy use. Studies focusing on anxiety and apprehension to writing (k= 10), examined varying samples almost entirely consisting of undergraduate students (Martinez, Nock, & Cass, 2011; Sanders-Reio, Alexander, Reio, & Newman, 2014; Stewart, Seifert, & Rolheiser, 2015; Vanhille, Gregory, & Corser, 2017; Woodrow, 2011). Furthermore, a growing body of literature examining motivation (k= 8) in writing that commonly incorporates multiple motivational components (e.g. goals, strategies, cognitive mechanisms) examined by newer statistical methods (namely structural equation modeling (SEM)) was found (De Smedt, Graham, & Van Keer, 2018; De Smedt, Merchie, Barendse, Rosseel, Van Keer, & De Naeghel, 2017; Limpo & Alves, 2017; MacArthur, Philippakos, & Graham, 2016; Prat-Sala & Redford, 2010; Troia, Harbaugh, Shankland, Wolbers, & Lawrence, 2013).

Measurement of Writing Self-Efficacy (post 2007)

Across the past ten years, the measures used to capture writing self-efficacy have greatly varied. Of the 60 studies, 21 studies used or created 16 different measures, while 27 studies adapted 15 different measures to their own needs. Only 23 (39%) of the 60 studies included the measure they used in their article for the reader. However, despite the suggestion from Klassen (2002) to include the actual measure, some authors did include sample items that offered some description and view of each measure's characteristics. Therefore, 36 studies were able to at least partially be assessed for *specificity, level, strength, conceptual congruence, and correspondence*.

Measure Focus

Throughout the studies, writing skills and tasks continued to be a focus of measurement (Klassen, 2002). For example, Garcia and Fidalgo (2008) focused their created scale towards writing skills that included the quality of text, ideation, having an understandable logic, mechanics, and spelling and punctuation. Similarly, work by Martinez, Nock, and Cass (2011) focused their created measure on writing tasks like writing short essays, their quality, speed, and efficiency. Capturing writing self-efficacy of self-regulation seemed to also evolve into measurement into the 2000s with both Bruning's measure and work by many other authors thereafter that focused their instruments at capturing efficacious beliefs of self-regulation in relation to writing (Bruning et al., 2013; De Smedt et al., 2016, 2017, 2018; Ekholm et al., 2015; MacArthur et al., 2015; Zumbrunn et al., 2016). Other works, however, continued to focus more on tasks of writing, such as Prat-Sala and colleagues' work (2010, 2012) that captured both reading and writing self-efficacy while focusing primarily on efficacious beliefs of writing essays, such as demonstrating knowledge and providing evidence therein.

Adherence to Bandura's Theoretical Guidance

In all, I assessed each article's *specificity*, *level*, *conceptual alignment*, and *correspondence* (when applicable). Over the past ten years, it appears researchers have improved their use and alignment to theory in terms of their measurement. Collectively, when assessing *specificity, level, conceptual alignment,* and *correspondence* (when applicable), articles from the past 10 years were 53%, 74%, 85%, and 54%, respectively aligned. Although these are an improvement comparative to Klassen and Usher's findings (2010), many studies continue to lack *specificity*, which is problematic because efficacy beliefs are contextualized beliefs that focus on judgements of capabilities to perform some type of skill or task in a specific domain. Without proper *specificity* or *level*, that which a measure purports to measure and to what degree is gravely in question and purely left to the conceptualization of the participant. Nevertheless, a large majority of studies did conceptually align and ask participants to evaluate their confidence, as opposed to other forms of self-beliefs. In terms of *strength*, despite only 38% of the measures using Bandura's suggested rating response scale, it is difficult to find fault with any without further study. Aside from work by Engelhard and Behizadeh (2012) on the WSES, no other writing self-efficacy scale has been psychometrically assessed for proper item functioning (Smith, Wakely, de Druif, & Swartz, 2003). Furthermore, no measure to my knowledge has undergone rigorous psychometric investigations to determine if the proclaimed group differences are truly meaningful, whereby the constructs (writing self-efficacy) are interpreted and conceptualized equally among groups (measurement invariance). Without such analyses, group differences could be completely meaningless and invalid (Horn & McArdle, 1992; Meredith & Teresi, 2006). In relation to groups interpreting latent construct differently, Vandenberg and Lance (2000) stated it “may be tantamount to comparing apples and spark plugs” (p. 9). Nevertheless, the state of writing self-efficacy measurement, at least in my eyes, continues to be flawed and risks limiting and perpetuating a weakly theoretically grounded means of understanding a vital component to human motivation. Of note, however, the review of each article was completed by only myself and may, given another vantage from an additional researcher, be both more robust and substantiated. Despite this limitation, this section has provided an initial view of how the last 10 years of research in writing self-efficacy has adhered to the theoretical guidance provided by Bandura (1997).

Scale Development

An important component to any substantive field is consistent and evolving scale development, especially so when a measure is often predictively assessed to particular criterial tasks, such as self-efficacy (Bandura, 1997, 2006, 2018). In other words, as the use, function, and evolution of writing has consistently changed in the world, especially among K-12 institutions that have consistently evolved their standardized tests to capture a developing world that uses writing in different ways across the curriculum, so too should the measures used to capture efficacious beliefs.

Based on this notion, it was important to capture and explicitly describe here what steps have been taken to advance the measurement of writing self-efficacy in the past 10 years. In other words, which studies have provided, at the least, some type of validation, psychometric analysis, or statistical examination of the properties of the measures used. Therefore, this section will outline those studies which did just this and highlight those that strictly developed and presented new measures.

In total, 18 studies included some form of advanced measure validation, typically consisting of confirming the factor structure data fit through confirmatory factor analysis (CFA). Although this provides factorial validity evidence the data fit a predefined factor structure, it is, often enough in most studies, simply supportive and substantiating in nature. For example, work by De Smedt and colleagues (2016) provided CFA evidence on multiple measures, including writing self-efficacy, to substantiate the data to a predefined model prior to extensive multilevel analyses to predict writing performance. Similarly, work by Limpo and Alves (2017) used CFA to ensure each construct they assessed, fit an 'a priori' factor structure prior to examining how each construct associated by latent path analyses. In a step to using advanced statistical methods in a creative and novel way, Jones (2008) used exploratory factor analyses to compare the conceptual overlap of both self-efficacy and locus of control items. Further, work by Engelhard and Behizadeh (2012) used Rasch measurement theory, a form of item response theory, to assess Pajares, Miller, and Johnson's (1999) WSES scale for psychometric quality, while also examining how related self-efficacy beliefs were to teacher grades, gender differences, and the alignment writing self-efficacy judgements through qualitative inquiry. Collectively, the use of these statistical validative methods has both demonstrated the advent and ease of advanced statistical procedures to assess model data fit and provided researchers with validative evidence across broad samples, contexts, and situations.

Apart from demonstrating a measure's conceptual fit, few studies over the past 10 years have directly focused on presenting a new measure, as opposed to just adapting a previous measure. Even within these studies, trends in the measurement of writing self-efficacy present themselves. For instance, Schmidt and Alexander's (2012) work developed a college-level writing self-efficacy scale, the Post-Secondary Writerly Self-Efficacy Scale (PSWSES), to be specifically used in a university writing center. In doing so, they focused on its function across multiple tutoring sessions to assess the measure's reliability, consistency, factor structure (in which they found three: local and global writing process knowledge, physical reaction, and time/effort), and validity. Together, both MacArthur, Philippakos, and Graham (2016) and Troia and colleagues (2013) took a more global and all-inclusive approach to developing a measure that included multiple forms of motivation, of which self-efficacy was included. In terms of scale development and self-efficacy's place amongst other variables, much can be illuminated. MacArthur and colleagues' (2016) study with college students, which developed a motivational questionnaire that included self-efficacy, achievement goal orientation, beliefs about writing, and affect toward writing sub-scales, found a single factor writing self-efficacy sub-scale that exhibited no significant correlation to writing achievement, however, they did align self-efficacy with the level of class in which students were sampled (either high or low level developmental class). That is, students in lower level classes exhibited lower levels of self-efficacy. Nevertheless, they primarily presented the measure through individual sub-scale exploratory factor analyses. Troia and colleagues (2013) took a multidimensional motivation approach to capturing writing motivation with the "Writing Activity and Motivation Scales." In doing so, they related motivation and activity to writing performance, finding that females and older students performed better, while motivational beliefs, self-efficacy for writing skills and tasks, interest, value, and attributions for writing success, mediated the relationship between specific writing activity and performance (Troia et al., 2013).

Most recently, Bruning and coleagues' (2013) work, to whom drew from early work by MacCarthy, Shell, Pajares, and Zimmerman, sought to shift gears on how writing self-efficacy was conceptualized. In other words, they focused on conceptualizing and measuring writing's "psychological, linguistic, and behavioral challenges" inherent to the writing process, to provide both a theoretically sound, yet pragmatic and useful measure (Bruning et al., 2013, p. 27). In doing so, they sought to capture a writer's judgement of their ability to use common mechanics of writing such as spelling, grammar, and punctuation, generate ideas and use them, and stay focused and control frustration (Bruning et al., 2013). Through two studies therein, one of middle and one high-school students, they assessed the measure's factor structure, yielding a three factor model fit, and, like both the MacArthur and Troia studies, assessed the measure's relationship to other variables, such as liking writing, self-reported writing grades, the state's writing assessment (SWA), and the type of English/language art class enrollment. Of note, their first study, which examined the factor structure of the SEWS, was conducted with middle school students and further validated with high-school students in study two. From study two, they found writing ideation and self-regulation to be significantly related (r = .707) and a strong relation to affect for writing. However, they found a stronger relationship between writing's conventions to the SWA than either ideation or self-regulation and significant differences between the levels of ELA classes and all factors of the SEWS.

Over the past 10 years, the SEWS has been the most used and adapted scale, with 11 studies employing it among various ages, languages, and locations across the globe (see De Smedt et al., 2016, 2017, 2018; Ramos-Villagrasa et al., 2018). Among those adapting the scale, Ekholm, Zumbrunn, and Conklin (2015) reduced the scale from the original 16 items, to 9, in a study to examine the predictive and mediational roles of college students' writing self-efficacy and feedback perceptions on writing self-regulation aptitude. In doing so, results indicated that feedback perceptions partially mediated the relationship between writing self-efficacy and writing self-regulation aptitude. In a similar study that also demonstrated a partially mediated relationship between writing self-efficacy and writing self-regulation, although with middle and high school students, Zumbrunn, Marrs, and Mewborn (2016) further adapted the 9-item version of the SEWS to be more cognitively and developmentally appropriate (see Cowan, 2010; Weil et al., 2013), by replacing the 101-point response scale originally championed by Bandura (1997), with a 1-4 (*Almost never - Almost always*) scale.

Most recently, this adapted version of the SEWS (9-item, 1-4 response scale) has also been further psychometrically assessed through a robust model comparison between a 3-factor model and a 1-factor model using confirmatory factor analyses (Zumbrunn, Broda, Varier, & Conklin, 2019). Additionally, this work also used structural equation modeling (SEM) to assess the predictive relationship between the specific factors of the adapted SEWS and that of both writing achievement and student writing self-regulation. In doing so, Zumbrunn and colleagues (2019) found *conventions,* and only *conventions*, to be significantly predictive for both elementary and high school students' writing grades and writing self-regulation. Also recently, mixed methods work by DeBusk-Lane, Lester, and Zumbrunn (2019) used the adapted SEWS' latent factor scores from a 3-factor confirmatory model in a latent profile analysis. In doing so, they found three profiles of students largely differentiated by level differences (described as *doubtful, average,* and *confident* writers) across the three specific factors of the adapted SEWS. Interestingly, when students' qualitative responses related to their beliefs about writing improvement were explored through the 'sources' of self-efficacy (see Usher & Pajares, 2008), the profiles reported predominantly mastery experiences and feedback as indicators of their improvement, yet largely differed in terms of the kind of mastery experiences and feedback they reported (DeBusk-Lane et al., 2019).

Future Directions and Gaps in the Literature

This section will provide both a brief overview of what common trends, themes, and findings have been found and identify particular gaps inherent to the measurement of writing self-efficacy that need further study.

To summarize, 60 articles across the past 10 years were reviewed. Two main substantive areas were found; writing self-efficacy's relationship with writing performance and gender differences in writing self-efficacy. Within these, the field has focused newer statistical methods and effort on better understanding how similar motivational variables interact and relate with writing self-efficacy and its relationship to performance, but also that this landscape continues to be wildly limited and unknown. Building upon over a decade of progressive work prior that has examined gender, the past 10 years continued this trend well, yet also continues to purport differences that are inconclusive and dynamic across both age and developmental grade.

Although ample research prior to 2008 focused on adolescent participants (see Klassen, 2002), this focus in the last ten years has been fairly limited. Furthermore, very little research has specifically examined early high-school students, especially in the United States. Considering such, there is ample room for future research to examine the current status of adolescent students' writing self-efficacy, especially considering the dynamic changes that have occurred throughout the writing landscape (e.g. computers, writing's use in society, ect.). That is, not only is the landscape in which composing consistently changing, but through the period of adolescence, the use and function of writing dynamically changes across the curriculum (Graham & Perin, 2007; Pajares, 2003). In other words, the amount, way in which it is used, and how it is assessed changes throughout the adolescent developmental period (Applebee & Langer, 2011). Writing increasingly becomes vital in knowledge building and creating connecting networks between disparate forms of information, while also serving as a strong metric that teachers use to judge understanding (Appleebee & Langer, 2011). Without a better understanding how self-eficacy exists, changes, or shifts during this period, educators and researchers alike are limited in how best to approach and foster student motivation towards writing. Given these structural environmental changes, it is vital to continue to better understand how students respond, how their motivation and confidence changes, and how they navigate such changes to best support them, foster writing motivation, and provide opportunities to become better writers. Together, adolescence is a tumultuous developmental period that has presented writing self-efficacy researchers with many disparities worthy of further inspection (e.g. opposing gender differences, weakened relationship between efficacy and writing achievement, declining strength of writing self-efficacy). Therefore, although adolescence has previously been a focus of writing self-efficacy research, the dynamic changes over the past ten years greatly warrants further research.

Considering the degree to which these studies, and presumably those that came before, adhere to the theoretical guidance Bandura (1997) provided, it is surprising little research has aimed at determining to what degree measures, at any level of *specificity*, capture both domain specific variability and some type of global sense of efficacy. I single out *specificity* because it is consistently a point of contention among scholars (Bandura, 1997, 2006; Klassen & Usher, 2010; Marsh & Hau, 2007; Marsh et al., 2018). Understandably, the assessment of a measure's *specificity* is subjective and debatable on many levels, while no attempts have been made to further assess a measure's ability to capture a particular form of *specificity* or to better understand how it portrays it. Said another way, critical assessments of measurement have strictly focused on obvious degrees of *specificity*, but not precisely how a measures is able to capture both a global sense and domain specific facets of its particular scope. This gap or line of inquiry aligns well with Marsh and colleagues' (2018) directions of future research that called for a more nuanced understanding and testing of multidimensional measures of self-efficacy that may exhibit hierarchical trends similar to self-concept. Simply, it is vital to ensure how we measure writing self-efficacy is fully understood prior to extending its use to examine much needed substantive gaps still left to explore. Without a solidified grasp of measurement, faulty interpretations, jingle-jangle fallacies, and theoretical missteps can be expected.

That said, to my knowledge, no scale has been psychometrically assessed with factor structures other than common confirmatory factor analyses. Considering the studies over the entire body of literature, a multifactorial depiction of writing self-efficacy has commonly been used, however, a number of cases have found or portrayed it as a unidimensional construct. For example, works by Shell and colleagues (1989) developed task related scales (e.g. essay, novels) and skills related scales (e.g. spelling, punctuation), while Pajares, across multiple lines of research, has commonly operationalized writing self-efficacy as unidimentional (Pajares, 1996; Pajares & Valiante, 2006), he also identified multidimensional cases related to skills (e.g. spelling, grammar) and behaviors while composing (e.g. structuring paragraphs) (Pajares, 2007). A single factor (unidimensional) take has also been found by Zimmerman and Bandura (1994) that sought to capture a writer's strategic efficacy through three areas of planning, organizing, and revision. Now, it must be said, many of these scales have conceptualized, and therefore operationalized, self-efficacy in different ways and assessed beliefs on different skills, tasks, and cognitive processes inherent to writing. This, among other conflating factors, has muddied the waters in fully understanding both the scope of writing self-efficacy and its dimensionality.

Furthermore, the degree to which each scale accurately captures *specificity* inherent to the scale's intention and premise has yet to be fully discussed or examined in the field. Newer statistical methods enable researchers to more accurately depict the extent to which a multidimensional measure captures a global construct or targets more focused facets of a domain (specific factors). This is not to say or argue that measures should be more or less specific, but that there are advanced statistical methods that permit a closer inspection than has been historically presented in the literature to date.

Considering such, Bruning and colleagues' (2013) original and adapted SEWS (see Zumbrunn et al., 2016) has been commonly depicted as a multidimensional three factor scale (DeBusk-Lane et al., 2019; De Smedt et al., 2016, 2017, 2018; Zumbrunn et al., 2019), however, the adapted version has also been depicted, through item reduction, unidimensionally (Ekholm et al., 2015; Zumbrunn et al., 2016). Most recently, work by Zumbrunn and colleagues (2019) extensively compared a three factor model against a one factor model of the adapted SEWS. In doing so, the three factor model presented much better data model fit and demonstrated invariance between elementary and high-school students. Similarly, taking some of the items from the original SEWS, MacArthur and colleagues (2016), who conceptualized and operationalized writing self-efficacy to capture writing tasks, strategies, and self-regulation (see MacArthur et al., 2016), initially extracted two factors, yet also identified a strong first eigenvalue, which may suggest the presence of a general factor relatable with all items. Furthermore, shared loadings across both factors led them to only fit a one-factor model that explained 55% of the variance. Similarly, recent work by Graham and colleagues (2017) using 11 items (of 13 total to capture writing self-efficacy) from the original SEWS, sought to examine the factor structure of a questionnaire focused to capture a writer's strategic approach to writing, attitude toward writing, and self-efficacy for writing, found an ill-fitting three factor model (one factor per area) that suggested (through modification indices) three *convention* items covary within the self-efficacy scale, which may indicate some level of multidimensionality (Brown, 2015; Kline, 2016). In work to extend the original SEWS to the Spanish language, Ramos-Villagrasa and colleagues (2018) presented a rather extensive item analysis that included EFA and factor correlations. In doing so, the EFA had a strong first factor eigenvalue that explained 48.65% of the variance and factor correlations that ranged from .50 to .63. Similarly, DeBusk-Lane and colleague's (2019) presented the adapted SEWS' latent factor correlations ranging from .44 to .89. Furthermore, the work by Ramos-Villagrasa and colleagues (2018) also demonstrated Pajares' (2001) unidimensional *Self-efficacy for Writing* scale (translated into Spanish) significant correlated to all three factors of the original SEWS. Similar work by Limpo and Alves (2017) with a Portuguese translation of the original SEWS found similar factor correlations and an extremely good three factor model data fit (e.g. CFI = .992).

These trends are quite understandable, as a great majority of measurement instruments are developed and designed to capture multiple areas or facets of a given construct or domain of focus (e.g. self-efficacy -> self-regulation, conventions, ideation; engagement -> social, cognitive, behavioral, affective). Considering the SEWS was specifically constructed to do just that, it is plausible that such closely related domains likely exhibit variability between the indicators that is described by some amount of a hierarchical or global facet. Furthermore, across the studies that have employed the adapted SEWS, clear dimensionality clues (varying best fitting factors, 1 or 3, and correlated latent factors) suggest further inspection is warranted and needed, especially considering its use in SEM and as a pragmatic developmentally appropriate tool for educators. Together, these assumptions and the presence of suggestive indications of some type of hierarchically organized instrument and/or global/general construct of the adapted SEWS necessitate further inspection of the adapted SEWS to what is referred to as construct-relevant multidimensionality.

The following sections will present how, through the use of a substantive-methodological synergy, a full inspection and analysis can be fostered to examine whether the adapted SEWS exhibits construct-relevant multidimensionality. First, I will present how such an approach can influence theory by focusing on a substantively important topic--writing self-efficacy. Second, I will present how unique, novel, and newer statistical methods can appropriately and rigorously examine both sources of construct-relevant psychometric multidimensionality.

Substantive-Methodological Synergy

In seeking a response to both the measurement issues outline above and the calls for future research, the field is primed to extend what has been done with newer statistical methodologies and approaches that can further examine construct-relevant multidimensionality, and continue to refine and mold the theoretical basis for what we know about self-efficacy. In a prolific introduction to Contemporary Educational Psychology in 2007, Herbert Marsh and Kit-Tai Hau presented and established a directive that thrust Educational Psychology towards a pursuit of research that employs advanced methodological tools towards answering important substantive applied issues--"substantive-methodological synergies." In other words, through a synergistic relationship of creative and unique applications of advanced statistical methods, vital and important substantive issues can be further examined and explored. The field of writing self-efficacy, as I see it, is prime for this approach and would do well to examine established, yet clearly undefined and ambiguous findings, in relation to scale structure and dimensionality and even the extent to which particular domain specific measures actually portray domain *specificity* as defined by Bandura (1997). To be explicitly clear, this is not to say current measurements are are not important or in question to theory, but perhaps a better understanding of how they operate, construe *specificity,* and relate to other constructs can provide a pathway to a better and more theoretically grounded instrument.

Although particular methodological advancements enable us to statistically capture general or hierarchical concepts inherent to multidimensional scales, this concept in the theoretical positioning of self-efficacy is nothing new. In fact, Bandura (1997) describes it well and clearly: “A multidimensional approach does not mean that there is no structure or generality to efficacy beliefs.” (p. 50). In aligning this statement to social cognitive theory and its depiction of human adaptive functioning, Bandura described six processes in which perpetuate, and therefore we should anticipate, the production of generality within domain specific measures. First, any multitude of activities, skills, or tasks likely include, to some extent, *similar subskills.* That is, in learning or developing there is an application of similar, and often familiar, aspects that enable a transfer of perceived efficacy derived from prior experience and reflective thought, as such beliefs are not “simply a disjoined collection of specific self-beliefs.” (p. 51). Next, *codevelopment* is described as another process that promulgates generality. Especially evident in educational contexts, the development of disparate skills, such as mathematics and music, can provide cross-over influence in perceived self-efficacy, whereby focused development that leads to high efficacy in one domain transfer to another simply by it occurring. Furthermore, *self-regulatory skills* provide ample footing and regulative processes that transfer amongst differentiated skill domains. For example, efficacious beliefs of learning, often developed from mastering a number of difficult skills across domains, can be incorporated into self-appraisals and therefore be a source of generality. Similarly, *generalizable coping skills*, of which are common approaches to dealing with and controlling threats, can influence one's approach, and therefore beliefs, to new or foreign situations. Finally, Bandura (1997) describe *transformational restructuring of efficacy beliefs*, which depict the power of mastery experiences that illicit and 'transform' efficacy beliefs. That is, great successes and experiences perpetuates the generalization of a belief that "one can mobilize whatever effort it takes to succeed in different undertakings." (Bandura, 1997, p. 53).

Although these processes, which instigate and permit generalizations of efficacious beliefs, likely in and of themselves overlap and are obviously similar, it is within this logic that instruments like the SEWS may statistically exhibit construct-relevant multidimensionality. For example, when the SEWS queries a self-appraisal about one's ability to write a "complete sentence," it would be conceptually appropriate to anticipate it to illicit beliefs also associated with appraisals related to one's ability to "concentrate" or "keep writing when it is difficult." These actions are, as Bandura explained, related among deeper generalized skills, cognitive processes, and behaviors that exhibit conceptual overlap between efficacious beliefs. Furthermore, because the SEWS is focused upon three inherent dimensions of the writing process, it is likely they hold commonalities based upon individual student experiences that collectively influence each generally. Therefore, it is plausible to expect such and there may be great utility in capturing a hierarchical or general factor such that reflects "the full range of task demands within the activity domain," while also clarifying specific factor differences (Bandura, 1997, p. 52). To this end, a substantive-methodological synergy focused to inform theory, particularly involving the degree to which a multi-factorial instrument models or exhibits construct-relevant multidimensionality, is of great utility, timely, and an appropriate pathway to forge the field forward.

Construct-relevant Psychometric Multidimensionality

As alluded to earlier, there are two areas relevant to examine; namely, the need to clarify the SEWS' factor structure and to determine to extent to which the items exhibit true score association with non-a-priori factor constructs (Morin et al., 2016). Considering the SEWS and a majority of writing self-efficacy scales are portrayed as multifactorial and presented as a set of conceptually related, yet distinct, and correlated factors (ICM-CFA), the aforementioned evidence highlights a series of potentially critical questions for the measurement of writing self-efficacy: (a) whether individual specific factors illicit meaningful specificity over and above a global construct, (b) whether such a global construct exists alone with these specificities included, or (c) whether such specific factors exhibit distinct correlated constructs without this global foundation (Morin et al., 2017).

To sufficiently establish the methodological footing needed to present this thesis, this section will aim to explain the background and methodological premise of both the variable- and person-centered approaches.

A Variable-Centered Perspective

At its core, CFA is depicted as an Independent Cluster Model (ICM), whereby each item targets its respective conceptually aligned latent factor while fixing all other cross-loadings to exactly zero. Within multidimensional measures, this assumption is both theoretically and statistically difficult to establish (Marsh, Ludtke, et al., 2010; Marsh et al., 2009; McCrae, Zonderman, Costa, Bond, & Paunonen, 1996).

Theoretically, and drawing from what has already been explained, it is a conceptual leap to suggest a multi-factorial scale employed to specifically capture clustered or distinct like constructs within a particular latent domain are, due to the nature of fixed item factor relationships, not conceptually related. Such individual specific factors likely conceptually overlap and it is often the actual intent of practitioners to capture related subdomains to portray a given latent domain. Although historically convenient to portray such relationships in CFA, it does not fully align with reality or common describe it theoretically (e.g. domains are connected, have overlap).

Statistically, the ICM-CFA contends each item to be associated to only one source of true score variance, the factor (or factors). Comparatively, however, Classical Test Theory (CTT) and similarly Generalizability Theory (see Brennan, 2010), contends there are multiple forms of score variance: random measurement error, construct-irrelevant sources of true score variance (validity), and construct-relevant sources of true score variance (Morin et al., 2016; 2017). That is, random measurement error, which is often depicted by measures of reliability and often described as natural or innate fluxuations inherent to the measures themselves, construct-irrelevant true score variation, which is "…excess reliable variance that is irrelevant to the interpreted construct.", and construct-relevant true score variance, which can be depicted as variation inherent to those answering the questions such as aptitude or experience, can be expected to influence true scores (Messick, 1989, p. 13). In an ICM-CFA, random measurement error contributes to the uniqueness of the indicator itself, while construct-relevant sources of true score variation contributes to factor loadings. Although in unidimensional models, construct-irrelevant sources contribute to item uniqueness, in multidimensional models that have conceptually related factors (that are correlated), construct-irrelevant sources of true score variance that depict actual true associations between items (e.g. conceptual overlap) is required to be absorbed into fixed item-factor associations, therefore biasing factor correlations (Morin et al., 2016; 2017). Reviewing statistical simulation studies, Asparouhov, Muthen, and Morin

(2015) found that when cross-loadings even as small as .100 exist in population models, relying on traditional CFAs ICM item-factor associations results in significant biased estimates of factor correlations. This prompted Morin and colleagues (2016) to posit that at least two sources of construct-relevant psychometric multidimensionality are not captured by traditional ICM-CFAs, and therefore bias estimated parameters.

First, multidimensional measures may exhibit a hierarchical or global relation whereby all items associate with their own domain specific factor, as well as a global or hierarchical construct. Furthermore, as described earlier, items may have conceptual alignment to conceptually related factors and due to the fallible nature of ICM-CFA, relate to more than one factor. Neither of which are explicitly captured or modeling in traditional ICM-CFAs. This differentiates to more well-known forms of construct-irrelevant psychometric multidimensionality, such as methods-effects or reversed item effects, that are often captured by the addition of special methods factors (Marsh, Scalas, & Nagengast, 2010). Nevertheless, this, most obviously, has implications for the large amount of recent structural equation modeling work that has been completed and published focused on the facets of the adapted SEWS and other motivational constructs.

Second, multidimensional measures may also exhibit dispersed degrees of true score association between items and non-targeted latent factors. Within the ICM-CFA framework, this true score association is fixed to a priori factor associations and absorbed within the aligned latent factor. As explained earlier, it is likely a lofty assumption to believe items aligned to given conceptual factors do not, at least minimally, cross-load on related latent factors in a multidimensional frame among domain related factors.

Newer methodological approaches have been developed and proposed, or as Reise (2012) stated, “rediscovered,” such as exploratory structural equation modeling (ESEM) and bifactor models (or both) (Asparouhov & Muthen, 2009; Marsh, & Nagengast, & Morin, 2013, Morin et al., 2016, 2017). The following sections will further describe how these two approaches can be employed to best capture construct-relevant psychometric multidimensionality and how their findings may contribute to theory.

Hierarchically ordered and global constructs. The question that most obviously relates to arguments from Bandura (1997) and a host of others (see Usher & Klassen, 2010; Pajares, 2007) throughout the history of self-efficacy work, regards the extent to which a measure is focused upon a given domain, how focused it is, its *specificity*. To be clear, and to overtly situate this presentation, I am not arguing that measures should be more general or that they should be more domain specific to capture efficacious beliefs. The present effort is directly focused on capturing a hypothesized global or hierarchical construct inherent among the multiple dimensions included in the SEWS. There is a clear foundation and collective theoretical opinion that measures should be specific, domain focused, and arranged such that they capture the full extent of the inherent difficulties and range of the skills or tasks in which they are employed (Bandura, 1997, 2006, 2018). To this end, the presence of a global or hierarchical construct may simply better model the data and extensively aid in examining differences among the facets or dimensions of writing self-efficacy.

First, a higher-order CFA, often referred to as a second-order CFA (although there can be any number of higher-orders), is specified such that each item is aligned to only its conceptually aligned factor (just like a CFA), and each factor is then specified to load on a common factor or second-order factor. This second order factor, which is defined by the first-order dimension latent factors, reflects the correlation exhibited among the first-order factors (e.g. Rindskopf & Rose, 1988). Although seemingly useful and still within the ICM-CFA framework, second-order factor models, or higher-order models in general, only re-parameterize and model earlier-order factor correlations, which results in a limited reconceptualization of the relationships between the first-order factors. Nevertheless, the proportionality restraints inherent to ICM-CFA are further compounded, as inflated or biased first-order factor correlations further saturate the higher-order factor, making external uses and interpretability limited and in question. Despite this, the interpretation of the second-order factor can be meaningful, as it simply represents a collective representation of the first-order factors while the first-order variances (commonly defined as disturbances) resemble domain specific factors (Chen, West, & Sousa, 2006). Nevertheless, in the event a domain specific factor only reflects a general factor, as opposed to a specific factor over and above a general factor (or once the general association is removed), it may not easily be detected in a second-order model, as the second-order factor model will naturally absorb this into the first-order disturbance (although it may present as a non-significant first-order disturbance). On the other hand, such an occasion in a bifactor model will cause model convergence issues, whereas it is not likely to cause any model fit problems in a second-order model. This trend continues further with difficulty in using second-order disturbances (specific factor variances) in SEM models (see nonstandard SEM models in Bentler, 1990; Gustafsson & Balke, 1993) and examining measurement invariance of domain specific factors is not possible (Chen et al., 2006). Although, as will be seen, second-order factor models are limited in clearly distinguishing domain specific factors over and above a general factor and similarly expressing the general factor apart from the domain specific factors, bifactor models are much less limited and provide much more information, seemingly answering all of the aforementioned limitations inherent to second-order factor models.

Bifactor models are traditionally estimated such that all items simultaneously load on a global factor and on their a-priori construct related specific factors (Chen et al., 2006; Reise, Moore, & Haviland, 2010). In the ICM framework, all items are freely estimated on the general and specific factors, while fixing cross-loadings to exactly zero. For interpretational purposes, all latent factors are estimated as orthogonal, as this allows the covariance to be partitioned such that it is absorbed by the general factor representing all the items while the specific factors explain the residual covariance therefore not explained by the general factor. This unique distribution of covariance allows for an easy interpretation and further allows for both measurement invariance analyses and each latent factor to be used as either a predictor or outcome in SEM models (Reise, 2012). Therefore, the global factor represents a unitary construct inherent to all items, while the specific factors express “meaningful specificities” over and above a commonly held construct among the items (Morin et al., 2017, p. 397).

It should be noted, to my knowledge, there have been no models other than traditional CFAs estimated in the domain of writing self-efficacy research. Furthermore, only one published work could be found that examined self-efficacy as a bifactor model (Török, Tóth-Király, Bőthe, & Orosz, 2017). Although from organizational research, their examination of the Career Decision Self-Efficacy Scale Short Form clearly demonstrated the utility in using a bifactor model and further illuminated how similar specific factors can be better modeled by incorporating a general factor.

Nevertheless, examining the extent to which either a second-order or bifactor model best depicts the data captures the first type of construct-relevant psychometric multidimensionality, which hypothesizes the presence of a global or hierarchically oriented conceptual relation among all items or the factors themselves. Although a step forward in examining the dimensionality, these models are still hinged on using the restricted ICM-CFA framework, which may limit the true-score association cross factor and inherently has biased factor correlations.

Conceptually related constructs. The second source of construct-relevant multidimensionality commonly not captured within the ICM-CFA framework is that the items fixed to a priori conceptually aligned factors do not always capture the full extent of true score association in the model. That is, due to the fallible nature of indicator items and the given presence of conceptually related constructs between factors, it is expected that items will naturally, to some extent, be conceptually related to other similar latent factors. In these cases, and within the ICM-CFA framework, the true score association is forced to be absorbed by the fixed factor, resulting in inflated factor correlations. To capture true score association, recent methods have integrated exploratory factor analysis (EFA) with CFA and SEM, referred to as exploratory structural equation modeling (ESEM) (Marsh et al., 2014; Morin et al., 2013). Furthermore, a semi-confirmatory approach has also recently been integrated that allows ESEM models to employ target rotation (Asparouhov & Muthen, 2009). Therefore, an ESEM allows all items to cross-load, similar to EFA, while target rotation allows for the pre-specification of target item factor loadings to freely vary and non-target items to originate from zero (be as close as possible) in both traditional first-order factor models and, most recently, in bifactor models (Morin et al., 2017; Reise, 2012). For example, a bifactor ESEM model is similarly interpreted as an ICM-CFA bifactor model, with the added benefit of allowing cross-loadings among the specific factors to better or more accurately model real-world conceptual overlap. Recent research has highlighted this fact, as ignoring cross-loadings in population models leads to inflated general factors, while similarly ignoring generalized factors results in inflated cross-loadings (Morin et al., 2016). Therefore, without examining the extent to which these models accurate depict the data, traditional ICM-CFA models can clearly be limited and may not accurately reflect the latent construct they are intended to portray.

To recap, this variable centered approach follows Morin and colleagues' (2016) framework to identify often overlooked and unexplored sources of construct-relevant psychometric multidimensionality through a series of model comparisons. Drawing from a vast history of rudimentary, though advanced for their time, measurement work, the field of writing self-efficacy, and self-efficacy in general, can greatly benefit from a deeper examination and rigorous approach to model data fit. Following Morin and colleagues' (2016) framework and determining a model that best fits the data, understanding and modeling how students disaggregate into groups, what predicts students into such common response trends, and the extent to which these groups differ on a number of outcomes if greatly beneficial to both validating the SEWS and helping push the field forward by establishing an alternative, yet complementary, approach to modeling multidimensionality.

A Person-Centered Perspective

The previous variable-centered approach assumes a population that is homogenous or in which relationships between variables is said to hold for all members within a population (Laursen & Hoff, 2006; Masyn, 2013; Morin & Marsh, 2015), differentiates from a person-centered approach, that assumes the population is heterogeneous and may contain any number of subpopulations. Although different, they are not, as many have alluded and portrayed, disparate camps or the antithesis of the other--they are complementary, dynamically useful in their own unique vantage, and ultimately draw upon the same data, and "provide alternative views of the same reality." (Morin et al., 2017, p. 400). That is, throughout the entire person-centered approach, I will discuss profiles in aggregate, often as a set of means, compared to means, and, of course, compared the profiles through variable-centered analyses. To further connect the two approaches, Bauer and Curran (2004) may have said it best; "the common factor model decomposes the covariances to highlight relationships among the variables, whereas the latent profile model decomposes the covariances to highlight relationships among individuals." (p. 6). Latent profile analysis (LPA) seeks to groups participants by common response trends among a set of input variables within a probabilistic model-based framework (Masyn, 2013; Nylund-Gibson & Choi, 2018; Lubke & Muthen, 2005; Morin & Marsh, 2015). This provides a certain granularity view of typology of participants such that can then be described both quantitatively, by their input variable profile descriptives, and qualitatively, as they differ between profiles.

As described in Morin and colleagues' (2017) synergistic example, accurately capturing and modeling construct-relevant multidimensionality (hierarchical or global factors and cross-loadings with ESEM) can then be better disaggregated through person-centered analyses to best understand which profiles exist, how participants differ therein, and ultimately provide a clearer vantage of domain specific differences inherent to the sample. Furthermore, a person-centered approach, at least herein, can specifically provide two general purposes; to validate the SEWS across varying profiles and provide a more nuanced understanding of how students exhibit particular dimensions of writing self-efficacy, potentially over and above a hypothesized generally held efficacious belief towards writing. These can then both aid in advancing theory, as understand differences or subpopulations within the population may deviate from commonly held theoretical assumptions and findings (e.g. writing self-efficacy is inversely related to writing anxiety, when it may only be true for certain profiles), and provide practitioners and researchers alike a detailed view of those who may most be at risk within the sample or that could most benefit from targeted writing interventions.

Potential shape versus level effects. Important to note and understand, by accurately depicting and examining construct-relevant multidimensionality, it is possible to more accurately depict both *shape* and *level* effects in the LPA. Here, *level* effects are defined as differences between profiles, often described as being high, medium, or low, on all indicators, whereas *shape* effects are the "tendency for a given person or profile to have a distinct pattern of indicators on which they are high, medium, or low." (Morin & Marsh, 2015). As will be described in Chapter 3, I will conduct both a statistical and substantive assessment to determine the correct number of profiles that also entails assessing the heuristic value, theoretical conformity, and generalizability to new samples (Marsh, Ludtke, Tautwein, & Morin, 2009, Morin & Marsh, 2015).

Commonly, one of the most frequent criticisms associated with conducting person-centered approaches is how they depict the data differently or provide value over and above common variable-centered methods. Although a detailed account of this argument among scholars is outside the scope of this thesis (see Morin & Marsh for a review, p. 40), the essence is that both approaches use the same underlying covariance structure and without particular reason to believe a person-centered approach would add heuristic value, a variable-centered approach would be sufficient. That is, without sufficient evidence of qualitative *shape* effects among the profiles, there is little value using LPA to simply display *level* effects (e.g. high, medium, and low among all indicator items within profile) (Morin & Marsh, 2015). Therefore, the theoretical backing and plausibility that a general factor exists, permits *level* effects to be uniquely modeled as an indicator item, highlighting domain specific *shape* effects over and above global or *level* effects (Morin et al., 2016). To be sure, although this is simply hypothesized, in the event a bifactor (bifactor CFA or bifactor ESEM) model is the best model, also enables researchers to identify domain specific item responses that parse out domain general or, in this case, globally held efficacious beliefs of writing. This perspective can aid in understanding how groups of participants differ, their peculiarities, and how their beliefs demonstrate unique belief structures as a function of their probabilistic membership in groups defined by common response trends. Nevertheless, in the event a global construct is not present among the variable-centered models, I will use the results from either the CFA or ESEM representation, as domain specific differences will naturally be evident with the absence of a global factor (Morin et al., 2016). Either way, and as Morin and colleagues (2016) stated, this squarely reinforces the importance and benefit of establishing a proper variable-centered measurement model prior to conducting person-centered approaches.

Validity components. Integral to the premise and utility of person-centered approaches is their ability to also disaggregate validative predictors among profiles. Herein, I will employ two means of validating the SEWS, aside from common correlations. This effort will include both predictors and outcomes associated with the profiles. Doing so will provide both a meaningful interpretation of the profiles, based upon prior works that have related self-efficacy to other constructs, and continue to build measurement validity to demonstrate the profiles have (a) heuristic value, (b) theoretical alignment or value, (c) anticipated and meaningful relationships to covariates, and (d) generalize, over subsequent replications to new samples (Bauer, 2007; Marsh et al., 2009; Meyer & Morin, 2016; Morin, Morizot, Boudrias, & Madore, 2011; Muthén & Asparouhov, 2008).

To establish further criterion-related validity evidence of the SEWS, I will use both Writing Self-Efficacy Scale (WSES; Pajares, 2007) and the Writing Apprehension Scale - 12 (WAS-12; Limpo, 2018). This will provide both concurrent and divergent/discriminant validity evidence. To be sure, although the prediction of membership into differentiated profiles does not provide common or traditional forms of such associations, it does provide ample forms of evidence to build upon. Although examining how predictive, say, scores on the WSES are to profiles of SEWS responses, is not a traditional form or way to assess such an anticipated relationship, it does, based upon the history of the WSES and its own conceptual alignment, provide a nuanced and unique look at how it aligns. In other words, because both the SEWS and the WSES purport to capture beliefs inherent to writing grammar, punctuation, and self-regulative skills, it is expected that profiles of the SEWS that demonstrate stronger beliefs would be positively predicted by higher scores on the WSES. This relationship will further be validated by assessing how each profile does in terms of writing performance. Drawing from the long history of the items within the WSES across various samples spanning almost 30 years (see Pajares & Valiante, 1999, 2001; Pajares et al., 2001; Shell et al., 1989), and where Pajares (2007) formally introduced and presented the underlying factor structure along with its own construct validity, it has a well-established validated record among the extant literature. Notably, it is positively aligned with both self-regulation and writing achievement, which would presuppose its hypothesized relationship herein (Pajares, 2007). Furthermore, this scale is the only scale to my knowledge that has undergone further psychometric analyses with Rasch measurement analyses, also purporting a positive relationship to writing achievement (teacher grades). Therefore, there is reason to believe a positive predictive relationship between the SEWS and the WSES and a positive relationship between higher efficacious profiles and writing performance.

To demonstrate divergent validity, I intend to use the WAS-12 (Limpo, 2018), as writing apprehension and anxiety have long been established to have an inverse relationship to writing self-efficacy (Klassen, 2002; Klassen & Usher, 2010, Pajares, 2007; Pajares, Johnson, & Usher, 2007). Writing apprehension is defined as “a person’s general tendencies to approach or avoid situations perceived to demand writing accompanied by some amount of evaluation” (Daly, 1978, p. 327). I hypothesize that those with greater writing apprehension would be more likely to be found in less efficacious profiles. This relationship is theoretically aligned, as efficacious beliefs portray confidence and the strength of self-belief, seeming inverse to avoidance (Bandura, 1997; Limpo, 2018; Pajares & Valiante, 1997). Nevertheless, much like the relationship between the WSES and the SEWS, this relationship could be differentially explained within the LPA such that a more nuanced understanding of this relationship may be exhibited between profiles. Although recently published (Limpo, 2018), the WAS-12 demonstrated adequate initial psychometrics and retained its given factor structure of two factors conceptually related to writing ‘affect’ and ‘concern.’

Conceptual Framework

Collectively, the theoretical, measurement, and methodological literature and foundation explained here will serve as a guide to the conceptual framework of this thesis, which ultimately seeks to examine construct-relevant psychometric multidimensionality in the SEWS (Morin et al., 2016). The present conceptual framework is derived from a robust analysis of the literature that identified ample room and conceptual space to examine and investigate the deeper psychometric properties of the adapted SEWS, how the concept of self-efficacy is modeled, and how that precisely translates to existing theory. Herein, theory supports the contention that the adapted SEWS captures and may model a hierarchical or general factor and likely exhibits related concepts between factors, both of which would be beneficial to explicitly model (Bandura, 1997; Pajares, 2006). This assumption is predicated upon both extant literature, which consistently depicts correlated factors and conceptually related specific factors of the adapted SEWS, and the specific theoretical positioning of self-efficacy as a domain specific belief that can be depicted by specifics inherent to particular domains, yet also have somewhat relative and generalized beliefs surround that domain. Bandura (1997) was clear in contending that "Domain particularity does not necessarily mean behavioral specificity," for which there commonly exists a multitude of behaviors, skills, and tasks in which individuals partake within a given domain (p. 49). Therefore, a multidimensional measure can "reveal the patterning and degree of generality of people's sense of personal efficacy." To best model the data, it is plausible the traditional (ICM-CFA) depiction of the adapted SEWS includes construct-related components to true score variation and is therefore limited and likely biased, by not acknowledging any degree of generality existent in the data. Following the framework provided by Marsh and Hau (2007), the theoretically supported hypothesis of both a hierarchical or global structure and overlapping conceptual factors inherent in the adapted SEWS is prime for a substantive methodological synergy. This research uses evolving statistical approaches targeted and applied to substantively important research questions. As such, this work has broad relevance to the motivational literacy field and practitioners alike by providing a robust use of variable- and person-centered analyses that can be used to more accurately and theoretically depict efficacious beliefs involving writing and the relationships it has to other motivational characteristics. Therefore, based upon this conceptual framework and the reviewed theory and variable relationships, I propose the following research questions and hypotheses:

1. Are the items of the SEWS conceptually related across a priori factors?

2. Does the SEWS exhibit hierarchically-ordered constructs?
3. What specific quantitative profiles of writing self-efficacy emerge?
4. What forms of validity evidence is found for the profiles of the SEWS?
    a. Do the profiles exhibit concurrent validity evidence based on responses to the WSES?
    b. Do the profiles exhibit divergent/discriminant validity evidence based on responses to the Writing Apprehension Scale (WAS-12)?
    c. Do the profiles exhibit predictive validity?

I hypothesize the following:

1. Based on both preliminary work in this area (DeBusk-Lane, Lester, & Zumbrun, 2018; Zumbrunn et al., 2019) and a number of studies demonstrating latent factor correlations (Bruning et al., 2013; Limpo & Alves, 2017; Ramos-Villagrasa et al., 2018), I hypothesize that the adapted SEWS will depict a global construct that represents writing self-efficacy as a product of the three domain specific factors together.
2. Drawing from both the conceptual framework for which the original SEWS was derived (see Bruning et al., 2013) and the degree to which the individual domain specific factors are conceptually and theoretically related, I hypothesize items will cross-load and improve model fit.
3. Because no published study has examined a person-centered approach to examine profiles of the SEWS, a limited hypothesis will be provided. Based upon preliminary person-centered work using the SEWS (see DeBusk-Lane et al., 2018), three profiles likely exist without taking into account a general factor. Furthermore, based upon these analyses, the three profiles exhibit clear *level* differences and some degree of *shape*

effects. Based upon my experience with other bifactor ESEM models and using their factor scores for person-centered analyses, I would posit at least three profiles to exist and the *shape* effects to be far more pronounced and evident amongst the profiles.

4.1. Because both the adapted SEWS and the WSES purport to capture beliefs inherent to writing grammar, punctuation, and self-regulative skills, it is expected that profiles of the adapted SEWS that demonstrate stronger beliefs would be positively predicted by higher scores on the WSES, especially among items and factors that are conceptually similar. Although, it should be noted, the WSES does not explicitly seek to capture efficacious beliefs of ideation and may be less related to the adapted SEWS in this facet, however, collectively, the scores should provide predictive utility for validative purposes.

4.2. Drawing from ample literature that purports an inverse relationship between writing anxiety/apprehension and writing self-efficacy (Chen & Lin, 2009; Goodman & Cirka, 2009; Martinez et al., 2011; Stewart et al., 2015; Klassen, 2002; Klassen & Usher, 2010, Pajares, 2007; Pajares et al., 2007), I hypothesize a similar relationship will be found here. As mentioned earlier, because no person-centered work exists of the SEWS to date, other profile relationships may exist aside from an anticipated negative linear relationship. That is, lower profiles or profiles that exhibit unique *shape* effects between the adapted SEWS' indicators may have either weaker predictive relationships or not hold to commonly found variable-centered findings between anxiety and efficacy. Furthermore, other interactive relationships may exist (e.g. sex, grade, ethnicity) that differentially could play a role in also providing validative evidence.

4.3. The relationship between writing self-efficacy and writing achievement is well established across various samples and instruments throughout the last 40 years of research (Klassen, 2002; McCarthy et al., 1985; Pajares, 2003; Pajares & Johnson, 1996; Pajares & Valiante, 1997, 1999; Shell, Murphy, & Bruning, 1989; Zimmerman & Bandura, 1994). Therefore, it is plausible to expect a similar finding among the profiles that exhibit stronger efficacious beliefs.

Although clearly outside the initial hypotheses presented here, such a dynamic hypothesized modeling change warrants, at least, a brief discussion of how these such findings may influence existing theory, pragmatic use, the field at-large, and the future of writing self-efficacy research. In essence, these modeling approaches do not explicitly alter existing theory, as prescribed by Bandura's (1997) notion and explanation of muti-dimensional conceptual overlap. Furthermore, these approaches do not question the vantage or *'specificity'* of the measure, but may provide a unique perspective that allows reality to better me understood. In other words, it has long been argued that instruments must be *specific* enough to accurately and precisely capture efficacious beliefs (Bandura, 1986, 1997, 2006, 2018; Bong & Skaalvik, 2003; Klassen & Usher, 2010; Marsh et al., 2018; Pajares & Usher, 2008; Pajares, 1996; 2006; Usher, 2015) and that not doing so diminishes its validity in relation to the target domain. That said, the existence of a global factor would provide researchers with a vantage that expresses the commonality amongst the dimensional factors. For the SEWS, a global factor would represent the collective disposition of efficacious beliefs among writing's *mechanics, ideation,* and *self-regulation* inherent to the writing process. Noticeably, and it must be stressed, this does not represent a global sense of writing self-efficacy that is collective to the entire domain of writing, and never can. A global perspective herein simply purports only that which has been measured within each specific factor. Therefore, a measure such as the SEWS retains its given *specificity*, yet may now be better understood to represent, or measure, a certain collective sense of efficacy. Although not expressively influential to the theoretical understanding of writing self-efficacy, a global factor may provide practitioners a better grasp of which students likely exhibit targetable beliefs for intervention. Students to whom exhibit markedly low global efficacious beliefs, over and above their, perhaps, specific factor strengths, may be more easily identified for teacher, parent, or coach led discussions, help, and guidance. In respect to future research, it goes without saying, that if there is a better way to model an instrument, more accurate latent relationships can be derived that can better inform theory. In other words, because global factors assume variability and leave residual variability to be assumed by specific factors, such specific factors can then be more explicitly determined to influence other motivational constructs. For example, the growing body of research from De Smedt and colleagues (2016, 2017, 2018) has demonstrated how each individual factor of the SEWS differentially predicts and is related to cognitive and motivational strategies.

As described earlier, disaggregating participant responses with latent profile analysis after capturing a global facet will permit *shape* effects to be clearly evident. In doing so, the use of latent profile analysis will provide a robust mechanism to determining how common response tendencies, and the beliefs therein, relate to other motivational areas. Said another way, because LPA groups participants in commonly existing profiles or groups (based on response trends), examining their relation to other motivational constructs provides a unique over and above traditional regression by assessing of extrinsic variable relations outside the presupposed linear assumption. Taken together, the presence of a global facet and the use of LPA may offer a completely redefined perspective that can inform theory, enable easier and a more nuanced perspective to practitioners, and ultimately help clarify to what degree efficacious beliefs associated with the writing process relate to other motivational constructs.

Chapter 3: Methodology

This chapter details the variable- and person-centered methodology this study employed to examine whether the adapted Self-Efficacy for Writing Scale depicted sources of construct-relevant psychometric multidimensionality. This study was guided by the following research questions that first assess the presence of two sources of construct-relevant multidimensionality, with RQ1 and RQ2, and then further examine dimensionality and profile validity in a person-centered approach with RQ3 and RQ4.

1. Are the items of the SEWS conceptually related across a priori factors?
2. Does the SEWS exhibit hierarchically-ordered constructs?
3. What specific quantitative profiles of writing self-efficacy emerge?
4. What forms of validity evidence is found for the profiles of the SEWS?
    a. Do the profiles exhibit concurrent validity evidence based on responses to the WSES?
    b. Do the profiles exhibit divergent/discriminant validity evidence based on responses to the Writing Apprehension Scale (WAS-12)?
    c. Do the profiles exhibit predictive validity?

Included are descriptions of the research design, power analysis, population, sample participants, measures, data collection procedures, and analytic plan.

Research Design

As described above in the research questions, this study was focused on investigating and better understanding the overall factor structure of the SEWS to further assess construct-relevant psychometric multidimensionality. That is, although the scale was constructed to query three main constructs inherent to writing self-efficacy (self-regulation, ideation, and conventions), this study was concerned with understanding, modeling, and deciphering their relation, structure, and dimensionality. This study used a non-experimental quantitative research design to examine the adapted SEWS' factor structure using both variable- and person-centered analyses. To date, little work has been published to examine the factor structure of the adapted SEWS beyond traditional independent cluster model confirmatory factor analysis (CFA). Therefore, the purpose of this study was to further examine the factor structure amongst other factor models and examine how they disaggregate through person-centered analyses to provide ample validity evidence.

Power Analyses

This study consisted of two analytic phases, a variable-centered approach that consists of comparing a number of factor models and a person-centered approach that builds on the variable-centered approach to disaggregate the factors further. To determine the minimum number of participants for this study, I conducted a statistical power analysis to guide the sample design and ensure adequate statistical power. To garner an adequate number of participants to sufficiently detect a result, if that result actually exists, is referred to as power (Cohen, 1988). Therefore, it is the probability of rejecting the null hypothesis when the null hypothesis is *false*, which is often described as the inverse of the probability of Type II error ($1 - \beta$) (Cohen, 1988).

Based upon prior factor structure findings, three power analyses across the major model measurement structures were conducted in Mplus version 8.2 using Monte Carlo simulation with 5000 replications while iteratively decreasing the sample size to approach approximately 80 percent significant parameter recoveries across the replications (at the .05 level in a two-tailed test with a critical value of 1.96). (Muthén & Muthén, 2002, 2012). First, a prototypical three factor, three item per factor (adapted SEWS a priori arrangement), confirmatory factor analysis (CFA) measurement model was simulated using factor averaged standardized item factor loadings, latent factor covariances, and residual item variances. These parameter estimates were generated from a similar target sample of 544 6-11th grade students from a comparable size and demographically distributed school division. To adequately recover parameter estimates, approximately 50 participants are needed. Next, from the same preliminary data example a bifactor confirmatory factor analysis (b-CFA) was simulated and suggested approximately 200 participants to adequately recover parameters of similar strength and relation. Finally, a higher-order (second-order) confirmatory factor analysis (h-CFA) was also simulated with the same parameter estimates and suggested approximately 300 participants to adequately recover parameter estimates. Specific details can be found in the Mplus syntax accessible through Appendix G for each simulation.

To my knowledge, no published study or recommendation has clearly outlined Monte Carlo simulation power analyses within the exploratory structural equation modeling (ESEM) framework. Extant literature on sample size recommendation in traditional exploratory factor analysis (EFA) continues to be mixed, often differentiating by either a minimum participant recommendation (e.g. 200 or 250; Cattell, 1978; Guilford, 1954) or a ratio of participants to items (20:1; MacCallum, Widaman, Zhang, & Hong, 1999). Nevertheless, more recent simulation studies have contended and demonstrated minimum participant sample sizes are dependent on many data characteristics, such as how many factors, the number of items per factor, magnitudes of the loadings, and the strength of item to factor cross-loadings (de Winter, Dodou, & Wieringa, 2009; Gagne & Hancock, 2006; McNeish, 2017). However, when conditions are favorable (e.g. strong factor loadings, less factors, strong communalities), true factor structures can be recovered with as few as 20 participants. Despite this conundrum, this study used ESEM models that employ the partial confirmatory approach of target rotation (Asparouhov & Muthén 2009; Browne, 2001). Target rotation allows for the prespecification of target items to load on a priori factors, while also targeting cross-loadings to be minimal (approximately zero). This approach poses less of a risk to being underpowered than traditional EFA, as the magnitude of both factor loadings and item communalities should bolster adequate parameter recovery and require less power to capture. Because targeted item factor loadings can likely be expected to be of less magnitude than those found in a traditional CFA, a conservative approach to ensure adequate power was used. Therefore, for all ESEM models, a minimum sample size of at least 600 participants is conservative enough to provide adequate power to capture expected reductions in targeted item factor loadings.

Participants

All participants were 8th through 10th graders in a large southeastern school division. In the present 2018-2019 school year, this division is made up of 48.5% female, 32.0% identified as economically disadvantaged (which includes those eligible for Free/Reduced Meal, or receives Temporary Assistance for Needy Families (TANF), is eligible for Medicaid, or Identified as either migrant or experiencing Homelessness), 9.8% English Learners, and 12.5% disabled (those who receive services under the Individuals with Disabilities Education Act (IDEA) according to an Individualized Education Program (IEP), Individual Family Service Plan (IFSP), or service plan). The division is also racially diverse, including less than 1% American Indian or Alaskan Native, 3.3% Asian, 25.6% Black or African American, 49.3% White, 16.4% Hispanic, less than 1% Native Hawaiian or Other Pacific Islander, and those who identified as non-

Hispanic, but two or more races 4.9%. Demographics across grades 6 through 10 are comparable to the overall averages.

Recruitment

Data was collected as part of the participating school division's initiative to capture student writing motivation to better focus teacher efforts and prepare for standardized statewide writing assessments.

IRB and Consent

I obtained both division approval for secondary research and VCU IRB approval prior to commencing this research study.

Measures

Demographic Variables

To both accurately describe the sample and provide validity evidence of profiles, I requested a number of demographic and prior performance measures and information. These will include participants' sex, ethnicity, first quarter grades, and standardized writing scores for all participants.

Writing Self-Efficacy

The adapted Self-Efficacy for Writing Scale (SEWS; Ekholm et al., 2015; Zumbrunn et al., 2016), originally developed by Bruning and colleagues (2013), was the primary measure for this study. The modified version of this scale consists of 9 items that ask students to rate, on a scale form 1 (*Almost never*) to 4 (*Almost always*), how confident they are that they can perform specific writing processes. Preliminary work on this scale from two 'under review' studies consistently report McDonald's Omega for each factor; conventions, ideation, and self-regulation at .65, .79, and .80, and .61, .77, and .75, respectively (DeBusk-Lane et al., 2018; Zumbrunn et al., 2018) (Deng & Chan, 2017; McNeish, 2017). The full scale is provided in the Appendix C.

Validity Building Predictors and Outcomes

To support a substantive interpretation and develop validity evidence of the profiles, the employed person-centered approach a number of predictors and outcomes. This effort was to provide both a meaningful interpretation of the profiles based upon prior works that have related self-efficacy to other constructs, as well as continue to build measurement validity to demonstrate the profiles have (a) heuristic value; (b) theoretical alignment or value; (c) anticipated and meaningful relationships to covariates; and (d) generalize, over subsequent replications, to new samples (Bauer, 2007; Marsh et al., 2009; Meyer & Morin, 2016; Morin et al., 2011; Muthén & Asparouhov, 2008). In addition to assessing the demographic variables, I also examined two other measures to provide additional criterion-related validity evidence: The Writing Self-Efficacy Scale (WSES; Pajares, 2007) and a shortened version of the Writing Apprehension Scale (WAS; Bline, Lowe, Meixner, Nouri, & Pearce, 2001; Daly & Miller, 1975; Pajares & Johnson, 1994), the 12-item Writing Apprehension Scale (WAS-12; Limpo, 2018). First, the WSES was chosen, based on both its broad usage in prior literature and the extent to which it has been statistically evaluated, to provide concurrent validity evidence to the SEWS (Pajares, 2007). The WAS-12 was chosen, also based on its extensive use and statistical reliability, to provide concurrent divergent/discriminant validity evidence. In a later section, these two measures will be described in full and methods inherent to their relational value will be presented.

Participant demographics, the WSES and WAS-12, and first quarter grades were used as predictors of profile membership. Furthermore, to better understand how these profiles differentiate across profile, a standardized writing assessment across the grades was examined as the primary distal outcome. Secondary distal outcomes, for validity building purposes, were the WSES and WAS-12.

Standardized writing assessment scores (8th and 10th grade). Both the 8^{th} and 10^{th} grade participants participated in a statewide standardized writing assessment. In both occasions (8^{th} and 10^{th}), the first component required students to correct errors embedded in sections of a nominal rough draft of student writing. The second component required students to write a short paper in response to an expository or persuasive prompt, which are graded holistically on both composing/written expression and usage/mechanics. For scoring, all papers were scored by two trained readers on a scale of 1 to 4 based on the provided rubric (see Appendix D). The composing/written expression domain is counted two times and the usage/mechanics scores is counted once towards the overall score. In the end, three scores are reported, a total, which encompasses both a multi-choice component and both the 'research, plan, compose, and revise,' and 'editing' reporting categories.

Project based assessment (PBA) for writing. The 9th grade writing PBA consists of a standardized persuasive writing prompt independently completed within a timed writing environment. Writing samples were scored with a state-developed high-school writing rubric (see Appendix E) by readers who have been trained on the application of the rubric. As outlined in the rubric, writing samples will be scored in the domains of "composing, written expression, and usage/mechanics." Teacher serving as scorers shall not score their own students' writing samples.

Writing Self-Efficacy Scale (WSES; Pajares, 2007). The WSES scale consists of 10-item scale asking students how sure they are at performing a specific skill on a scale of 0 (*no* *chance*) to 100 (*completely certain*) (Pajares, 2007). Pajares (2007) reported a two-factor solution representing basic grammar skills and advanced composition skills, individual factor Cronbach alpha coefficients of .88 and .86 respectively, and similar factor and reliability findings at the elementary, middle school, and high-school levels, among 1,258 students from grades 4-11.

Writing Apprehension Scale-12 (WAS-12; Limpo, 2018). The WAS-12 is a 12 item shortened version of the 63 item Writing Apprehension Scale originally presented by Daly and Miller (1975) that was, through item reduction, reduced to 26 items that represent a single factor. Similarly, through item reduction techniques, 12 items that represented two salient factors, concern and affect, were presented with Cronbach's alphas for each facet greater than .85 (Limpo, 2018).

Importantly herein, the WAS-12 was previously presented with concurrent validity to Pajares and Valiante's WSES (1999), where the 'affect' (*I like writing*) facet was positively correlated (although not significantly) and the 'concern' facet was inversely significantly related. These findings are in-line with previous work that has examined writing anxiety and writing self-efficacy (Pajares & Johnson, 1994; Pajares & Valiante, 1999; Goodman & Cirka, 2009; Limpo, 2018; Martinez et al., 2011; Sanders-Reio et al., 2014).

Conflict of Interest Consideration

All data was requested from a large southeastern school division. To be clear, I am employed by this division as an educational researcher in the Office of Research and Evaluation. As part of my employment, I have been involved with a multidisciplinary team of educators charged with capturing student writing motivation prior to the implementation of state mandated standardized writing assessments. A portion, but not all, of the data collected as part of this endeavor was used in this study. That said, to clarify both my roll as a student using data for research and as an employed educational researcher, I followed all procedures naturally associated with conducting research on secondary data for both VCU's institutional review board and that required by the school division. There are no financial conflicts of interest associated with this study.

Procedure

All survey data was collected in January 2018 as part of a division priority to assess student writing motivation. Data was collected online with a survey, requiring each student to answer each item before moving on. Each item was presented iteratively with the overall directions for each applicable section as a header. There was no time limit to complete the measures. Teachers were instructed to not provide help in clarifying or explaining the directions or items. All measures were collected in one sitting in each student's English class.

Data Analytic Plan

The data analytic plan encompassed two phases, a variable-centered approach that consisted of multiple factor model comparisons, and a person-centered approach that consisted of a Latent Profile Analysis (LPA) and subsequent analyses.

Variable Centered Analyses

All analyses, unless otherwise noted, were estimated in Mplus version 8.2 using the robust weighted least square estimator using diagonal weight matrices for the factor models (WLSMV; Muthén & Muthén, 2018). The WLSMV estimator is more appropriately suited to the nature of ordered-categorical Likert response categories and has been shown to outperform maximum likelihood estimation/maximum likelihood estimation with robust standard error (ML/MLR) when there are fewer than five response categories, both of which the adapted SEWS uses (Beauducel & Herzberg, 2006; Barendse, Oort, & Timmerman, 2015; Finney & DiStefano, 2013; Sass, Schmitt, & Marsh, 2014). Furthermore, the use of MLR with categorical outcome variables requires numerical integration and is computationally taxing with 3 points of integration (1 point of integration per latent factor) and does not provide model fit indices. Although WLSMV uses listwise deletion for cases with missing data, the anticipated sample size and method of data collection should limit concerns.

To explicate RQ1 and RQ2, which focus on examining he SEWS' hierarchical and item cross-association, a number of model comparisons were needed. Therefore, in total, participant responses on the SEWS were represented with seven models: exploratory factor analysis (EFA), confirmatory factor analysis (CFA), hierarchical CFA (h-CFA), bifactor CFA (b-CFA), exploratory structural equation modeling (ESEM), hierarchical-ESEM (h-ESEM), and a bifactor-ESEM model (b-ESEM). I reported, for all models, item descriptive statistics (distribution, polychorical correlation coefficients (Finney & DiStefano, 2006), model-based omega coefficients of composite reliability (Deng & Chan, 2017; McNeish, 2017), standardized factor loadings, and model fit indexes. When necessary, I also computed omega hierarchical (omegaH), which compared to alpha or standard omega that estimate the proportion of variance exhibited to all sources of common variance, omegaH estimates the proportion of variance in total scores that can be attributed to the global factor (Rodriguez, Reise, & Haviland, 2016). Furthermore, I also computed omega hierarchical subscale (OmegaHS), which assesses the unique variance associated with each group factor while attenuating and accounting (partitioning out) for the variance associated with the global factor. By comparing original omega values of each factor and omegaHS, I was then able to compute the exact amount of variability accounted for by the global factor for each specific factor (Reise, Bonifay, & Haviland, 2013; Rodriguez et al., 2016)

First, to examine a base model that assess item cross loadings, I used an EFA with a Geomin (oblique) rotated solution. This model allows all items to cross-load and allows each latent factor to be correlated. Based upon an assumption of the common factor model, each latent factor exerts a linear causal effect on the measured variables (often referred to as an effects indicator model; Edwards & Bagozzi, 2000; MacCallum & Browne, 1993). That is, I assumed the observed measured variables are effects of the latent variable and likely evidenced by inter-factor item correlations and conceptual similarities among like factor indicator items (Fabrigar & Wegener, 2011). The use of the Geomin oblique rotation allows latent factors to correlate, which can plausibly be theoretically anticipated and warranted. Therefore, the use of EFA in this study was to establish an initial perspective of the interpretable factors needed to describe or explain the correlations among the variables. Furthermore, through an inspection of the eigenvalues, an initial examination of a possible global factor will be discerned, as relatively large first eigenvalues may suggest both a global factor and multidimensionality (Reise et al., 2010).

I then estimated a traditional independent cluster model CFA, whereby all items are forced to load only on their conceptually respective latent factor, without cross-loadings. In this case, each of the three factors will have their respective three observed indicator items loaded. In total, this model will include three correlated factors representing writing self-efficacy self-regulation, self-efficacy of ideation, and self-efficacy of conventions. In the h-CFA, all three SEWS factors will be specified such that they related and were a product of a higher-order latent variable. More specifically, to fully examine the extent to which these three individual factors collectively represent a higher-order factor, one of the first-order factor loadings was set to one. This allowed for the estimation of the higher-order factor variance, which represents the commonality among the first-order factors (covariance explained by the higher-order factor).

Finally, in this series of confirmatory models, I then estimated a bifactor CFA (b-CFA). This entails allowing each observed variable to simultaneously load on a 'general' (G) factor and on their conceptually respective latent 'specific' (S) factor. To allow the S-factors to reflect the variance unexplained by the G-factor, the G-factor and each of the S-factors will be specified as orthogonal (Chen et al. 2006; Reise 2012).

Although these models may offer a better glimpse of the reality of writing self-efficacy, they are, by the very nature of being within the traditional independent cluster model CFA framework, not often pure indicators of the constructs they are constrained to be associated with (Marsh, Lüdtke, et al., 2010; Marsh et al., 2009; McCrae et al., 1996). To better capture expected conceptual and statistical cross-loadings, a series of exploratory structural equation models (ESEM) was also estimated based on the oblique target rotation (Asparouhov & Muthén , 2009; Browne, 2001). Target rotation allows for the prespecification of target items to load on a priori factors, while targeting cross-loadings to be minimal (approximately zero).

First, I estimated a base ESEM model using oblique target rotation, which allows all "targeted" cross-loadings to be as close to zero as possible, while allowing the main loadings to be freely estimated. Next, I estimated a hierarchical-ESEM model, although allowing all three latent factors to be related to a single higher order factor, with no residual correlations between the first-order factors. Finally, I estimated a bifactor-ESEM model with appropriate bifactor assumptive 'orthogonal' target rotation (Reise, 2012; Reise, Moore, & Maydeu-Olivares, 2011). That is, each item will be defined by the G-factor, while also being similarly arranged in the base model ESEM. All confirmatory and ESEM models are depicted in Figure 2.

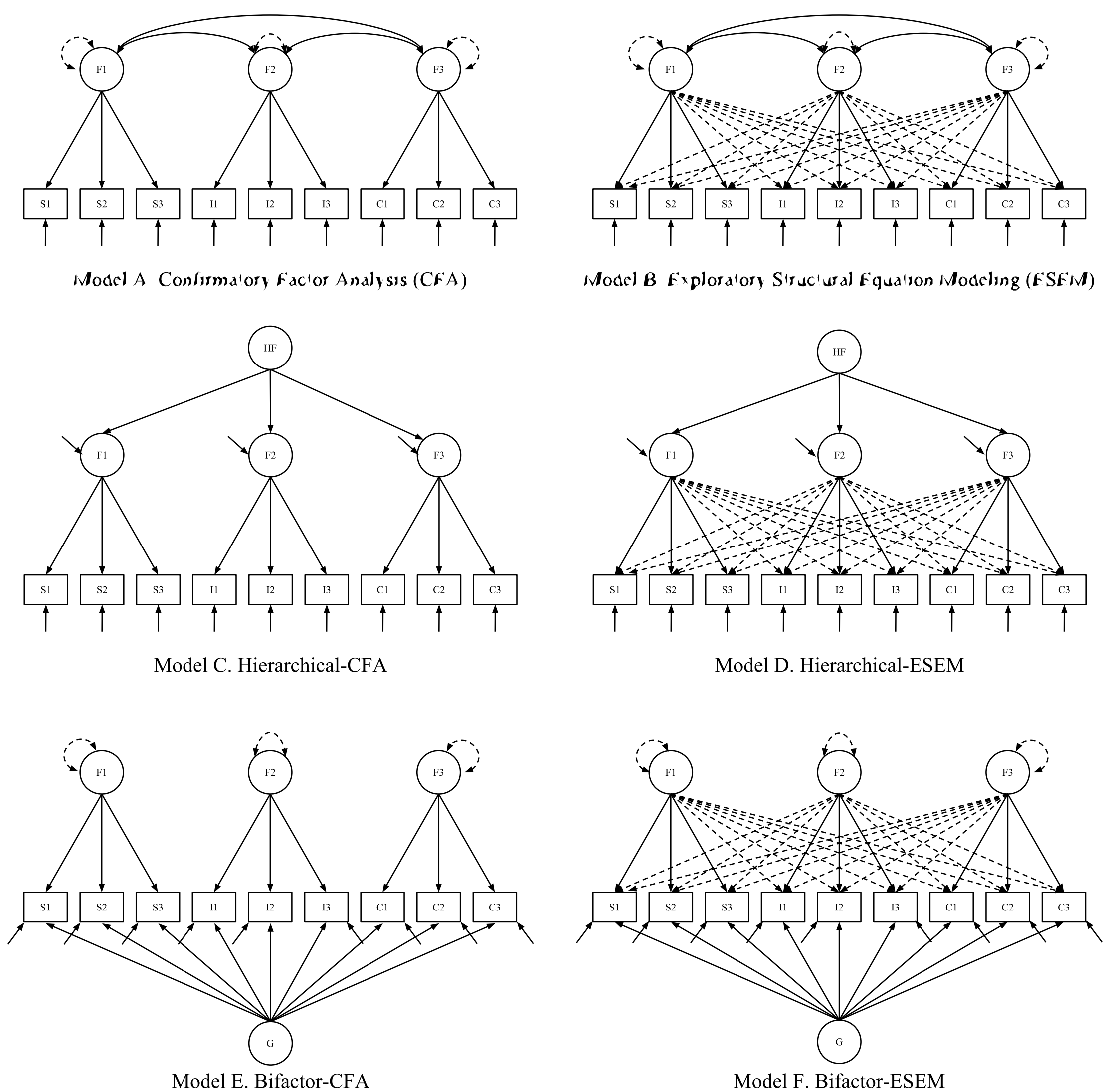


Figure 2. Variable-centered models discussed and to be estimated in this study.

Model evaluations. Model evaluations in this study partially relied on goodness-of-fit indices to describe and compare the fit of all alternative models, as the use of the chi-square test of exact fit and the chi-square differences test is biased due to sample size and model misspecifications--although they will be reported (Kline, 2006; Marsh, Hau, & Grayson, 2005). Therefore, I used the following: the comparative fit index (CFI; Bentler, 1990); the Tucker-Lewis index (TLI; Tucker & Lewis, 1973); the root-mean-square error of approximation (RMSEA; Steiger, 1990; and its 90% confidence interval); and the standardized root-mean-squared residual (SRMR; Asparouhov & Muthén , 2018). Following typical interpretation guidelines (e.g. Kline, 2016; Marsh et al., 2005), CFI and TLI greater than .9 and .95 was considered indicative of excellent fit to the data, respectively. For RMSEA and SRMR, values less than 0.05 and 0.08 are contended to be of excellent fit to the data, respectively (Asparouhov & Muthén , 2018; Hu & Bentler, 1999).

These guidelines do not come without limitations. First, model fit comparisons have been well-established solely within the individual cluster model CFA framework. The adequacy and ability to detect deviations between models, based upon fit change, has not been fully explored among ESEM models. Furthermore, these indices can be influenced by design and model complexity, which limits their ability to detect meaningful differences and generalize beyond the simulation studies they were draft on (Fan & Sivo, 2005, 2007). Considering these limitations, each model comparison included inspections of parameter estimates, statistical conformity, and theoretical adequacy (Fan & Sivo, 2009). This approach has also been echoed in prior ESEM studies (e.g. Grimm et al., 2013; Marsh et al., 2009; Morin et al., 2013; Morin et al., 2016; Morin et al., 2017).

Using the aforementioned comparison guidelines, and as suggested in Morin and colleagues (2016), the CFA and ESEM model was first compared. Assuming the ESEM target factor loadings remained strong and well-established (similar to CFA), the precision for which the factor correlations are modeled will likely be superior (as the cross-loadings will more accurately depict the data) (Asparouhov et al., 2015). Comparatively, an observation of unexpected and theoretically difficult to explain cross-loadings in the ESEM model could suggest needed changes at the item level. Next, depending which initial model fit the data best (CFA vs. ESEM), its corresponding hierarchical and bifactor model was compared. Although the h-CFA/ESEM model is asymptotically equivalent to a first-order factor model (as factor correlations are replaced with higher-order factor loadings), which results in equal degrees of freedom and model fit, it was still assessed within this thesis. In the b-CFA/ESEM model comparison, the presence of reduced factor loadings to the S-factors suggests a bifactor model representation is favorable.

Person-Centered Analyses

Extending the vantage of variable-centered analyses, person-centered statistical approaches can provide a unique lens, while negating some of the given, yet limiting, assumptions of variable-centered approaches. Although variable-centered approaches rely on the assumption that all participants are collected from a uniform population from which *averages* are derived, person-centered approaches assume the sample may include a number of sub-populations (Masyn, 2017). To be specific, variable-centered approaches (factor models) "decompose" covariances to describe relationships between and among variables, while person-centered approaches (latent profiles) uses them to explain and describe the relationships between individuals (Bauer & Curran, 2004). Latent class/profile models can be used to observe latent heterogeneity within a sample, such that individuals are classified into a probabilistic model-based typology that is uniquely defined by response trends (Lubke & Muthén, 2005; Masyn, 2017). Therefore, latent profile analyses can provide a unique vantage from a sample that is presupposed to garner some variability on both a global factor (*level* effects) and S-factors (*shape* effects).

As explained by Morin and colleagues' (2017), both variable- and person-centered approaches can simultaneously be equivalent when a model with *k* profiles is compared to a common factor model with *k*-1 latent factors ("identical covariance implications"; Steinley &

McDonald, 2007). “Variable- and person-centered analyses are thus considered as complementary approaches, as both provide alternative views of the same reality.” (Morin et al., 2017, p. 400). Nevertheless, a person-centered approach has great practical utility in identifying individuals commonly associated with particular response trends across the input variables, which may prove useful to practitioners, especially in educational settings that can greatly benefit from early identification of students for interventions and supplementary instruction.

Despite this utility, the statistical value of person-centered approaches is often questioned. That is, models that only contain *level* effects (e.g. high, medium, or low on all indicator items) are often contended to overcomplicate what common variable-centered mean analyses simply portray (Morin & Marsh, 2015). To optimize the utility, meaningfulness, and interpretation of person-centered approaches, it is advantageous to focus on or accentuate profile *shape* effects (discernible patterns or difference on the indicator items within profile) (Morin & Marsh, 2015). This distinction is equivalent to *level* effects representing differences in a ‘global’ construct sense, while *shape* effects represent ‘S-factor’ differences. Therefore, to examine construct-relevant multidimensionality, the model comparisons employed to capture hypothesized hierarchical or global and cross-construct sources of construct-related multidimensionality are applicable in this context. Capturing a hypothesized ‘global’ construct enables a clear vantage (controlling for a G-factor) to best explicate qualitatively distinct profiles on indicator item differences (e.g. S-factor differences).

Using this logic, I used factor scores derived from the variable-centered measurement model that best depicted the data from the variable-centered approach as indicator items in the person-centered approach. For example, if a b-ESEM model is chosen, I would then use factor scores derived from each S-factor (3) and one from the global factor to represent both the global construct (*level* effects) and each S-factor (e.g. writing self-efficacy of self-regulation, ideation, conventions; *shape* effects) (Morin et al., 2017). This process will model qualitative differences between profiles over and above any globally held attribute of writing self-efficacy, while also providing clarity of G-factor differences between profiles.

Using this approach, I extracted profiles, based on factor scores saved from the variable-centered approach, using Mplus 8.2's (Muthén and Muthén, 2017) MLR estimator, 10,000 random starts, 1,000 iterations for the randoms starts, and 500 final stage optimizations (Hipp & Bauer, 2006). Factor scores were then derived from Mplus, which uses the maximum a priori method (e.g. regression method) to derive scores (Muthén and Muthén, 2017). To generate iterative profiles of increasing profiles, I used MplusAutomation, which is an R package used to systematically execute a number of Mplus input files, to arrange and run all enumeration files (Hallquist & Wiley, 2018; R Core Team, 2017).

During enumeration, I estimated LPAs with 2 to 7 profiles using the aforementioned factor scores (Nylund-Gibson & Masyn, 2016) derived from the traditional CFA model and whichever model fits the data best among the model comparisons. Following the split-sample cross-validation procedures outline in Masyn (2013), I randomly split (stratified) the sample (both the CFA factor scores and whichever model fits best) approximately equally into 'calibration' and 'validation' sets, representative to sex, English Language Learners, and grade level (all R syntax used to perform this can be found in Appendix G). All other covariates were not representative to this split due to sample size considerations (e.g. some were too small to adequately split and resulted in abnormal displacements). Once split, the following enumeration process was performed on the calibration data.

To enumerate these data, I selected models based on multiple statistical indices, theoretical interpretability, and substantive meaningfulness (Marsh et al., 2009; Nylund, Asparouhov, & Muthén, 2007). Statistical indices included minimum values of Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and sample-size adjusted BIC (aBIC). Smaller values of AIC, BIC, and aBIC estimates indicate more parsimony when comparing models (Collins & Lanza, 2013; Geiser, 2013). The entropy value and classification probabilities was also examined, with values closer to 1 indicating higher precision and reliability of classification (Jung & Wickrama, 2008). Although entropy alone was not used as a determinant metric, it offers valuable information about how the profiles relate and are distributed (Lubke & Muthén, 2007). I also employed the bootstrapped likelihood ratio test (BLRT), and the Vuong-Lo-Mendell-Rubin likelihood ratio test (VLMR-LRT) to compare nested models (Muthén & Asparouhov, 2012). These model comparison tests compare the model with *k* latent classes to the model with *k*-1 latent classes, whereby a non-significant p-value indicates the *k*-1 class should be favored (Muthén & Asparouhov, 2012). It should be noted that these indices and tests are heavily influenced by sample size (Marsh et al. 2009). In such cases, these indices will continually suggest an increasing number of profiles, as AIC and BIC will continue to decline as profiles increase, suggesting each is a better fitting model. To mitigate this, I used elbow-plots to graphically depict information criteria, where the point after the slope flattens is recommended as the optimal number of profiles (Morin et al., 2011; Petras & Masyn, 2010). Although it is possible to control for the non-independence of classroom observations or clustering in schools (using Mplus' Type = Complex), I did not employ this during enumeration, as it restricts the computation of BLRT. Nevertheless, not controlling for nesting has been shown to not influence the statistical decisions regarded the number of profiles (Morin et al., 2016).

Once a final number of profiles has been determined, I accounted for clustering, as not accounting for it can bias standard errors and classification accuracy in subsequent analyses (Chen, Kwok, Luo, & Willson, 2010).

After selecting the most appropriate profile solution from the calibration data enumeration, I followed the split-sample double cross-validation procedures (Masyn, 2013) by retaining and saving all parameter estimates from the final *k*-class. Using these saved calibration dataset parameter estimates, I then fit an LPA with the same *k*-classes with these parameter estimates fixed using the ‘validation’ dataset. If the model fits well, then I further compared this fixed ‘validation’ dataset LPA to a freely estimated ‘validation’ LPA using an MLR corrected nested-model chi-square loglikelihood ratio test following equations outlined in Satorra and Bentler (2010). Provided the comparative model fit does not significantly decrease, the model was supported and usable for the entire sample (Collins, Graham, Long & Hansen, 1994; Masyn, 2013). If the comparison test identifies a significantly different fit, a double (or twofold) cross-validation was conducted. This entails using the validation data to freely establish parameter estimates and then fix them in the calibration data to examine (MLR corrected LRT) the reverse fit congruence (Masyn, 2013). If both of these comparative tests indicate a decrease in model fit, a more substantive approach was conducted. In doing so, profiles that are substantively similar between the calibration and validation data models were further assessed for *similarity* (Morin et al., 2016). Following Morin and colleagues’ (2016) multi-group LPA similarity procedures, both models were statistically compared in a series of analyses whereby equality constraints were imposed and increasing become more restrictive between the models—similar to common measurement model invariance testing (Morin et al., 2016). Four models were assessed: configural, structural, dispersion, and distributional. The configural model establishes a base model fit for subsequent analyses, the structural constrains profile means equal, distributional equates within profile variability across all profiles, and distributional, which renders proportional profile sizes invariant between profiles. Iterative model information criteria indicates equality between models (Morin et al., 2016).

Predictor analyses. Each predictor (sex, ethnicity, grade, and prior year standardized assessment (when applicable)) was assessed on its predictive influence on profile membership both individually and together as a whole. Additionally, both the WSES and the WAS-12 were also included as predictors to add validity evidence to the profiles. I used Mplus' R3STEP procedure, which examines covariate influence upon the latent class variable by first estimating an unconditioned (without covariates) model using only the manifest observed indicator variables. Next, the nominal most likely class variable is generated using the posterior distribution estimated from the latent class model. Lastly, this modal class assignment is used as an indicator variable fixed with uncertainty rates which were derived from the logits of the classification error. Therefore, this decouples the covariates from the classification model, accounts for the probabilistic deviations of nominal modal class assignment, and assesses covariate influence upon the latent class variable (Bolck, Croon, & Hagenaars, 2004; Vermunt, 2010). This will result in a series of multinomial logistic regressions to examine how each predictor alone, and accounting for the others, influenced the likelihood of membership in the profiles. Specifically *k* - 1 regression coefficients are generated in relation to a reference profile in the form of log odds (Muthén & Muthén , 2017). To aid in interpretation, I transformed each log odds into odds ratios to present the likelihood of profile membership in the target profile comparative to the reference.

Outcome analyses. Each outcome (WSES, WAS-12, and standardized writing assessments) was assessed across the profiles. Using a similar statistical approach as R3STEP, Mplus' BCH method evaluates the means of outcome variables across profiles (Bakk & Vermunt, 2014; Vermunt, 2010). To be specific, when applicable (for those with prior year achievement) I may employ the manual version of this process outlined in Asparouhov and Muthén (2018) to account for prior achievement and earlier identified predictors by profile to assess the profiles' predictive utility over and above prior year performance. In other words, by using the classification weights in an unconditioned latent profile analysis, I could then manually account for prior performance and other predictors when assessing the profile's predictive utility towards writing performance outcomes.

Chapter 4: Results

This chapter presents the findings from both analytic phases described in Chapter 3. For ease of interpretation and understanding the progression of this study, the results are presented in the order of the research questions. To facilitate this, the chapter will begin with a reiteration of the research questions, descriptive statistics of all involved variables, and an initial exploratory factor analysis. Importantly, this chapter will explicitly only include results and illuminating information required to best understand the decisions needed to navigate the methods executed to answer the below research questions. Therefore, all model substantive, practical, and theoretical interpretations and discussions will be reserved for chapter 5.

Research Questions

1. Are the items of the SEWS conceptually related across a priori factors?
2. Does the SEWS exhibit hierarchically-ordered constructs?
3. What specific quantitative profiles of writing self-efficacy emerge?
4. What forms of validity evidence is found for the profiles of the SEWS?
    a. Do the profiles exhibit concurrent validity evidence based on responses to the WSES?
    b. Do the profiles exhibit divergent/discriminant validity evidence based on responses to the Writing Apprehension Scale (WAS-12)?
    c. Do the profiles exhibit predictive validity?

Descriptive Statistics

This section will be broken up into a number of sections to best describe both the sample and their responses among both the predictors and outcomes herein.

Demographic Descriptive Statistics

Table 1 displays all disaggregated demographic data for sex, ethnicity (minority), and grade level for the total sample of 1,466 students in grades 8 through 10. To mitigate the risk of identifying students, disaggregated data for those who receive special education services (n = 189, 12.9%), participate in a gifted program (n = 210, 14.0%), are English language learners (n = 56, 3.8%) or from smaller ethnicity groups is omitted. To be clear, all of these students were used in the analyses results presented here, just their descriptive statistics were not explicitly displayed, as tabulating some demographic categories could aid in reverse identifying them. Minority was arranged such that non-minority represented both White and Asian students, while minority was assigned to those traditionally under-represented and identified by federal guidelines from Title VII of the Civil Rights Act of 1964. Compared to the school division's overall student distribution in the 2018 school year, across all demographic variables presented, the present participant sample is within approximately six percentile points, as this sample includes 3.8% English language learners, whereas the division serves 9.8% (e.g. all other variables are < 6% difference to the division total).

Table 1. Descriptive Statistics for Demographic Variables

			Sex				Minority			
	N %		Male		Female		Non-Minority		Minority	
N %	1466		727	0.50	739	0.50	810	0.55	656	0.45
Grade										
8	203	0.14	117	0.08	86	0.06	152	0.10	51	0.03
9	488	0.33	213	0.15	275	0.19	252	0.17	236	0.16
10	775	0.53	397	0.27	378	0.26	406	0.28	369	0.25

Polychoric correlations and non-categorical item variability can be found in Appendix F for all non-performance variables. Scale frequencies and descriptive statistics for the SEWS can be found in Table 2. Overall, item response distributions were commonly negatively skewed, yet still within normally accepted ranges of -1 to 1 (Kline, 2016). The 'conventions factor,' however, was obviously negatively skewed (se1 = -2.277) and exhibited fairly strong kurtosis. Omega values for the SEWS' original 3-factor structure were adequate (ω = .58 to .76) and similar to past studies that have reported omega composite reliability (DeBusk-Lane et al., 2018; Zumbrunn et al., 2019). Other measure (e.g. WSES, WAS-12) descriptive statistics will be acknowledged in subsequent sections.

Exploratory Factor Analysis

To fully assess the multidimensionality of the SEWS, I first conducted an exploratory factor analysis (EFA) for models with 1 to 4 factors using a Geomin oblique rotation for categorical variables using the WLSMV estimator (Browne, 2001; Yates, 1987) and accounting for the natural class clustering by using the Mplus Complex option, as demonstrated in Appendix G. Fit and descriptive statistics for models with one through four factors were estimated and are reported in Table 3. The fit of models with one and two factors was suboptimal compared to that of three factors, while the 4-factor model did not converge. The sample correlation matrix extracted eigenvalues of 3.854, 1.279, 0.852, and 0.704, which represent the sum of the squared factor loadings for each subsequent factor. The chi-square statistic significance *p*-value increased only for the model with three factors (p = 0.0081), suggesting less of a difference between the actual covariance matrix and the proposed three factor model in explaining the matrix. Additionally, the ratio of the chi-square value and the degrees of freedom are only below three for the three-factor model (Kline, 2016). Nevertheless, due to the sample size, the chi-square value should be interpreted with caution (Brown, 2015; Kline, 2016).

Examining the goodness-of-fit statistics, the three-factor model should be retained, as its fit indices are excellent (Brown, 2015; Kline, 2016). As depicted in the output (Appendix G), the rotated loadings clearly demonstrate the a priori item to factor relationships derived and put forth by the adapted SEWS (Zumbrunn et al., 2016). However, notable significant cross loadings are evident. For example, a significant loading of .323 exists between item 9 ("I can keep writing even when it is difficult.") and the factor that is conceptually related to writing *ideation* (see Appendix G for EFA outputs for more details). Furthermore, this exploratory depiction also notes, despite capturing all cross-loading effects, significant latent factor correlation (r = .296-.598), which may also indicate a lack of further construct relevant multidimensionality inherent to the global nature of the construct.

Table 2. *Adapted Self-Efficacy for Writing Scale Response Frequencies and Descriptive Statistics*

	N	*Almost never* (1)		2		3		*Almost always* (4)		*M*	σ^2	skewness	kurtosis
Self-Efficacy for Ideation		*n*	*p*	*n*	*p*	*n*	*p*	*n*	*p*				
ω = 0.79 CI [0.763 , 0.805]													
2. I can think of many words to describe my ideas.	1466	27	.018	199	.136	691	.471	549	.374	3.216	0.241	-0.628	-0.039
6. I can think of many ideas for my writing.	1466	79	.054	313	.214	630	.430	444	.303	2.994	0.721	-0.482	-0.465
7. I can put my ideas into writing.	1466	46	.031	252	.172	619	.422	549	.374	3.149	0.650	-0.629	-0.276
Self-Efficacy for Mechanics													
ω = 0.62 CI [0.582 , 0.658]													
1. I can write complete sentences.	1466	4	.003	41	.028	245	.167	1176	.802	3.776	0.241	-2.277	5.253
3. I can punctuate my sentences correctly.	1466	21	.014	158	.108	580	.396	707	.482	3.359	0.513	-0.857	0.164
5. I can spell my words correctly.	1466	44	.030	190	.130	609	.415	623	.425	3.239	0.623	-0.809	0.085
Self-Efficacy for Self-Regulation													
ω = 0.78 CI [0.762 , 0.802]													
4. I can concentrate on my writing for a long time.	1466	116	.079	446	.304	603	.411	301	.205	2.742	0.761	-0.196	-0.682
8. I can avoid distractions when I write.	1466	235	.160	484	.330	545	.372	202	.138	2.485	0.832	-0.045	-0.811
9. I can keep writing even when it is difficult.	1466	186	.127	523	.357	548	.374	209	.143	2.541	0.774	-0.031	-0.710

Note. Omega coefficients of composite reliability were computed using 1000 bootstrapped samples along with bias corrected confidence intervals (see Zhang & Yaun, 2016). By scale response, both the sub-sample quantity (*n*) and the proportion ($\hat{p}$) are provided.

Table 3. *EFA with a Geomin oblique factor rotation fit statistics for the SEWS.*

Model	Chi-square	Chi-square df	Chi-square /df	CFI	TLI	RMSEA	RMSEA CI-low	RMSEA CI-hi	RMSEA *p*-value	SRMR
1-factor	645.480	27	23.907	0.923	0.897	0.125	0.117	0.133	0.000	0.103
2-factor	288.779	19	15.199	0.966	0.936	0.098	0.089	0.109	0.000	0.052
3-factor	26.874	12	2.240	0.998	0.994	0.029	0.014	0.044	0.992	0.017
4-factor										

Note. Estimator = WLSMV, EFA Factor rotation = Geomin (oblique). The 4-factor solution did not converge.

Variable-Centered Approach

For comparison, Table 4 presents the goodness-of-fit of the various models. In general, all confirmatory and ESEM models provide adequate fit to the data (CFI: 0.981-1.000, TLI: 0.971-1.000), however, as the models progress, they generally continue to improve. An exception, the hierarchical CFA's fit declined compared to the base 3-factor CFA. Additionally, RMSEA values for all confirmatory models (CFA, bCFA, and hCFA) were above 0.06, which exceed common recommendations (Hu & Bentler, 1998, 1999; Kline, 2015). Judging from these fit statistics alone, the bifactor ESEM model should be retained (Morin et al., 2016, 2017). However, as described in Chapter 3, a full detailed inspection of all parameter estimates, their relationship to each latent arrangement, and their theoretical conformity is necessary to determine the best model fit (Morin et al., 2016).

Table 4. *Goodness-of-Fit of all models.*

Model	Chi-Square	df	CFI	TLI	RMSEA (90% CI)		RMSEA *p*	SRMR
EFA 1	550.182	27	0.853	0.804	0.115	[0.107, 0.123]	0.000	0.068
EFA 2	337.031	19	0.911	0.831	0.107	[0.097, 0.117]	0.000	0.035
EFA 3	27.708	12	0.996	0.987	0.030	[0.015, 0.045]	0.989	0.012
CFA	180.045	24	0.981	0.971	0.067	[0.058, 0.076]	0.001	0.037
hCFA	225.819	24	0.978	0.967	0.076	[0.067, 0.085]	0.000	0.037
bCFA	163.020	18	0.984	0.968	0.074	[0.064, 0.085]	0.000	0.031
ESEM	26.874	12	0.998	0.994	0.029	[0.014, 0.044]	0.992	0.012
hESEM	26.874	12	0.998	0.994	0.029	[0.014, 0.044]	0.992	0.012
bESEM	0.176	2	1.000	1.003	0.000	[0.000, 0.019]	0.997	0.001

Note. RMSEA *p*: Probability that RMSEA is <= .05.

Research Question 1

To determine the extent to which the items of the SEWS exhibit construct relevant psychometric multidimensionality due to the presence of conceptually related constructs, I compared the CFA to the ESEM model. Overall, both models fit the data well, however, the ESEM model's goodness-of-fit statistics were marginally better. For example, the CFA exhibited an RMSEA of 0.067, while the ESEM model 0.029, suggesting the ESEM model has less error of approximation and has excellent fit (MacCallum, Browne, & Sugawara, 1996). Latent factor correlations are stronger for the CFA ($|r|$ = .510 to .808, M = .652) than the ESEM ($|r|$ = .428 to .704, M = .547), suggesting the ESEM model provides a more distinct vantage of the specific factors compared to the CFA. Standardized parameter estimates (factor loadings and residual variances) for both the CFA and the ESEM are presented in Table 5.

An examination of the parameter estimations across both the CFA and ESEM models suggests both models exhibit strong factor to item relations (CFA: $|\lambda|$ = .538 to .857, M = .756; ESEM (a priori items only): $|\lambda|$ = .549 to .970, M = .711), however, this is to be expected. In general, the a priori factor loadings across the ESEM model are weaker, suggesting a more accurate depiction of true score variation in comparison to the CFA, as the ill-modeled true score variability is more accurately extended to target cross-loadings (e.g. non-a priori item cross-loadings). Interestingly, target factor loadings across the factors (target only: $|\lambda|$ = -.195 to .221, M = .042) are commonly statistically significant, yet lack strength. This may demonstrate that a majority of the items exhibit a common theme and could better be exhibited by a general factor. Together, these findings suggest the ESEM model is more accurately depicting true score variation and accounting for construct relevant multidimensionality from conceptually related constructs inherent between the latent factors of the SEWS.

Table 5. *Standardized Factor Loadings and Residual Variance for the CFA and ESEM.*

Items	ICM-CFA λ (*SE*)		ICM-CFA δ	ESEM λ (*SE*) Ideation		ESEM λ (*SE*) Mechanics		ESEM λ (*SE*) Self-Regulation		ESEM δ
1. Ideation										
Item 2	0.728	(0.014)**	0.470	0.549	(0.041)**	0.311	(0.034)**	-0.001	(0.034)	0.429
Item 6	0.797	(0.015)**	0.364	0.877	(0.042)**	-0.142	(0.022)**	0.060	(0.031)	0.267
Item 7	0.857	(0.011)**	0.265	0.739	(0.038)**	0.043	(0.032)	0.111	(0.030)**	0.288
2. Mechanics										
Item 1	0.838	(0.034)**	0.298	0.190	(0.033)**	0.711	(0.043)**	-0.050	(0.038)	0.363
Item 3	0.717	(0.024)**	0.486	-0.023	(0.039)	0.732	(0.044)**	0.041	(0.031)	0.456
Item 5	0.538	(0.031)**	0.710	-0.106	(0.037)**	0.568	(0.035)**	0.107	(0.040)**	0.680
3. Self-Regulation										
Item 4	0.805	(0.016)**	0.351	0.157	(0.033)**	-0.003	(0.021)	0.673	(0.034)**	0.376
Item 8	0.724	(0.020)**	0.476	-0.195	(0.024)**	0.007	(0.019)	0.970	(0.035)**	0.282
Item 9	0.800	(0.015)**	0.360	0.221	(0.033)**	0.022	(0.020)	0.576	(0.031)**	0.423

Note. All a-priori item factor relationships are in grey.

Research Question 2

To examine if the SEWS exhibits construct relevant psychometric multidimensionality due to the presence of a hierarchically ordered construct, I compared the ESEM model (previously found to be superior to the CFA) to both the hierarchical ESEM and bifactor ESEM models. Because the ESEM model was chosen from RQ1, this section will omit comparison to both the hierarchical CFA and bifactor CFA (outputs of both models can be found in the online supplement found in Appendix G).

Overall, the fit of all three ESEM models is excellent. Of note, however, the hESEM model fit is asymptotic to that of the ESEM model, as the first-order factor correlations (now disturbances) from the ESEM model are modeled as factor loadings. Because of this, degraded fit, and the fact that second-order models are less interpretable and theoretically useful herein, I also omit a full comparison between the hESEM and ESEM model in this chapter (output for the hESEM model is available through Appendix G). The omission of the hESEM model will be discussed in Chapter 5.

Unfortunately, the bESEM model did not converge in its original configuration. In assessing the failed model, it was found that item-1, which is heavily negatively skewed, as 80.2% of all responses, or n = 1176 were for *'Almost always'* (I can write complete sentences). Using theta parameterization, which is an alternative estimation technique that models the latent variable distribution variability, y^*, differently (yet produces identical model parameter estimates as delta parameterization), allowed model convergence and demonstrated that item-1 abnormally aligned with the global factor. Additionally, this caused latent factor score computational problems for over 90% of all responses in which participants responded as a 1, or "*Almost never,"* and was reported as a minimization error in computing the factor scores.

Taking a substantive approach to this item, it is commonsensical to expect a vast majority of secondary students to respond more positively, which does not likely or well attenuate to their developmental level. In other words, it can be expected that a vast majority of students are capable and view themselves as capable of 'writing a complete sentences' and simply corresponded accordingly, obviously negatively skewing the respond distribution. This item also stands apart from the other two within factor items that did not reflect a similar response trend. Interestingly, on inspecting the initial confirmatory and base ESEM models, this item did not strongly or abnormally present itself, as the WLSMV is well known to control and handle non-normal item distributions (Finney & DiStefano, 2006). Therefore, identifying that this item's response distribution as problematic only in a bifactor exploratory structural equation scenario is both statistically and pragmatically relevant and useful to future research in this area. To my knowledge, no readily available published works have presented a similar issue with this type of model.

Upon removing this item, the bESEM model adequately converged and a full parameter inspection was conducted to ensure the specific *mechanics* factor displayed normal functioning and adequately represented a meaningful latent factor from the two remaining freely estimated items that well differentiates from the other specific factors and target items (Brown, 2015; Kline, 2016). In doing so, the specific *mechanics* factor displayed expected a priori and target parameter estimates, clearly delineating a unique and meaningful factor. That is, for this factor alone, a priori factor loadings ranged from .375 to .724, while target (as close to zero as possible) loadings ranged from -.084 to .033 and global factor loadings ranged from .326 to .474 (see Table 6). Therefore, despite dropping item 1, the bESEM adequately models the data and will be used in comparison to the ESEM model.

Compared to the ESEM model, the bESEM model goodness-of-fit indices are superior (see Table 4). The bESEM's G-factor exhibits strong significant factor loadings for all items ($|\lambda|$ = .326 to .820; M = .625). In most cases, the strength of the factor loading on the G-factor exceeds that of the S-factors. Although factor loading significance is derived from the ratio between the loading strength and its standard error and simply provides a statistical test to determine if the loading is significantly different than zero, it does suggest which loadings likely provide practical significance. For example, although item 6's target loading is statistically significant on the *mechanics* factor, the strength of the loading itself suggests it may not be practically significant. Alternatively, item 4's loading of .439 and standard error of .036 suggest it to be a meaningful item factor loading. Therefore, one must interpret both the strength and significance of the loadings when determining their practicality. Nevertheless, a majority of the S-factor loadings ($|\lambda|$ = .087 to .724; M = .409) are markedly stronger than the target loadings ($|\lambda|$ = -.009 to .154; M = -.002).

Although the strength of the S-factor loadings are commonly less than that assumed by the G-factor, it can be expected that the factor correlations reported for the ESEM model ($|r|$ = .428 to .704, M = .547) are somewhat consumed and re-expressed by increased factor loadings on the G-factor due to having an orthogonal latent factor arrangement. In particular, items 2 and 7 exhibited weak loadings on their a priori factor (λ = .087 and .179, respectively), yet strong loadings on the G-factor (λ = .723 and .820, respectively), suggesting these items relate stronger to global efficacious beliefs towards writing than specific efficacious beliefs towards writing *ideation*. Said another way, earlier models likely exhibited these items' variability as relating to between latent factor correlations, yet once the global factor was introduced and the specific factors were disallowed to correlate, the variability is consumed and represented by the global factor. Ultimately, the *ideation* factor appears to contribute less specific relation within the model (1.91% of the reliable variance) than either the *mechanics* or *self-regulation* factors, which exhibit some items that provide stronger parameter estimates towards the S-factor than the G-factor. Additionally, as depicted by OmegaH, the global factor assumed approximately 87% of the reliable variance, suggesting there is a robust theme that runs congruent amongst all the variables therein. Therefore, this model provides a superior depiction of and fit to the data, as suggested by both the goodness-of-fit indices and the extent to which the parameter estimates are generally supportive of a general factor, while also exhibiting specific factor variability over and above that depicted by target loadings. Furthermore, the strength of the G-factor substantiates the need to more accurately model construct-relevant psychometric multidimensionality in relation to globally structured concepts and is clearly needed in this case as the items collectively load on it.

Of note, I also examined all models without item 1 to assess model fit and interpretability in an attempt to permit a full and complete model comparison (again the bESEM). Unfortunately, model fit and interpretability among all previous models before establishing the alternative (without item 1) bESEM model did not either converge, exhibit acceptable fit, or provide a readily interpretable solution. Despite this, it was determined that accepting the alternative bESEM model without item 1, considering its fit and interpretability, was important to report and a vital contribution to writing self-efficacy research.

Table 6. *Standardized Factor Loadings for Bifactor Exploratory Structural Equation Modeling Solution of the Self-Efficacy for Writing Scale (-se1)*

Items	λ (*SE*)				δ
	Ideation	Mechanics	Self-Regulation	G-Factor	
1. Ideation					
Item 2	0.087 (0.121)	0.154 (0.056)	-0.063 (0.039)	0.723 (0.038) **	0.442
Item 6	0.511 (0.259) *	-0.099 (0.031) **	0.047 (0.021) *	0.750 (0.032) **	0.164
Item 7	0.179 (0.155)	-0.036 (0.033)	0.025 (0.038)	0.820 (0.038) **	0.294
2. Mechanics					
Item 3	-0.084 (0.077)	0.375 (0.110) **	-0.045 (0.062)	0.474 (0.046) **	0.625
Item 5	-0.017 (0.069)	0.724 (0.192) **	0.033 (0.026)	0.326 (0.033) **	0.367
3. Self-Regulation					
Item 4	0.081 (0.047)	-0.013 (0.024)	0.439 (0.036) **	0.654 (0.022) **	0.373
Item 8	-0.025 (0.045)	0.036 (0.021)	0.623 (0.043) **	0.563 (0.032) **	0.294
Item 9	0.009 (0.042)	-0.028 (0.029)	0.336 (0.038) **	0.690 (0.029) **	0.411
ω	0.866	0.654	0.838		
ω_H	0.017	0.039	0.061	0.788	
ω_{HS}	0.082	0.432	0.292		
% Var. Ind. G-Factor	9.46%	65.94 %	34.86%		
% Reliable Var.	1.91%	4.31%	6.77%	87.01%	

Note. ** $p < 0.01$, * $p < 0.05$. All target factors are in greyscale. % Var. Ind. G-Factor = Percent variation independent of the G-Factor; % Reliable Var. = Percent of reliable variance ($\omega_H \div (1 - \text{total error})$). ω: Coefficient omega; ω_H: Coefficient omega hierarchical; ω_{HS}: Coefficient omega hierarchical subscale.

Person-Centered Approach

To section will present findings pertaining to a person-centered approach. Herein, I will use factor scores derived from the CFA, ESEM, and bESEM to fully examine how the presence of construct-relevant multidimensionality disaggregates and provides further validity evidence for the SEWS. This is not to negate the clear fact that a bifactor ESEM model best depicts the data, but that it is informative to collectively present latent profiles derived from a traditional model (CFA), a superior model that well captures conceptual overlap (ESEM), and an ultimate model that captures both conceptual overlap and a global concept inherent to all items (bESEM).

Research Question 3

To establish the extent to which the data disaggregates into discernable, meaningful, and interpretable profiles, I first enumerated a calibration data set of the CFA, ESEM, and bESEM factor scores. All profile unconditional enumeration indices are reported in Appendix H. For comparison purposes, a detailed data and visual product was created and can be found through Appendix G for each configuration.

Examining the CFA calibration enumeration indices, no clear information criteria leveled off, suggesting a logical stopping point in enumeration (Morin et al., 2011; Petras & Masyn, 2010). However, a rather strong 5-profile non-significant aLMR p-value suggested the k-1 profile was superior. Assessing the enumeration profile substantively, the 4-profile solution exhibited logical and practical profiles, while the 5-profile solution (which was also suggested by the k+1 non-significant aLMR p-value) exhibited a small ($\hat{p} < .04$; $n \sim 24$ of the calibration dataset) extreme negative profile. Although I'm not discounting it as less than meaningful, adding extra profiles that simply provide more extreme versions of already existing profiles can be less informative and problematic as profile size decreases and variability increases during post-estimations (Masyn, 2013). Ultimately, because the calibration 4-profile solution was both statistically supported and provided a meaningful and interpretable solution, it was chosen to represent the CFA measurement model.

Using the starting values of the calibration 4-profile solution, I fixed the starting values of the validation dataset and assessed the difference against a freely estimated validation dataset 4-profile solution. Testing at the $\alpha = .05$ significance level using the MLR corrected chi-square LRT, I did not reject the null model (TRd = 21.98, $df = 18$, $p = 0.23$). This method was also inversely replicated by fixing the starting values of the calibration dataset LPA with those derived from a freely estimated validation dataset and then assessing the difference against a freely estimated calibration model (TRd = 26.32, $df = 18$, $p = 0.09$). Therefore, the fit of the 4-profile is statistically found to validate well and be replicated stable across the two subsamples and will be used as the final enumerated unconditional CFA LPA model moving forward. Table 7 reports the means and standard errors, while Figure 3 provides a visual depiction.

Table 7. *Profile Indicator Means and Standard Errors (CFA)*

	Ideation		Mechanics		Self-Regulation		$\hat{p}$
Profile	*M*	SE	*M*	SE	*M*	SE	
1	-1.031	0.048	-0.934	0.060	-1.059	0.053	0.136
2	-0.321	0.038	-0.264	0.037	-0.353	0.042	0.377
3	0.293	0.038	0.202	0.033	0.317	0.039	0.353
4	0.990	0.034	0.824	0.039	1.085	0.047	0.144

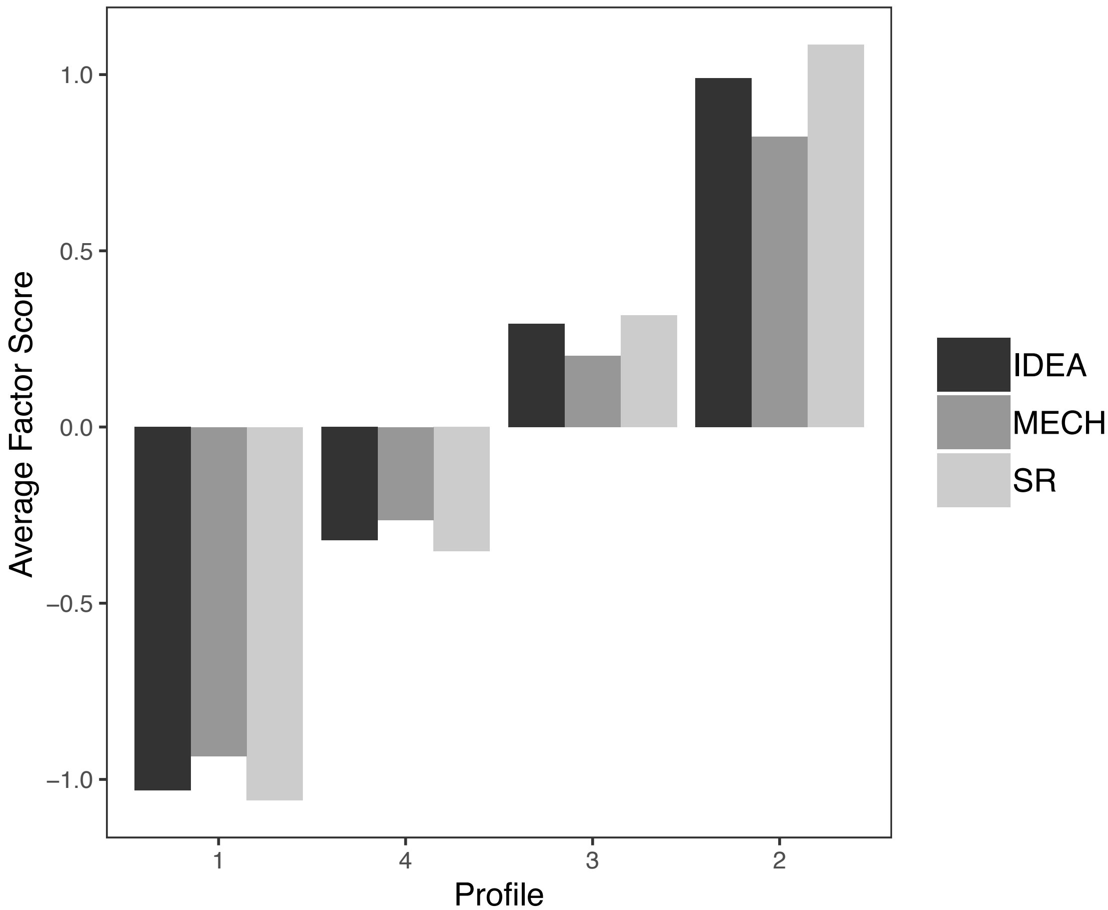


Figure 3. CFA Latent Profile – 4-Profile

In a similar case, the ESEM model followed suit in exhibiting no information criteria level-off and reporting a non-significant k = 5 profile aLMR p-value. Additionally, the 5-profile calibration solution also exhibited a very similar extreme low profile with a small proportion of the sample ($\hat{p}$ < .05; n ~ 32). Using the split-sample cross-validation method, both analyses resulted in p > .01, suggesting the profile validates across the entire sample. Therefore, the fit of the 4-profile model will be used as the final unconditional ESEM LPA model. Table 8 reports the ESEM LPA profile means and proportions, while Figure 4 depicts this visually.

Table 8. *Profile Indicator Means and Standard Errors (ESEM)*

	Ideation		Mechanics		Self-Regulation		$\hat{p}$
Profile	*M*	SE	*M*	SE	*M*	SE	
1	-1.379	0.070	-1.000	0.085	-1.259	0.075	0.135
2	-0.420	0.069	-0.269	0.048	-0.403	0.062	0.368
3	0.414	0.066	0.222	0.047	0.381	0.061	0.354
4	1.303	0.045	0.916	0.050	1.317	0.070	0.133

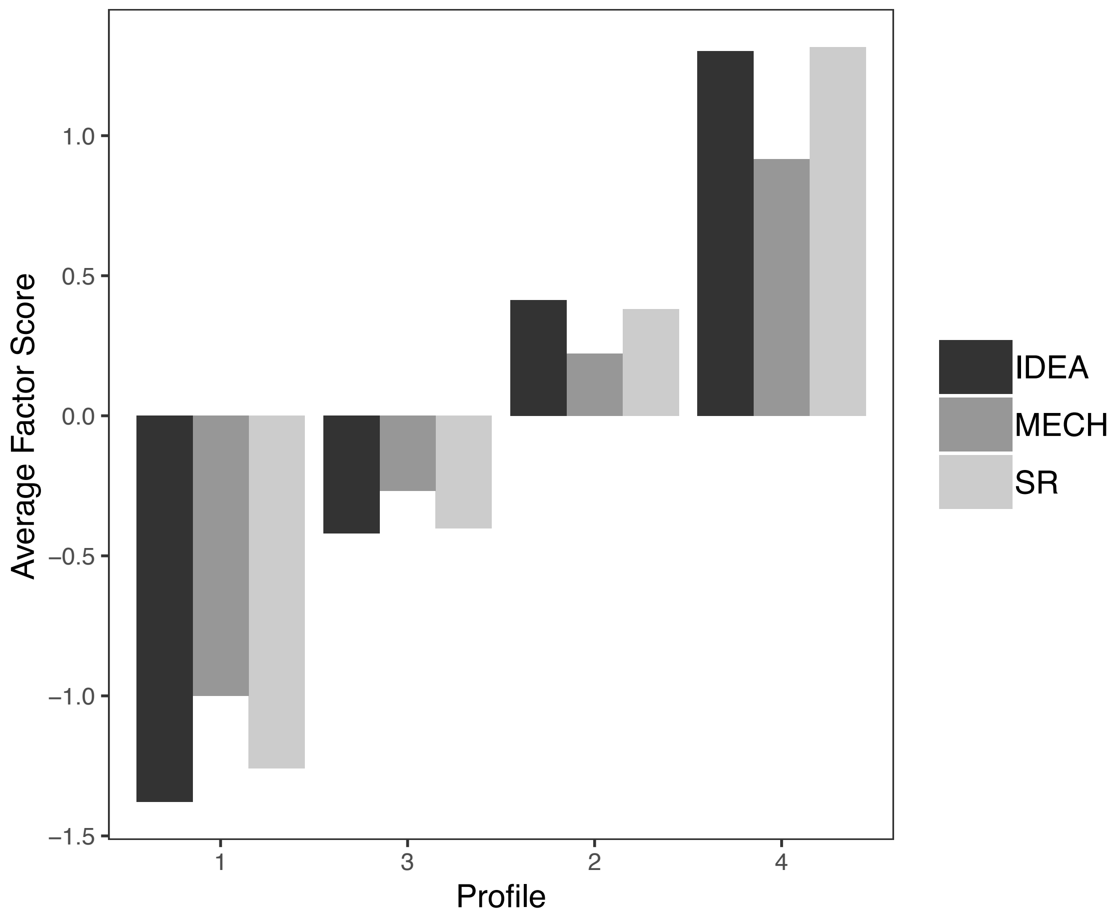


Figure 4. ESEM Latent Profile – 4-Profiles

Examining the bESEM calibration enumeration, a clearly non-significant aLMR *p*-value indicated the 3-profile model was favored. The double split-sample cross-validation method, however, suggested the 3-profile solution was not congruent across the entire sample (p = .0001 & .0026, respective to both cross-validation adjusted chi-square LRTs; see Appendix I). This sample split-sample cross-validation method was then deployed to the 4-profile, 5-profile, and 6-profile calibration and validation data, also with no success in replicating the profile configurations across the entire sample.

Despite this, I substantively inspected both the calibration and validation 3-profile solutions and found they had very similar profile means, variances, and proportions. Therefore, I assessed the profile similarity using Morin and colleagues' (2016) multi-group tests of *similarity*. As evidenced by continued model fit improvements from CAIC, BIC, and aBIC, it was determined that the two samples met *configural, structural, dispersion,* and *distributional* similarity, validating the 3-profile solution across the entire sample. From this, I also looked at the 4-profile solution substantively to ensure a 3-profile provided a better vantage, despite it not having a non-significant aLMR *p*-value for the *k*+1 profile.

In doing so, the 4-profile solution replicates the major profiles exhibited by the 3-profile solution, but also includes a profile that exhibits low *global* and *ideation* (-0.393, -0.653 factor score averages, respectively) averages and a markedly higher (0.653) self-regulation average. This extra profile, which is substantively interesting, seemingly replicates the 3-profile solution's lowest profile, but with stronger and positive self-regulation. Therefore, it appears this new profile may allow self-regulation to be separated between the other profile to a further extent, as the other profiles, which mimic those found in the 3-profile configuration, exhibit slightly stronger and extreme values associated with self-regulation.

Nevertheless, in determining the most appropriate profile configuration, a number of both statistical and substantive criteria are necessary and used (Masyn, 2013; Morin et al., 2016). In this case, there is little statistical evidence to select the 4-profile solution over and above the 3-profile solution, as the 4-profile solution is not supported by aLMR *p*-values and the information criteria continue decline, which is expected given the sample size (Marsh et al., 2009). Conversely, the 3-profile solution is supported by both a non-significant aLMR *p*-value for the *k*+1 profile and a notable and obviously elbow plot decline in information criteria (e.g. AIC, BIC, aBIC) at the 3-profile configuration (Morin & Marsh, 2015; Petras & Masyn, 2010). Substantively, both the 3- and 4-profile solutions are informative and interpretable, given their adequate profile proportions and indicator mean delineations (e.g. 4-profile solution exhibit a unique profile). Consistent with prior enumeration work and previous recommendations that guide enumeration decisions, a more parsimonious profile solution was retained as the final model herein given the statistical support (Marsh et al., 2005, 2009, Muthen, 2009). Of note, however, future research should not negate the 4-profile solution and may be a fruitful avenue to better understand students' differentiations in relation to writing self-efficacy of self-regulation. Table 9 reports each profile's mean, standard error, and proportions, while Figure 5 depicts this visually.

Table 9. *Profile Indicator Means and Standard Errors (bESEM)*

	Global		Ideation		Mechanics		Self-Regulation		$\hat{p}$
Profile	*M*	SE	*M*	SE	*M*	SE	*M*	SE	
1	-0.725	0.077	-0.496	0.049	0.128	0.064	-0.040	0.044	0.267
2	-0.219	0.092	0.566	0.151	-0.414	0.159	-0.516	0.079	0.151
3	0.484	0.112	0.073	0.033	0.021	0.040	0.224	0.098	0.582

Note. $\hat{p}$ = proportion of sample.

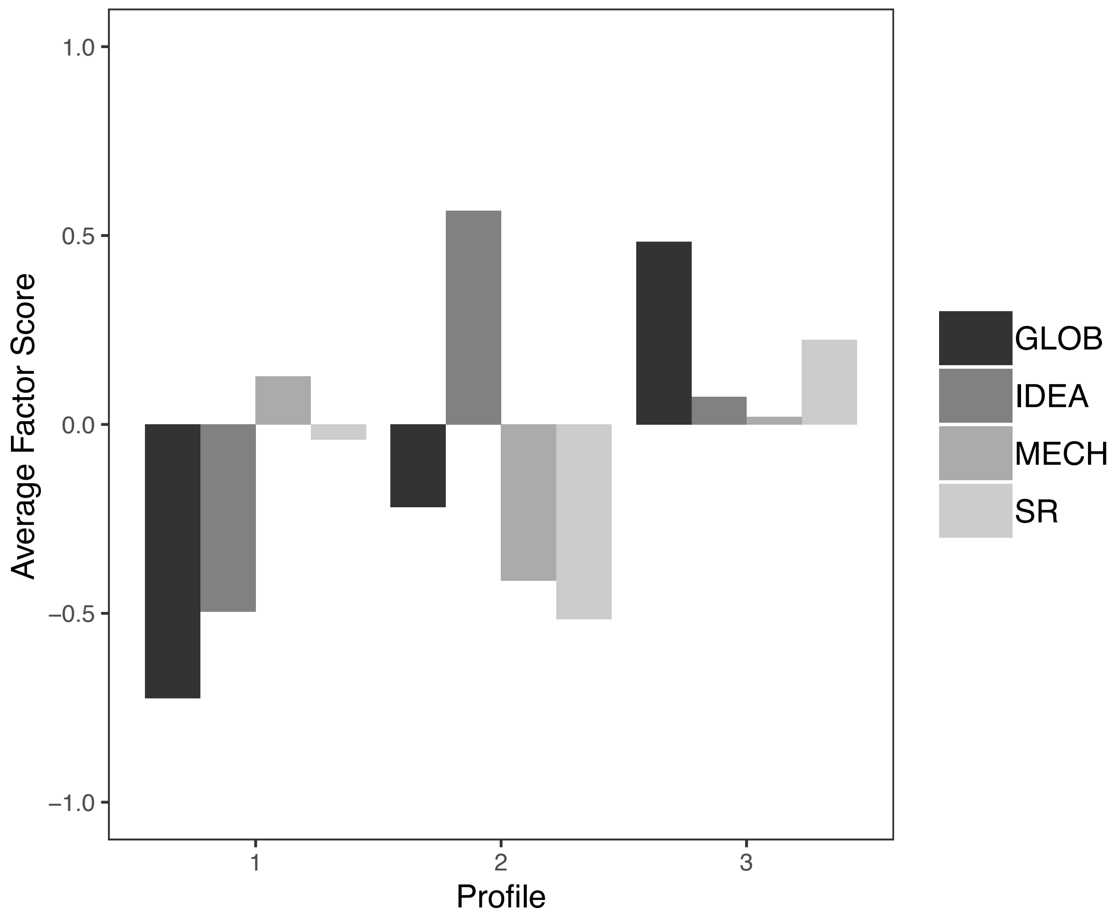


Figure 5. bESEM Latent Profile – 3-Profile.

Final bESEM LPA Model Descriptive Statistics

This section will provide an overview of the profiles exhibited by the 3-profile bESEM model LPA, descriptive statistics, and assess other aspect that will help shape the narrative to best understand what each profile means, what validity evidence it can provide, and especially how it adds to our theoretical understanding of writing self-efficacy.

Demographic descriptive statistics are reported in Table 10. Therein, each profile is well represented by the demographic variables and some trends begin to take shape that may prove predictive. For example, there are clear differences among how the 10th graders are distributed across the profiles. Additionally, it appears ELL participants are more prevalent in the lower profiles. Nevertheless, I will later assess each demographic variable to determine the extent to which it predicts profile membership.

Table 10. *Demographic % by Profile*

	Profile 1	Profile 2	Profile 3
Total *n* (1466)	26.67	15.14	58.19
Sex (female)	48.85	48.20	51.70
Minority	46.29	50.45	42.56
8th	10.23	18.02	14.42
9th	36.32	36.49	31.07
10th	53.45	45.50	54.51
ELL	6.39	4.05	2.58
Disabled	13.55	15.32	11.96
Gifted	13.04	9.91	16.06

Note. Each percentage represent the percent of each variable represented in each profile.

Despite the 3-profile model expressing Morin's *similarity*, the model does exhibit rather low entropy (.583), at least comparative to other confirmatory factor score derived models (~ > .8). Although entropy should not be used in determining the most optimal profile configuration (Lubke & Muthen, 2007), it does provide a metric representative to how well the cases are classified into their respective profiles. That said, low entropy in this case is likely a product of the loss of *level* information between the profiles (see Morin et al., 2017). Nevertheless, the lowest profile by indicator variable average, denoted herein as 'profile-1' (see Figure 3) exhibited a .711 classification probability for the most likely latent class, profile-2 - .652, and profile-3 - .907, suggesting the smaller profiles exhibit more classification error, or have less probability of being assigned, based on their response trends, to the profile in which they are modally assigned (profile in which they have the highest probability of membership). Despite this, each profile is clearly denoted by different profile means and is discernible apart from the others, therefore providing substantive meaning that is uniquely modeled and meaningful. Together with the aforementioned *similarity* analysis, a low entropy value herein is but descriptive and of little concern, especially considering two of the three profile exhibit entropy values greater than .7 and it is logical to assume entropy is reduced due to a lack of *level* differences.

Nevertheless, each profile's mean latent factor score derived from the bESEM model and the profile standard error are reported in Table 9. To be clear, for identification and descriptive purposes, the lower profile, exhibited by the lowest global indicator average, will be referred to as 'profile-1,' and continue iteratively as the global values increase. Although I will fully describe this profile in more detail in Chapter 5, it is useful to clearly identify this profile now to not confuse later as this configuration will remain throughout this dissertation.

Research Question 4

To assess the concurrent and divergent/discriminant validity of the SEWS, a number of predictors and outcomes were assessed for their relation to the final enumerated profiles derived from RQ3. Although the CFA derived LPA enumeration was presented earlier for comparison purposes, results for predictors and outcomes will not be assessed further, as the ESEM model is, undoubtedly superior. Furthermore, all ESEM model predictor and outcome results can be found in Appendix G.

First, a series of demographic predictors were both individually and collectively tested to examine the extent to which they predicted profile membership. All individual demographic predictor regression coefficients, standard error, and odds ratios are reported for the bESEM 3-profile LPA in Table 11 and collectively in Table 12.

Table 11. *Individual Predictor Coefficients and Odds Ratios for Demographic Variables*

Predictor	Profile 1 vs. 3				Profile 2 vs. 3				Profile 1 vs. 2			
Sex	Coef.		SE	OR	Coef.		SE	OR	Coef.		SE	OR
	-0.162		0.160	0.850	-0.225		0.229	0.799	0.063		0.241	1.065
Grade	Coef.		SE	OR	Coef.		SE	OR	Coef.		SE	OR
	0.047		0.119	1.048	-0.351	*	0.169	0.704	0.398	**	0.120	1.489
Minority	Coef.		SE	OR	Coef.		SE	OR	Coef.		SE	OR
	0.232		0.147	1.261	0.498	**	0.171	1.645	-0.267		0.204	0.766
Gifted	Coef.		SE	OR	Coef.		SE	OR	Coef.		SE	OR
	-0.357		0.281	0.700	-0.914	*	0.411	0.401	0.556		0.453	1.744
Disability	Coef.		SE	OR	Coef.		SE	OR	Coef.		SE	OR
	0.224		0.261	1.251	0.445		0.447	1.560	-0.221		0.307	0.802
ELL	Coef.		SE	OR	Coef.		SE	OR	Coef.		SE	OR
	1.395	**	0.461	4.035	0.946		0.627	2.575	0.449		0.501	1.567

Note. * $p < .05$; ** $p < .01$.

Table 12. *Predictor Coefficients and Odds Ratios for Demographic Variables*

Predictors	Profile 1 vs. 3				Profile 2 vs. 3				Profile 1 vs. 2			
	Coef.		SE	OR	Coef.		SE	OR	Coef.		SE	OR
Sex	-0.176		0.173	0.839	-0.188		0.245	0.829	0.013		0.258	1.013
Grade	0.008		0.122	1.008	-0.439	**	0.169	0.645	0.446	**	0.124	1.562
Minority	0.155		0.144	1.168	0.533	**	0.185	1.704	-0.378		0.208	0.685
Gifted	-0.290		0.284	0.748	-0.901		0.471	0.406	0.611		0.491	1.842
Disability	0.225		0.291	1.252	0.354		0.538	1.425	-0.129		0.362	0.879
ELL	1.397	**	0.486	4.043	0.906		0.704	2.474	0.491		0.512	1.634

Note. * $p < .05$; ** $p < .01$.

Next, measurement model (CFA in both cases) factor scores from both the WSES (basic skills factor: ω = .89, CI [.879, .902]; advanced skills factor: ω = .92, CI [.911, .929]) and the WAS-12 (affect: ω = .88, CI [.867, .890]; concern: ω = .84, CI [.828, .855]) were assessed for their predictive utility towards the likelihood of profile membership. All regression coefficients, standard errors, and odds ratios are reported in Table 14. Of note, all measurement model goodness-of-fit indices are reported in Table 13 for both the WSES an WAS-12. Although the fit of the WAS-12 was marginal, the use of factor scores derived from the two latent factors is often regarded as more optimal than composite scores (Morin et al., 2016). Despite this lack of general fit, because the factor scores were only being used as a metric to provide validity, the model's goodness-of-fit was not a focus of concern.

Table 13. *Goodness-of-Fit of all Validity Models (WSES & WAS-12)*

Model	Chi-Square	*df*	CFI	TLI	RMSEA (90% CI)		RMSEA *p*	SRMR
WSES	211.124	34.000	0.976	0.969	0.060	[0.052, 0.067]	0.019	0.028
WAS-12	1522.550	53.000	0.784	0.731	0.138	[0.131, 0.144]	0.000	0.106
WAS-12*	829.652	51.000	0.886	0.852	0.102	[0.096, 0.108]	0.000	0.100

Note. * Residual variances were allowed to correlate for items 4 and 5 and items 8 and 9 to improve model fit.

Table 14. *Predictor Coefficients and Odds Ratios for WSES and WAS-12 Latent Factor Scores and First Quarter English Grades*

Predictors	Profile 1 vs. 3				Profile 2 vs. 3				Profile 1 vs. 2			
	Coef.		SE	OR	Coef.		SE	OR	Coef.		SE	OR
WSES - Basic	0.365		0.24	1.441	-0.497	*	0.207	0.608	0.863	**	0.179	2.370
WSES - Advanced	-1.719	**	0.217	0.179	-0.887	**	0.269	0.412	-0.832	**	0.189	0.435
WAS12 - Affect	-2.168	**	0.232	0.114	-1.165	**	0.25	0.312	-1.003	**	0.188	0.367
WAS12 - Concern	1.545	**	0.198	4.688	0.983	**	0.235	2.672	0.562	**	0.183	1.754
Q1 Eng Grades	-0.041	**	0.011	0.960	-0.046	**	0.011	0.955	0.005		0.006	1.005

Note. * $p < .05$; ** $p < .01$.

Next, the WSES (basic and advanced writing skills), WAS-12 (affect and concern), and both 8th and 10th grade standardized tests were assessed as outcomes across the profiles. All, by profile, mean outcomes are reported in Table 15.

Table 15. *bESEM LPA Outcomes by Profile*

	Profile 1	Profile 2	Profile 3	Summary of significant differences
	M	*M*	*M*	
Total *N*	391	222	853	
WSES - Basic	-1.014	-1.332	1.077	1 = 2 < 3
n	391	222	853	
WSES - Advanced	-1.400	-1.303	1.301	1 = 2 < 3
n	391	222	853	
WAS-12 - Affect	-0.661	-0.209	0.474	1 < 2 < 3
n	391	222	853	
WAS-12 - Concern	0.496	0.213	-0.376	1 < 2 < 3
n	391	222	853	
Grd 8 Total Performance	446.189	436.446	476.044	2 < 3
n	38	38	117	
Grd 8 Category 1	34.218	34.279	37.065	1 = 2 = 3
n	38	38	117	
Grd 8 Category 2	34.770	32.560	37.067	1 > 2 < 3
n	38	38	117	
Grd 10 Total Performance	444.216	431.196	477.077	1 = 2 < 3
n	191	93	432	
Grd 10 Category 1	35.063	34.236	38.167	1 = 2 < 3
n	191	93	432	
Grd 10 Category 2	34.594	32.276	38.516	1 > 2 < 3
n	191	93	432	

Note. Category 1 = "Research, plan, compose, and revise for a variety of purposes." Category 2 = "Edit for correct use of language, capitalization, punctuation, and spelling". Significant differences are $p < .05$ from a Wald chi-square difference test. Total performance, category 1, and category 2 are standardized writing scores.

Chapter 5: Discussion, Conclusions, and Implications

The purpose of this study was to examine construct-relevant multidimensionality within the adapted SEWS and provide further validity evidence (Ekholm et al., 2016; Zumbrunn et al., 2016). This chapter includes a discussion of the major findings for each research question, how such conclusions reported here influence both theory and our understanding of how writing self-efficacy should be measured, and how these findings may shape future research. This chapter will conclude with a discussion of limitations, future directions, and a concise summary.

This study was guided by the following research questions that first assess the presence of two sources of construct-relevant multidimensionality, with RQ1 and RQ2, and then further examine dimensionality and profile validity in a person-centered approach with RQ3 and RQ4.

1. Are the items of the SEWS conceptually related across a priori factors?
2. Does the SEWS exhibit hierarchically-ordered constructs?
3. What specific quantitative profiles of writing self-efficacy emerge?
4. What forms of validity evidence is found for the profiles of the SEWS?
    a. Do the profiles exhibit concurrent validity evidence based on responses to the WSES?
    b. Do the profiles exhibit divergent/discriminant validity evidence based on responses to the Writing Apprehension Scale (WAS-12)?
    c. Do the profiles exhibit predictive validity?

In summary, the SEWS exhibited evidence of construct-relevant multidimensionality as a product of both latent construct overlap and the existence of a global factor. Using a bifactor exploratory structural equation model as the final model, three latent profiles of response trends were discovered that exhibit strong relationships that were in-line with hypothesized expectations given previous research.

Interpretations of the Findings

To best facilitate the interpretation of these findings, the following sections will be organized by research question and will provide an overview and discussion that will include how they relate and inform existing theory. Limitations, recommendations for future research, and possible implications for educators will also be provided.

Research Question 1

To determine the extent to which the items of the SEWS exhibit construct relevant psychometric multidimensionality due to the presence of conceptually related constructs, I compared the CFA to the ESEM model. Prior to doing so, it was beneficial to examine both common descriptive statistics and an initial exploratory factor analysis. Collectively, the sample used was socially diverse, equally defined in terms of sex, and was well representative across all three grades. Additionally, the sample approximates the division's state reported demographics within ~6 percent, suggesting generalizability is adequate and appropriate.

During preliminary item analysis of the SEWS (Table 2), it was identified that item 1 ("I can write complete sentences") exhibited strong negative skew, whereby over 80 percent of the participants answered at the maximum of the scale. In prior work, this item was also found to exhibit strong negative skew and kurtosis and have the highest average among both elementary and high school students (M = 3.51 and M = 3.62; Zumbrunn et al., 2019). Similar to the models reported here, all models executed by Zumbrunn and colleagues (2019) handled this item's non-normality accordingly and resulted in adequate CFA models with the MLR estimator. Although the WLSMV estimator used herein does not make distributional assumptions of observed indicators, the item does stand apart from the others. That is, the WLSMV estimator is robust against non-normality of observed variables and should adequately model such occurrences (Finney & DiStefano, 2006). Upon further inspection, the item, “I can write complete sentences,” is likely too easy of an item and not squarely developmentally appropriate for secondary K-12 students. Despite this, all initial models (e.g. EFA, CFA, ESEM, bCFA) captured this well and were not problematic.

The initial EFA supported a 3-factor CFA model with strong goodness-of-fit indices, eigenvalues, and a rising chi-square statistic *p*-value. As noted in Chapter 4, the EFA also suggested significant cross-loadings on non-a priori item factor relationships and fairly robust factor correlations (r = .296-.598). Together, although depicting a 3-factor arrangement, the EFA did provide evidence that an ESEM model would provide a better fit and represent the data more appropriately. Of note, although the EFA suggested a better fit by an ESEM model, it still exhibited factor correlations that may better be represented and modeled by some type of global or hierarchical factor. Therefore, RQ1 was focused on determining if the ESEM model better represented the data, compared to the CFA model. Based on fit, reduction in latent factor correlations, and a parameter analysis, the ESEM model better represented the data.

Theoretically, Bandura (1997) suggested that multidimensional measures constructed to capture efficacious beliefs would likely exhibit conceptual overlap. Until now, at least from my vantage, no other study has examined if this is truly the case. Being the case here, this study will well inform researchers at large that efficacious beliefs, if at least conceptually related across a multidimensional measure, can be better modeled by an ESEM. That is, it is commonly found in recent writing self-efficacy literature (Bruning et al., 2013; DeBusk-Lane, Lester, & Zumbrunn, 2018; Limpo & Alves, 2017; Ramos-Villagrasa et al., 2018; Zumbrunn et al., 2019) that efficacious beliefs exhibit latent factor correlations that suggest conceptual overlap. The present study, however, provides statistical evidence that such correlations are, in some part, actually unmodeled non-a priori item-factor relationships. Although this is common in the social sciences, especially in psychological measures (see Morin et al., 2016), it does indicate that there is shared variability across latent factors and, given new statistical approaches (e.g. ESEM), may better be modeled to represent reality.

Statistically, the decrease in latent factor correlations between the CFA model ($|r|$ = .510 to .808, M = .652) and ESEM ($|r|$ = .428 to .704, M = .547) demonstrate this well and are clearly evident in the factor loadings on target (situated non-a priori loading starting value to zero; $|\lambda|$ = -.195 to .221, M = .042) loadings that represent true cross factor relationships. Although this comparison has likely been previously evident in earlier EFAs, traditional CFA methods that test a priori conceptual models have restricted all true score variation not absorbed by the latent factor to be depicted as latent factor correlations, as opposed to being assumed, in some part, by existing cross-loadings.

Theoretically, the ESEM model reported here provides the current theoretical understanding of writing self-efficacy important updates. For example, items focused to capture efficacious beliefs of *ideation*, in some part, are also influenced by self-beliefs associated with how well one can perform common writing *mechanics*. This is commonsensical, as it should be expected that beliefs associated with "…put[ing] my ideas into writing" (e.g. item 7) likely tap into and relate to beliefs associated with common writing mechancs such as punctuation, spelling, or forming complete sentences. In this case, as item 7 is phrased, to "put" ideas into writing implies the use and performance of the "generally accepted standards for expressing ideas in writing" (Bruning et al., 2013, p. 28). These cross-concept influences exist for all factors. Therefore, such cross-concept relations support the notion that efficacious beliefs exist not in extreme *specificity*, but that they prevail broadly in relation to writing as a whole.

In relation to the adapted SEWS, this suggests that efficacious beliefs associated with the "psychological and linguistic features of the writing process" (Bruning et al., 2013, p. 25), likely exist and can be modeled, in some part, by a global factor, as latent factor correlations still remain ($|r|$ = .428 to .704, M = .547).

Research Question 2

General discussion. To examine if the SEWS exhibits construct relevant psychometric multidimensionality due to the presence of a hierarchically order construct, I compared the ESEM model to both the hierarchical ESEM and bifactor ESEM models. It is worth noting that both the hCFA and bCFA were omitted in the full analysis and results from Chapter 4, however, all model results can be found through Appendix G. Following Morin and colleague's (2016) procedures, the best fitting model from RQ1 was compared to the like (CFA/ESEM) hierarchically or globally situated model (hESEM/bESEM). Although the hESEM model was estimated, it is simply a reconceptualization of the ESEM's latent factor correlations and offers little extra information. To be specific, the hESEM uses the starting values expressed in the ESEM model as starting values for the first order item factor relationships and models the ESEM factor correlations as higher order factor loadings, resulting in a mathematically equivalent model (Hershberger & Marcoulides, 2013). The decision to omit the hierarchical ESEM model was because these models do not offer additional information about how well the data represented a 'hierarchical' or global construct (for more details, see Gignac, 2016; Morin et al., 2016; Morin, Arens, Tran, & Caci, 2016; Reise, 2012).

Although the original 9-item scale did not adequately converge with a bESEM model, it was determined that there was sufficient cause to remove item 1 based on both statistical and developmental reasons. Furthermore, removing the item resulted in a converged model that continued to represent the *mechanics* S-factor well and established similar item to factor relationships as expected across all factors. Taking all of this into account, it was also determined that retaining this model and comparing it to the ESEM model is not only statistically meaningful, but important to report.

In comparing the ESEM model to the adapted bESEM model, the bESEM model exhibited superior overall goodness-of-fit and anticipated G and S-factor relations. That is, although most (all but one) S-factor a priori loadings exhibited stronger loadings for the G-factor, a majority of the factor loadings continued to provide significant strength over and above the G-factor, while continuing to model minimal target item relations across non-a priori item factor relationships. In this case, the continued latent factor correlations found in the ESEM model are re-expressed as the global factor. As described in Chapter 4, the *ideation* factor loadings suggest it contributed less to the S-factor than either of the other factors, which exhibited stronger collective loadings to the S-factor. It is important to recall that the G-factor represents the shared variability across all items, while the S-factors express shared variance among the a priori items controlling for the G-factor (Reise, 2013). To that end, these trends are uniquely clear in examining the omega coefficients and the percent of variation independent of the G-factor. For instance, for the *ideation* factor, only 9.46% of the reliable variance is independent of the global factor, suggesting the *ideation* factor is almost entirely captured by the global factor. However, despite dropping item 1, the *mechanics* factor models 65.94% of the reliable variance after accounting for the global variability, suggesting it is a unique factor

(Reise, 2013). *Self-regulation* exhibited the second highest amount of variance accounted for independent of the G-Factor (34.86%), while also accounting for the highest percent of reliable variability at 6.77%. Therefore, *self-regulation* also appears to be a strong unique factor, as it accounts for a large portion of variability after accounting for the G-factor and models the largest portion of reliable variability after accounting for error. The G-factor, which accounted for 87% of the total reliable variability, suggests that the global factor is both ubiquitous across the items and strong.

Theoretical implications. In terms of theory, the existence and prevalence of such a robust global factor extends the theoretical updates provided by the ESEM model. Capturing the shared or common variance exhibited by all items more readily expresses the conceptual overlap described by Bandura (1997), while also providing clear and present evidence that this multidimensional measure exhibits a strong common theme that runs throughout all variables and, although future research is needed, may extend to other facets commonly associated with writing self-efficacy. In other words, although efficacy beliefs are commonly understood to be domain specific (Bandura, 1986, 1997, 2006, 2018; Bong & Skaalvik, 2003; Klassen & Usher, 2010; Marsh et al., 2018; Pajares & Usher, 2008; Pajares, 1996; 2006; Usher, 2015), these findings suggest there is a strong common theme associated, at least, to the psychological attributes associated with the process of writing. This model suggests, apart from having a common belief system that relates to the writing process as a whole, students vary in some of the particular facets or S-factors. Said another way, although students may exhibit collectively high or low efficacious beliefs associated with writing, they still appear to vary between the specific factors therein. Although this seems logical, as there should be natural S-factor variation at any given point along the (global) continuum of writing beliefs, it may be that such variability is indicative to certain student characteristics, experiences, or methods of writing instruction, as it is well argued that a student's 'sociocultural' situation and collective experience greatly influence their self-efficacy development (Usher & Pajares, 2008; Usher & Weidner, 2018). This logic is squarely in theoretical alignment, as Bandura (1997) contended that while generalized self-efficacy is often stable, more specific efficacious beliefs become strongly influenced by contextual and experiential factors. Nevertheless, this model statistically affords researchers and theorists alike the opportunity to examine a more exact representation of *specific* factor variability over and above a general theme, seemingly providing ample avenues for future research (Morin & Marsh, 2015).

As the field progresses forward, this modeling vantage will offer a unique ability to not only understand how the specific factors relate to other latent motivational constructs (e.g. De Smedt's and Zumbrunn's most recent works), but how these specific facets relate over and above that which is naturally associated with beliefs of writing in general (Chen et al., 2006). In doing so, S-factor scores derived from a bESEM model permit researchers and practitioners alike to fully assess specific factor differences relative to their general self-referent beliefs associated with writing. As will be seen (e.g. RQ3 and RQ4), students often exhibit substantial differences among the S-factors relative to their position on the general factor, suggesting that without accounting for a global theme much less structural differences would be exhibited and a clear depiction of the variability therein would be lost. This would, in light of such a robust general factor, provide a more accurate and precise vantage of latent construct relationships. Not only does the bESEM model provide an easily interpretable depiction of overall writing self-efficacy, it allows an explicit analysis of whether the S-factors uniquely predict and relate to other motivational constructs over and above the general factor. Given recent research focused on using the SEWS in SEM frameworks, determining if the residual variability modeled by the S-factor is predictive can provide meaningful updates to how these specific facets of the writing process are influenced and influence other motivational constructs and other key covariates often related to writing self-efficacy (e.g. sex). It must be noted, however, that not all domain specific factors and the general factor should be assessed simultaneously, as this would introduce a linear dependency among the predictors (Chen et al., 2006).

Pragmatically, the presence of a well-defined global factor allows practitioners to use this information in a more nuanced manner to more specifically target particular facet constructs for intervention. Said another way, practitioners, and also researchers alike may be interested in targeting particular groups of students who, for instance, exhibit strong beliefs associated with ideation over and above general ability beliefs of writing. Relatedly, there is also utility in examining students who exhibit low self-regulation after controlling for a particular general writing self-efficacy level. Nevertheless, as either a researcher or educator, the bifactor model permits a unique and more detailed view of the extent to which efficacious beliefs actually exist.

Beyond the theoretical and practical benefits of using a bifactor model, RQ2 was focused at examining evidence to determine if the adapted SEWS exhibits construct-relevant psychometric multidimensionality due to the existence of a global or hierarchical facet. As reported in Chapter 4, there is clear evidence that the bESEM model best depicts the data, as it identified clear and discernable S-factors, a well-defined G-factor, useful (significant), yet minimal target (non-a priori) cross-loadings, improved goodness-of-fit, and the use of reliable variability and variance independent of the global factor across all three factors. Therefore, as described, the bESEM more accurately depicts reality and re-expresses correlated factors exhibited by the ESEM model, in part, because of the presence of construct relevant multidimensionality as a result of the presence of a global facet inherent to all items of the adapted SEWS.

Research Question 3

General discussion. Once the bESEM model was established as the final model that best depicted the data and best modeled the evident construct-relevant psychometric multidimensionality, I sought to examine how latent factor scores from the final bESEM model disaggregated into interpretable profiles to better grasp the measure's validity. In summary, the bESEM factor scores settled on a 3-profile configuration that provides ample disaggregation and meaningful profiles across the entire sample.

During the enumeration phase, LPAs were enumerated with factor scores derived from the CFA, ESEM, and the bESEM models for comparative purposes. As reported in Figure 3 and Figure 4, both the CFA and ESEM models suggested 4-profiles exhibited by very similar factor score averages for each S-factor. Although the ESEM measurement model provided better fit and accurate depiction of the data, the only discernable difference between the two final LPAs is in terms of average indicator score magnitude. For instance, the indicator variable range for *ideation* in the ESEM LPA ranged from -1.379 to 1.303, while the CFA LPA spanned from -1.031 to 0.990. Although this is likely attributable to the larger factor score variance, min, and max of the ESEM scores, otherwise the two LPA models are similar and do not result is substantively different profiles.

Importantly, both the CFA and ESEM LPA models overtly exhibited what is often referred to as *level* effects (Bauer, 2007; Morin & Marsh, 2015). This occurs when the profiles are uniformly high, medium, and low. For example, in both cases all three indicator items (*ideation, mechanics,* and *self-regulation*) clearly denote each profile vertically apart from each other (i.e. forming high, less-high, less-low, and low profiles), lacking *shape* effects (differences within a profile between indicator variables). Although models such as these would be better be represented by common variable-centered analyses, they are indicative of a strong underlying global construct (Morin et al., 2017). Without controlling for a global construct in these cases, it becomes increasingly difficult to discern qualitative differences between the profile indicators, aside from overall *level* differences. Identifying obvious *level* effects, without clear identification of *shape* effects, is, and has been, an indication that there is shared variability among the indicators that may better be captured by a global factor (see Morin et al., 2016; 2017).

In comparison, it is clearly evident that once capturing the global construct inherent to all the items and using it as a profile indicator, *shape* effects are allowed to be modeled. In this case, through enumeration a 3-profile model was chosen that best represents the data (see Figure 5). For clarity, I will denote profile-1 (*Strongly Inefficacious*) as that which is also depicted as profile-1 in Figure 5, profile-2 (*Moderately Inefficacious*) as profile-2, and profile-3 (*Efficacious*). To be clear, although there is often utility in 'naming' profiles, because these profiles exhibit fairly unique *shape* differences, no one name will likely encompass the full scope of a given profile well or adequately. This not to say the profiles themselves are not unique or stand apart, but that simply describing them with one- or two-word names likely limits the true description and risks under-describing them. This is especially important herein, as S-factor indicators are residual factor scores and should be interpreted accordingly. The decision to name these profiles similar to their G-factor indicator mean differences was both because the G-factor assumed so much variability and because the interpretation of the S-factors within a name would overcomplicate a mere naming convention. That said, the use of "profile-#" or the given name above will used interchangeably herein.

In judging and describing the profiles of the bESEM factor scores it is vital to remember that they should be interpreted similar to that of the bifactor model in which they were derived. That is, each profile's S-factors represent expressed factor score variability over and above that exhibited by the G-factor. Therefore, scores of *specific* factors are above (or below) that of the given global facet within profile. Said another way, values, or means as they may be here, can be thought of as values that control for and parse out collective beliefs associated with writing. Although this differentiates from how 'normal' LPA profiles are interpreted, it provides a unique vantage in terms of understanding groups or clusters of students and how they differentiate between both an overall sense of writing efficacy and beliefs aligned to each specific facet (e.g. ideation, mechanics, and self-regulation). Additionally, using factor scores, in which the mean is set to zero (and variance to, using the regression method, the squared multiple correlation), should be considered when assessing how different the profiles are and their inherent magnitude. To say this another way, the profiles should be interpreted at face value, such that dominant and clustered profiles may exist with opposing factor scores (e.g. high global score, with low *mechanics* scores). Nevertheless, profiles derived from bifactor models permit a unique perspective to not only assess how students group themselves relative to particular levels of collective writing self-efficacy and specific facets and what predicts membership in these groups, but also to examine the ways in which groups relate to well-established outcomes (e.g. standardized tests and both the WSES and WAS-12).

In general, the bESEM LPA produced three profiles well-differentiated by *level* differences of global writing self-efficacy. In this case, and relating to the common interpretation of bifactor models, profiles-1 and -2 exhibited low global writing efficacy, yet well-differentiate through all three of the *specific* factor responses. Profile-1 (, which includes approximately 26%

of the participants (n = 381), is characterized by low *global* writing self-efficacy, low *ideation*, moderate *mechanics*, and relatively average *self-regulation*. Therefore, participants in this profile are collectively doubtful, yet exhibit above average beliefs of their writing *mechanics* and much less confidence in their ability to develop and use ideas. Relative to their doubt, these students feel that they can employ common spelling and punctuation, yet overwhelmingly struggle to think of and use ideas. Comparatively, profile-2 portrayed participants who, despite having more than half the low global efficacy, exhibited strong beliefs associated with developing and using ideas, yet are less confident with managing the writing process and employing common writing conventions. Being the smallest profile, including approximately 15% of the participants (15.15%; n = 222), it is also the most obvious in terms of demonstrating the utility of capturing global writing self-efficacy while simultaneously capturing meaningful subscale specificity. Thus, without modeling the collective variability exhibited by all the items, such disparities and unique profiles are, given the demonstration from both the CFA and ESEM LPAs, not likely to be found. Profile-3, denoted by strong positive global beliefs, fairly average *ideation* and *mechanics*, and moderately strong *self-regulation*, is expressed as the normative profile by including almost 60% of participants (n = 853). Despite expressing strong global beliefs, these participants exhibit confidence in all specific facets, especially in their ability to manage the writing process.

Theoretical importance and relation to extant research. With the original intent of RQ3 to examine what specific quantitative profiles of writing self-efficacy emerge, the theoretical implications here relate directly to both the utility in modeling a global sense of efficacy on the creation of profiles and the particular differences between the profiles of the bESEM LPA.

First, without capturing and modeling a global factor, the profiles reported from both the SEWS' CFA and ESEM models lack substantive differences, aside from the obvious *level* effects. Provided the limited research in the field of writing self-efficacy using person-centered approaches, the current study extends DeBusk-Lane and colleagues' (2018) CFA derived LPA results. In doing so, the disaggregation of the four input variables (global factor, mechanics, ideation, and self-regulation) across the sample provides clear substantive differences between the profiles (e.g. both *level* and *shape* effects). Drawing from both the current study's reported CFA LPA 3-profile estimation (although a 4-profile configuration herein was found to fit best) and DeBusk-Lane and colleagues' (2018) CFA LPA 3-profile solution, which were both very similar, it appears the bESEM LPA's global factor attenuated the evident 'high,' 'medium,' and 'low' profiles accordingly. Although the results of DeBusk-Lane and colleagues' (2018) mixed methods study that examined the sources of writing self-efficacy between profiles seemingly mimic the present study's global factor profile differences, the very nature of the bESEM model likely situates participants in the LPA differently. That is, the factor scores of the bESEM model represent something entirely different, whereby participants not only differ between the specific factors but such that the specific factor scores represent differences left over from what the global factor modeled. This difference may result in a completely different modal classification for similar score configurations of the SEWS depending on the global factor score attached to each score configuration (e.g. a stronger set of scores across the factors results in a stronger global factor score; DeMars, 2013). Therefore, because each input variable in a latent profile analysis is treated equally in determining profiles, there are likely to be differences. This said, given the bESEM model better represents the data, a more accurate representation of the sources of writing self-efficacy may be needed.

In a similar thread as DeBusk-Lane and colleagues' (2018) research, the prevalence of profiles largely differentiated by generalized writing self-efficacy, and the inclusion now of identifiable specific factor differences, informs our current theoretical understanding of how students may exhibit differences in writing self-efficacy. Although the extant literature has largely focused effort to tease apart differences in writing self-efficacy by groupings between gender, race/ethnicity, and grade-levels, this current bESEM LPA approach uniquely provides a disaggregation to assess differences simply by how efficacious a student is (see Klassen & Usher, 2010; Pajares, 2003, 2007; Pajares & Valiante, 1999, 2001; Pajares, Valiante, Cheong, Hidi, & Boscolo, 2007; Usher & Pajares, 2008). This is not to negate prior research, as it has been greatly formative, but that the present research provides a novel lens for which to push the field forward in examining how students differ and to then also provide validity evidence based upon such past work.

It is important to remember while interpreting the profiles for how they impact our current theoretical understanding of writing self-efficacy, that the specific factors represent variability over-and-above the global factor (Chen et al., 2006). For instance, although profile-1 (*inefficacious*) exhibits a very low global factor mean, each specific factor mean represents scores derived while accounting for the global factor. To be clear here, these specific facet factor scores do not explicitly express higher scores on the SEWS, but higher scores relative to those with a similar global facet factor score. For example, in looking at the raw data, two participants that exhibit identical *ideation* factor scores of -1.133 actually have response patterns of [1, 0, 1] and [2, 1, 1] on the SEWS (for items 2, 6, and 7, respectively), and exhibit global facet factor scores of -1.37 and -0.304, respectively. Although these global factor scores represent the generalization across all 8 items, this example clearly demonstrates that the specific factor scores represent differences relative to or that which is not accounted for the global factor. Therefore, these profiles suggest that not only do student differentiate by basic *level* or global differences, but that the degree to which they exhibit differences on the specific factors is also different between profiles. Similar to the earlier stated theoretical implications and alignment to Bandura's (1997) contention that more specific beliefs are highly influenced by contextual and experiential factors, the results here further suggest that these differences are likely expressed differently throughout the continuum, if you will, of writing self-efficacy. Therefore, because efficacious beliefs are largely created and developed as a product of one's interpretation of the four sources (e.g. mastery experiences, vicarious experience, social persuasion, and physiological responses), the current findings suggest students within particular profiles undergo systematic or relatable experiences (Bandura, 1997; Usher & Pajres, 2008).

Therefore, these findings provide both theoretical support and evidence to extend theory. First, Bandura (1997) suggests that commonly held or generalized beliefs likely translate into more specifically held facets and the these two (generalized and specific beliefs) are inextricably connected. In other words, if a student generally holds less efficacy towards writing, they are also likely to naturally not be very efficacious towards more focused or specific skills associated with writing, such as punctuation or spelling. The present profiles demonstrate this well and support this notion, as both the *strongly inefficacious* and the *moderately inefficacious* profiles also exhibit less than average specific factor scores on a majority of specific factors. Despite this theoretical alignment, this study further suggests that within this connection or trend between generalized and specific beliefs, there exists rather cohesive groups of students who may exhibit systematic differences among the specific factors. Although this finding does not explicitly oppose theory, it suggests the relationship is not precisely linear within domain. Although future research would do well to examine *why* these profiles exhibit unique specific factor trends beyond their reported generalization of writing efficacy, I would posit that these unique profile trends are produced by differences in experience and interpretation. In other words, given the results from DeBusk-Lane and colleagues' (2018) work that found differences in not just the sources reported between profile, but the specific occasions or interpretations of sources they reported, it is likely that students who exhibit generally less (or more) efficacious beliefs of their writing ability interpret and develop their beliefs from disparate sources. From my perspective, to date no strong longitudinal evidence has been provided that examines the sources that create and develop efficacy beliefs in writing of students from varying levels of generalized beliefs. Provided the use of writing and how it is taught throughout the K-12 domain consistently changes, I would argue that this would be a worthwhile endeavor for future research to examine.

Ultimately, however, because person-centered approaches are a relatively new methodology in efficacy research at large, these results have little to compare against. To date, only one know study has been published and employed a bESEM LPA on efficacy data. Work by Perera, Calkins, and Part (2019) examined teacher efficacy profiles derived from a bESEM model (Perera, Wiens, McIlveen, Calkins, & McLenna, 2019). Although they state no major theoretical implications to efficacy research at large, their profiles largely resemble and exhibit similar *level* and *shape* effects as reported here. Together, both Perera and colleagues' (2019) study and the present study, support that efficacy exists and can be modeled both generally and specifically. Furthermore, their findings also provide some evidence of the connection between general and specific efficacy beliefs (e.g. in 4 out of 5 of their profiles, the general factor was associated with like valence specific factor means).

Despite these theoretical implications and connections to existing litetaure, this cursory view of the profiles only provides an initial understanding. Therefore, to better assess the extent to which these profiles translate and support well-established relationships and trends with other variables and constructs, a deeper look at what predicts membership into these profiles and how these profiles differentiate upon well-established outcomes is imperative to build further validity evidence of both the profile themselves and the adapted SEWS.

Research Question 4

RQ4 focused on examining concurrent and divergent/discriminant validity through a series of analyses that inspected predictors and outcomes. This section will first start with assessing the theoretical implication of the predictors and then the outcomes.

Theoretical importance and relation to extant research. Individually, each demographic predictor was assessed on the extent to which it predicted membership into each profile (see Table 11). Here, sex and disability were not significant predictors of profile membership, however, grade, minority, gifted, and ELL were all individually predictive. For instance, minority and gifted showed to be significant predictors of profile membership, whereby minority students were approximately 65% more likely to be in profile-2 relative to profile-3, and gifted students were approximately 60% less likely to be in profile-2 than 3. Depicted in Table 12, all demographic predictors were also assessed together to provide a more realistic depiction of which demographic variables predicts profile membership, controlling for the other demographic variables. In this case, sex, gifted, and disability were not significant predictors of profile membership. Minority students were reported as being approximately 70% more likely to be in profile-2 than profile-3, while ELL students were approximately 300% more likely, or about 4 times as likely to be in profile-1 than profile-3. Interestingly, for each one unit increase in grade, students have about a 50% greater likelihood of being in profile-1 relative to profile-2 and are approximately 35% more likely to be in the profile-3 when compared to 2, while controlling for all other demographics. This aligns well with the observation that writing efficacy beliefs and writing motivation in general tends to decline through the secondary school years (Klassen & Usher, 2010; Pajares & Usher, 2008; Pajares & Valiante, 1999; Pajares et al., 2007; Usher & Pajares, 2008), although the probability of membership into profile-1 versus profile-2 is an interesting point with the stark differences between *ideation*. If anything, because some students also exhibit higher probabilities of being in profile-3, relative to profile-2, by grade, perhaps this indicates beliefs diverge to some degree throughout these years of schooling. For sure, this data is cross-sectional and must be interpreted with caution in regard to differences between grades. Ultimately, these demographic predictors aligned in their anticipated directions, as ELL students would be expected to express less confidence in their writing ability and minority students, to whom are historically less confident in their writing ability (Pajares, 2003), would also be expected to have less confidence.

Despite the obvious trends that provide validity evidence of the SEWS, these results have implications for theory. Given the theoretical support and extensions provided by RQ3, these predictive trends even further the contributions to theory this study provides. In terms of predictors, the most obvious contribution is that these predictions largely replicate prior findings throughout literature and further substantiate the theoretical understanding of how personal factors influence the interpretation of one’s environment and therefore the promulgation of efficacious beliefs (Bandura, 2008; Pajares & Usher, 2008). Interestingly, the lack of statistical significance for sex, which has historically been reported to be a focal point in writing efficacy research (De Smedt et al., 2017; Pajares et al., 1999; Pajares & Valiante, 1999; 2001; Pajares et al., 2007; Vallalon et al., 2015), is, perhaps, the most surprising finding amongst the predictors.

This further substantiates DeBusk-Lane and colleagues' (2018) similar findings of non-significances to their CFA LPA, where they also assessed and controlled for ethnicity and grade. Although social cognitive theory does not ascribe gender specific properties of motivation or agency (Bussey & Bandura, 1999), expressed and reported differences throughout the literature suggest they may be a product of sociocultural stereotypes. Given this, a lack of predictive evidence herein suggests, perhaps, such stereotypes are declining. In other words, provided society, stereotypes, and the landscape of public education is consistently changing, may explain developments and differences exhibited by these findings. Alternatively, both DeBusk-Lane and colleagues (2018) and the present study found strong significant predictive effects associated with differences in grade. In both cases, those of higher grades are more likely to be in a less efficacious profile. However, the present findings also indicate a slightly stronger relationship of those in higher grades being predicted to be members of profile-3, the *high efficacious*. This may appear antithetical, but it also may suggest students become more differentiated as they progress through these grades. Nevertheless, social cognitive theory suggests developmental changes in efficacious beliefs are heavily influenced by ever changing, dynamic, and normative experiences that all mix with, inevitably, rapidly developing biological influences (Bandura, 1997). Considering the grade span of students in this study (8-10), these influences likely serve as rather robust influences, especially considering these students are experiencing rigorous standardized tests, the transition into high-school, and likely begin to become more specialized. In this last case, it would be expected that those who ascribe to and focus on more non-writing domains become less efficacious in their writing and account for some students of higher grades having a higher likelihood of membership in less efficacious profiles.

Aside from these demographics, I also assessed the predictive nature of those who were identified as gifted, having a disability, or being an English language learner. Surprisingly, neither those identified as being gifted or having a disability were significantly predictive, yet their coefficients trended in the anticipated direction (Frank Webb et al., 2016; Garcia & De Caso, 2004; Garchia & Fidalgo, 2008). English language learners, however, were significantly predictive of profile membership such that they had a higher likelihood of being members of the *strongly inefficacious* profile, as compared to the *efficacious* profile. Given prior literature in this areas, thought limited, these trends align and would be expected (Teng, Sun, & Xu, 2018).

Overall, the predictive associations reported here largely support both how theory purports self-efficacy to exist and develop and the ways in which existing writing self-efficacy research has portrayed similar relationships. This suggests, in a validity building effort, that the adapted SEWS is accurately, at least across these predictors, portraying writing self-efficacy accurately and in the same delineation as prior research. Although using bESEM derived factor scores in an LPA has, at face value, not specifically provided any outstanding theoretical updates or suggestions, using an improved model that more accurately depicted true score variability inevitably provides a more accurate depiction of profiles, especially considering the added utility the specific factor offer in interpretation. Nevertheless, it is also important to assess the profiles for further validity evidence across a number of other well known measures to continue to build an understanding of how well and accurately the SEWS is capturing writing self-efficacy.

To further provide validity evidence, I also examined the predictive value of both the WSES and the WAS-12 to profile membership. Interestingly, both measures were highly predictive across all profiles. More specifically, the WSES's basic skills factor seemingly echoed that of grade earlier, as those with higher basic skills were more likely to be in profile-1 compared to 2, yet also more likely to be in profile-3 than 2. This, along with grade differences, may suggest that as students gain more writing skills, they also become more efficacious and comfortable with, at least in regards to profile-1, writing *mechanics*. Comparatively, those with higher WSES advanced skills were more likely to be in profile-2 relative to 1, yet were similarly more likely to be in profile-3 relative to -2. This is to be expected, as the crosswalk between basic and advanced skills as operationalized by the WSES appears to translate well to the SEWS' *mechanics* and *ideation* factors, respectively. So in this case, it is logical for those with stronger 'advanced' writing skills beliefs to be more associated with membership in profile-2, relative to 1. Nevertheless, those with higher advanced skills scores were approximately 82% more likely to be in profile-3, relative to profile-1. As would be expected, diverging results of the WAS-12's affect (liking) and concern (writing anxiety), indicated that those with stronger affect towards writing exhibited stronger and significant predictions into more positive profiles (3>2>1). This inverse relationship is very much in-line with extant research between anxiety and writing self-efficacy (Pajares & Johnson, 1994; Pajares & Valiante, 1999; Goodman & Cirka, 2009; Limpo, 2018; Martinez et al., 2011; Sanders-Reio et al., 2014).

Collectively these results further substantiate the bESEM LPA profiles and provide validity evidence across demographic and other well-established metrics that have provided concurrent and divergent validity. Nevertheless, future research would do well to further investigate how self-efficacy changes across time, as these reported relations and predictions raise alarm as to the differences by grade.

To further establish validity evidence, I also assessed how the profiles responded to both the WSES and the WAS-12. As reported in Table 15, both factors of the WSES aligned with the global factor indicator in each profile. That is, participants reported less efficacy in less globally efficacious profiles. Interestingly, however, profile-1 and 2 exhibited fairly similar averages for both the basic and advanced factors (although basic was reported less efficacy for profile-2 than 1). Responses to the WAS-12's affect (liking) writing factor were in-line with the hypotheses, such that those with a stronger sense of efficacy towards writing exhibited a strong affliction towards writing. Conversely, those who reported less efficacy towards writing (members of lower profiles), exhibited a stronger relation to the concern factor of the WAS-12. These findings further provide validity evidence that the profiles, and the inherent bESEM model derived factor scores, are aligned to the well-established relationship between writing self-efficacy and writing apprehension, as well as to the most psychometrically established measure to-date. Furthermore, with the added specific factor differences between profiles, these results offer a unique relation not yet seen before. For instance, although it is logical and expected that those who express stronger 'concern' or apprehension for writing would be associated with the *strongly inefficacious* profile, these profiles now allow us to fully see specific factor variability and, perhaps, connect such apprehension to that profile's much lower than average affliction to efficacious beliefs associated with *ideation*. Although future researcher is needed, this level of inspection permits a new perspective on how writing self-efficacy and writing apprehension relate. Additionally, and because the other profiles also exhibit notable specific factor differences, these same type relations can be further assessed between these well-established predictors and outcomes and may offer new clues for future researchers.

In addition to these two scale predictors, I also captured and assessed the predictive value of each student's first quarter English grades. As would be expected, grades significantly predicted membership into efficaciously stronger profiles (1 vs. 3 and 2 vs. 3), however, no predictive relationship was found between higher grades and membership into either profile-1 or

2. Together, these results further substantiate the predictive findings and provide both concurrent and divergent validity evidence of both the profiles and the adapted SEWS' global indicator (Pajares & Johnson, 1994; Pajares & Valiante, 1999; Goodman & Cirka, 2009; Limpo, 2018; Martinez et al., 2011; Sanders-Reio et al., 2014). Aside from the utility in providing further validity evidence for the SEWS, these results suggest Q1 grades are largely predictive towards global efficacious beliefs, not a very robust predictor, and may be a less authentic metric of efficacy building mastery experiences. Said another way, teacher reported grades, though inherently limited in their own right due to teacher subjectivity (see Malouff & Thorsteinsson, 2016), may not well represent, and therefore predictive, student mastery experiences that readily translate into efficacious beliefs. To my knowledge no other study has fully examined the 'predictive' nature of prior English grades on writing self-efficacy, but a large body of literature does report mastery experiences as the most important and related sources used to create and develop efficacious beliefs. In examining source differences between profiles, DeBusk-Lane and colleagues (2018) found that mastery experiences were most found from those in the most efficacious profile, however, they did not assess the predictive nature of prior English grades. Therefore, the present study's findings may serve as a start to facilitate future research between prior 'grades,' what the grades mean, how they were derived, and the extent to which they may serve as adequate predictors of different levels of writing self-efficacy or profiles.

In a similar thread, a number of distal outcomes were assessed to establish predictive validity of the profiles. Although these analyses did not exactly perform a predictive statistical analysis (e.g. regression), the assessment of each variable's mean across each profile is telling. In this case, only grade 8 and 10 standardized writing test results were available at the time of writing this dissertation. Grade 8 total standardized writing scores mimicked earlier findings that have tended to find clear and statistically significant differences between profile-3, above, that of both profile-1 and 2. Although no clear differences were found among grade 8's Category 1 scores, Category 2 scores indicated that profile-1, which exhibited above average efficacious beliefs associated with writing *mechanics*, was significantly higher than profile-2 (which exhibited less than average *mechanics*). Considering Category 2 primarily involves editing for "…punctuation, and spelling," it is no surprise that those who exhibit stronger beliefs also perform better in this area. Grade 10 scores were reported in a similar manner across all three standardized test scores, also finding that Category 2 was higher for those who exhibited above average *mechanics*. In this case, using the bESEM model likely attenuated these differences and demonstrated the advantage of more accurately and precisely capturing specific factor differences among the profiles. As such, this implies that the relationships between writing self-efficacy and both grades' standardized writing scores may be more related to specific factor differences than generalized efficacy. This would make practical sense, as these standardized tests were largely focused on specific writing processes, such as editing. This highlights the importance of fully understanding that standardized tests may not fully tap into the entire writing process and may not relate differentially to students of varying levels of generalized efficacy beliefs associated with writing. This line of reasoning is not meant to negate that there were differences between the *efficacious* profile and the two lowest profiles, but that there were either no discernable differences between profile-1 and profile-2, or that profile-1 was exhibited stronger standardized category 2 scores than profile-2, despite profile-2 reporting stronger global efficacy. Although the predictive nature of the bESEM model was not assessed herein, the standardized test outcomes reported between profiles here may offer important clues as to the nature of such a prediction. Given writing self-efficacy has largely been positively associated with writing performance (see Pajares, 2003; Pajares et al., 2007), the present study adds further evidence of this, as there is a clear difference present between higher and lower efficacious profiles and the state-wide standardized writing scores. Furthermore, this also offers theoretical support for how scales should be developed. In other words, criterial alignment (*correspondence*), whereby the measures aligns with the performance outcome, often results in greater performance prediction (Bandura, 1997; 2006; Klassen & Usher, 2010; Marsh et al., 2018; Pajares, 1996). In this case, higher *mechanics* scores related to the performance outcome of the standardized test's category 2 outcome, which measured a student's ability to edit.

Collectively, these theoretical implications and relations to existing literature provide ample validity evidence for both the SEWS and the delineation of the profiles. Additionally, as compared to recent person-centered research, the present study's improved bESEM model LPA demonstrated and offered both practical utility and advancements in efficacy theory, as evidence through consistent alignment to well-established predictors and outcomes.

Limitations and Recommendation for Future Research

Although this study has employed a robust and analytically rigorous substantive methodological synergy towards examining construct-relevant psychometric multidimensionality in the adapted SEWS, it is not without limitations. This section will seek to identify characteristics of design or method that may have impacted or influenced the interpretations of the reported findings of the research. Furthermore, in most cases, the following limitations are aligned with suggestions for future research. These limitations and recommendations are reported in no specific order.

First, the data used in this study only represents student's beliefs encompassed within one school division, which likely limits external validity. Such localized data may inherently include particular environmental climate dispositions that differ from other school divisions locally or across the nation. Therefore, a more representative sample of participants from varying school divisions in different localities across the United States would likely better represent a normative student sample and therefore extend the interpretations of the findings herein.

Although the use of secondary data is often easier, it does not come without its own limitations. In this case, the precise form and way in which the data was collected was outside my full control, thereby potentially limiting the reliability due to outside sources of influence beyond my prevue. In these cases, such extrinsic influences are represented as measurement error (construct-irrelevant sources of error).

It is important to acknowledge and understand that removing an item to enable the bESEM to converge, limits its comparison to the other models in this study. Although a majority of the models in this research were derived from factor scores that represent the common variation among similarly themed items, I would be wrong to not acknowledge that all prior models in this study used that same item. On arguably equal footing, I would also be incorrect if I did not present an inspection of the failed bESEM model, the steps taken to examine the issues as they presented themselves, and ultimately further employ the alternative model (without item 1) considering the latent factor remaining meaningfully present and the item clearly lacked developmental appropriateness. Given this, future research should examine this item to ensure it is developmentally appropriate, functions adequately, and reliably expresses the construct in which it was developed. This is not to squarely recommend changes to the scale, but to strongly suggest that further psychometric investigation is clearly needed. Nevertheless, it must be acknowledged, that the interpretations and findings presented in this study are with a further adapted scale and therefore must be interpreted as such.

At the time of writing, the 9th grade PBA was not yet scored due to state level delays in finalizing scoring requirements and rubrics and was, therefore, unavailable for use as a distal outcome. Therefore, the validity evidence derived from outcomes across the profiles is limited to only examining standardized test results from 8th and 10th graders and may limit generalizability across the sample. Future research may incorporate this to further substantiate the findings across all grades.

Due to the obvious negative skew of the data across all items and by item means, whereby eight out of the 9 original item's raw mean above 2.5 (the calculated scale median), some amount of ceiling effects are present. Therefore, it may be unsurprising to see a larger profile that exhibits more positive beliefs. This may be a product of a reduced response scale (1-4), a lack of items that are developmentally appropriate, or items that do not provide appropriate levels of task demands that provide variability in responses. Future research should examine item functioning or offer comparisons with a similar sample with a larger or more broad response scale or further assess item appropriateness in terms of developmental appropriateness and the extent to which the questions provide ample and valid task demands to secondary students.

Although a robust global factor was evident and captured *level* effects, it must be understood that it only exists and represents variability across the given multidimensional scale. That is, the reported interpretations can only generalize to that which each factor and the overall scale were constructed to model, therefore perhaps limiting the full extent to which the results extend to writing self-efficacy as a whole. Because there exist many aspects associated with the various skills needed to write, it is logical that future research is needed to broaden and replicate these findings across other skill areas in writing or even other specific topic domains of writing (e.g. creative, research, or argumentative writing) to further establish the evidence to cumulatively inform the theoretical positioning of writing self-efficacy. On this same thread, other areas of efficacy research would also do well to follow suit, which would enable an even larger contribution to self-efficacy theory at large if, in fact, a global dimension is statistically evident.

Despite the 3-profile solution exhibiting statistical criteria and interpretability seemingly over that of the 4-profile arrangement, it is worth noting that the 4-profile solution may offer researchers and practitioners more detail in relation to writing self-efficacy of self-regulation. Picking and defending the final solution of profiles during enumeration is often not fully clear, lacks strong statistical reliance, and subjective. Future researchers would do well to fully examine the 4-profile solution, as it may offer a nuanced depiction of student writing self-efficacy not currently depicted in the 3-profile solution.

Although a well-fitting and interpretable bESEM model was reported, the validity and overall statistical extent to which the latent factors represented each set of items was not explored herein. Future research would do well to examine more robust statistical approaches to examining if each latent construct was reliable or exhibited construct replicability (Hancock & Meuller, 2001; Rodriguez et al., 2016). Such statistical tests as index *H*, which is defined as the sum of the ratios of the items' squared loadings (often explained to be the proportion of variance explained by the factor) on a particular factor to 1 minus the squared loading (unexplained variance), which represents a statistical method to examine construct reliability to judge how well a latent variance is represented by the items (Hancock & Meuller, 2001). Additionally, it would be beneficial to examine explained common variance (EVC), which assesses the unidimensionality of the common variance in a set of items to determine if a bifactor representation should actually, given a strong global factor, be treated as unidimensional (Reise, Scheines, Widaman, & Haviland, 2013; Ten Berge & Socˇan, 2004). Either way, future research would do well to fully statistically establish the appropriateness of a bifactor ESEM representation, as statistical support, aside from simply acknowledging the substantive value, is vital to ensuring the model is both accepted and appropriate to develop theory and be employed practically. Along this same initiative, future research would do well to also ensure the *ideation* factor is statistically meaningful. Using similar tests, research should examine whether this factor can be fully assumed by the global factor.

Even with a wide range of demographic predictors and various validity building outcomes, it will be important to determine if membership in the profiles themselves is predictive to the distal outcomes. That is, herein each outcome averages were examined between profiles, which is descriptive, but less informative than actually assessing the predictive value of the profiles themselves. Therefore, future research should include this analysis to better examine how well membership in certain profiles predicts higher or lower outcomes on important distal outcomes such as grades, standardized assessments, and future life success.

Furthermore, it is vital to examine the predictive value of each indicator to meaningful outcomes to better understand how influential both the G-factors and S-factors are. Future research, much like past research using SEM techniques, should examine how, once accounting for construct-relevant multidimensionality, each S-factor interacts and relates to other motivational constructs.

Future research would also do well to further assess the profile specific factor differences, especially among those of the *strongly inefficacious* and *moderately inefficacious* profiles. Delineating between these two, considering they are so different, would greatly aid in targeted interventions and better inform efficacy theory as to how students differentiate.

Implications for Educators

Although studies that employ, assess, and examine the psychometric properties of measures through advanced statistical techniques often lack clear and definitive implications for educators, the present study may offer important clues about the development and fostering of students' writing self-efficacy. As the findings demonstrate, students who exhibit strong confidence, or even appear doubtful, may also substantively differ on the extent to which they hold efficacious beliefs of writing's *mechanics*, the ability to develop and use ideas (*ideation*), or self-manage their writing process (*self-regulation*). Understanding these trends in the classroom may offer benefits in terms of targeting particular opportunities for students (Pajares, 1996, 2003, Villalon, Mateos, & Cuevas, 2015) to develop mastery experiences, while also acknowledging that students' efficacy beliefs may largely be held more generally towards writing. Although target versus more global interventions as a result of the present study's findings have yet to be fully actualized, this study may offer substantial clues that educators can act on today. For example, simply understanding that a rather substantial group of students who commonly view writing with less confidence simultaneously hold much less efficacious beliefs in relation to using and crafting ideas, suggests educators may do well to focus on creating, molding, developing, and employing ideas during writing tasks (more so than focus on writing *mechanics* or *self-regulation*).

Despite the statistical and theoretical value of determining which indicators from the bESEM model best predict meaningful and important outcomes, it has historically been held that simply improving writing self-efficacy should be "advanced as an explicit goal for writing instruction" (Bruning & Kaufman, 2015, p. 197; Usher & Pajares, 2008). This suggests that there is, without determining is these specific factors are highly predictive or not, great value in cultivating writing self-efficacy in general. As such, the present findings, which depict groups of students largely and importantly differentiated by a collective and global sense of efficacious beliefs towards writing, support the notion that most all efforts to foster stronger efficacy beliefs is viable to enhance students writing performance. This is not meant to denounce the present study's findings, but to clearly articulate that the robust presence of a global factor (that represents ~ 87% of the reliable variation) and the meaningful presence of the specific factors may suggest viable instructional pathways both globally and in a targeted sense that require future research to fully examine.

All of this together, these results further suggest adolescence is a dynamic period of development, but that there are clear trends that can be actionable for educators. Especially considering the diverging results as students increase in grade, better understanding the details or peculiarities of adolescent students' writing self-efficacy can be a powerful tool in identifying and supporting the transition to high-school, the marked changes in the use of writing, and the extent to which writing is relied upon by educators to capture student knowledge (Applebee & Langer, 2011). It is likely, given these robust changes, that many students transition from being confident writers, to not, simply because they lack the foundation of mastery experiences established from middle school. The results presented here further support this contention, but also offer ample avenues for future researchers to dive deeper, further assess why students may hold less efficacious beliefs towards writing, and arrange adequate interventions to re-establish and support stronger efficacious beliefs. Preparing students to employ writing, develop even further as writers, and become more competent as a result of such experiences is vital to their own learning, grades, success in college, and eventual work outcomes (Graham & Perin, 2007; National Commission on Writing, 2004, 2005). Without such, students are left at a disadvantage, often marginalized for future opportunities in classes and college, and often have lower grades (Graham, 2006). The results reported here help mitigate this by better understanding adolescent student writing confidence and offers various avenues for future researchers to further examine the nuances and trends exhibited by students as they enter high school.

Conclusion

> Modern statisticians are familiar with the notion that any finite body of data contains only a limited amount of information on any point under examination; that this limit is set by the nature of the data themselves, and cannot be increased by any amount of ingenuity expended in their statistical examination: that the statistician's task, in fact, is limited to the extraction of the whole of the available information on any particular issue. (Fisher, 1935, p. 44)

Along this theme, the present study demonstrated that current and widely used depictions of writing self-efficacy (e.g. CFA) are limited. This study showed that newer more representative models are better and offer a superior vantage to examining and accurately portraying more of the 'available information' provided by the adapted SEWS. To this point, a common and strong general factor exists among all the items of the SEWS, while the specific factors continue to be well represented. This suggests that writing self-efficacy simultaneously exists along both a collective spectrum of efficacious beliefs and expressed differentially among the original multidimensional factors of the SEWS. Furthermore, participants, though grouped into three profiles largely differentiated by global factor *shape*, exhibited unique and telling differences along the specific factors. Together, these findings provide ample evidence that the adapted SEWS contains construct-relevant psychometric multidimensionality as a product of both conceptual overlap between the specific factors and the existence of a global or generalized theme congruent to all items, therefore suggesting the often used, and arguably over-used, CFA depiction is less than optimal. These findings stand to greatly inform the theoretical understanding of writing self-efficacy, how it can best be fostered in classrooms, and ultimately how it relates to other motivational constructs to gain a more nuanced perspective of human motivational functioning.

References

Applebee, A. N. & Langer, J. A. (2011). A snapshot of writing instruction in middle schools and high schools. *English Journal, 100*(6), 14-27.

American Educational Research Association., American Psychological Association., National Council on Measurement in Education., & Joint Committee on Standards for Educational and Psychological Testing (U.S.). (2014). *Standards for educational and psychological testing*.

Anderson, D., Kahn, J. D., & Tindal, G. (2017). Exploring the robustness of a unidimensional item response theory model with empirically multidimensional data. *Applied Measurement in Education*, *30*(3), 163–177. https://doi.org/10.1080/08957347.2017.1316277

Asparouhov, T., & Muthén, B. (2009). Exploratory structural equation modeling. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(3), 397–438. https://doi.org/10.1080/10705510903008204

Asparouhov, T., Muthén, B., & Morin, A. J. S. (2015). Bayesian structural equation modeling with cross-loadings and residual covariances: Comments on Stromeyer et al. *Journal of Management*, *41*(6), 1561–1577. https://doi.org/10.1177/0149206315591075

Bakk, Z., & Vermunt, J. K. (2016). Robustness of stepwise latent class modeling with continuous distal outcomes. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(1), 20–31. https://doi.org/10.1080/10705511.2014.955104

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, *84*(2), 191.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *American Psychologist*, *37*(2), 122.

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist*, *28*(2), 117.

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York, NY: W H Freeman/Times Books/ Henry Holt & Co.

Bandura, A. (2001). Social cognitive theory: An agentic perspective. *Annual Review of Psychology*, *52*, 1–26.

Bandura, A. (2006). Toward a psychology of human agency, *1*(2), 17.

Bandura, A. (2008). An agentic perspective on positive psychology. In S. J. Lopez (Ed.), *Positive psychology: Exploring the best in people: Discovering human strengths*. (Vol. 1), New York: Praeger.

Bandura, A. (2012). On the functional properties of perceived self-efficacy revisited. *Journal of Management*, *38*(1), 9–44. https://doi.org/10.1177/0149206311410606

Bandura, A. (2018). Toward a psychology of human agency: Pathways and reflections. *Perspectives on Psychological Science*, *13*(2), 130–136. https://doi.org/10.1177/1745691617699280

Bandura, A., & Jourden, F. J. (1991). Self-regulatory mechanisms governing the impact of social comparison on complex decision making. *Journal of Personality & Social Psychology*, *60*(6), 941–951.

Barendse, M. T., Oort, F. J., & Timmerman, M. E. (2015). Using exploratory factor analysis to determine the dimensionality of discrete responses. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(1), 87–101. https://doi.org/10.1080/10705511.2014.934850

Bauer, D. J. (2007). Observations on the use of growth mixture models in psychological research. *Multivariate Behavioral Research, 42*, 757–786.

Bauer, D. J., & Curran, P. J. (2004). The integration of continuous and discrete latent variable models: Potential problems and promising opportunities. *Psychological Methods, 9*(1), 3-29. http://dx.doi.org/10.1037/1082-989X.9.1.3

Beauducel, A., & Herzberg, P. Y. (2006). On the performance of maximum likelihood versus means and variance adjusted weighted least squares estimation in CFA. *Structural Equation Modeling: A Multidisciplinary Journal*, *13*(2), 186–203. https://doi.org/10.1207/s15328007sem1302_2

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238-246. http://dx.doi.org/10.1037/0033-2909.107.2.238

Bline, D., Lowe, D. R., Meixner, W. F., Nouri, H., & Pearce, K. (2001). A research note on the dimensionality of Daly and Miller's writing apprehension scale. *Written Communication, 18,* 61-79.

Bolck, Croon, & Hagenaars, J. A. (2004). Estimating latent structure models with categorical variables: One-step versus three-step estimators. *Political Analysis*, *12*(1), 3–27.

Bong, M., & Skaalvik, E. M. (2003). Academic self-concept and self-efficacy: How different are they really? *Educational Psychology Review, 40*.

Brennan, R. L. (2010). Generalizability theory and classical test theory. *Applied Measurement in Education*, *24*(1), 1–21. https://doi.org/10.1080/08957347.2011.532417

Brown, T. A. (2015). *Methodology in the social sciences. Confirmatory factor analysis for applied research (2nd ed.).* New York, NY, US: The Guilford Press.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*(1), 111-150.

Browne, M. W. (2001). An overview of analytic rotation in exploratory factor analysis. *Multivariate Behavioral Research, 36*, 111-150.

Bruning, R., & Horn, C. (2000). Developing motivation to write. *Educational Psychologist*, *35*(1), 25–37. https://doi.org/10.1207/S15326985EP3501_4

Bruning, R., Dempsey, M., Kauffman, D. F., McKim, C., & Zumbrunn, S. (2013). Examining dimensions of self-efficacy for writing. *Journal of Educational Psychology*, *105*(1), 25–38. https://doi.org/10.1037/a0029692

Bussey, K., & Bandura, A. (1999). Social cognitive tehory of ender development and differentiation. *Psychological Review*, *106*(4), 676–713.

Cattell, J. M. (1893). Mental measurement. *The Philosophical Review*, *2*(3), 316. https://doi.org/10.2307/2175386

Cattell, R.B. (1978) *The scientific use of factor analysis in behavioral and life sciences*. Plenum, New York.

Chen, F. F., West, S., & Sousa, K. (2006). A comparison of bifactor and second-order models of quality of life. *Multivariate Behavioral Research*, *41*(2), 189–225. https://doi.org/10.1207/s15327906mbr4102_5

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

Chen, Q., Luo, W., Palardy, G. J., Glaman, R., & McEnturff, A. (2017). The efficacy of common fit indices for enumerating classes in growth mixture models when nested data structure is ignored: A monte carlo study. *SAGE Open*. https://doi.org/10.1177/2158244017700459

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233-255. http://dx.doi.org/10.1207/S15328007SEM0902_5

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Earlbaum Associates.

Collins, L. M., & Lanza, S. T. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences.* New York: Wiley.

Daly, J. A. (1978). Writing apprehension and writing competency. *The Journal of Educational Research, 72*, 10–14.

DeMars, C. E. (2013). A Tutorial on Interpreting Bifactor Model Scores. *International Journal of Testing*, *13*(4), 354–378. https://doi.org/10.1080/15305058.2013.799067

De Smedt, F., Graham, S., & Van Keer, H. (2018). The bright and dark side of writing motivation: Effects of explicit instruction and peer assistance. *The Journal of Educational Research*, 1–16. https://doi.org/10.1080/00220671.2018.1461598

De Smedt, F., Merchie, E., Barendse, M., Rosseel, Y., Van Keer, H., & De Naeghel, J. (2017). Cognitive and motivational challenges in writing: Studying the relation with writing performance across students' gender and achievement level. *Reading Research Quarterly*. https://doi.org/10.1002/rrq.193

De Smedt, F., Van Keer, H., & Merchie, E. (2016). Student, teacher and class-level correlates of Flemish late elementary school children's writing performance. *Reading and Writing*, *29*(5), 833–868. https://doi.org/10.1007/s11145-015-9590-z

de Winter, J. C. F., Dodou, D., & Wieringa, P. A. (2009). Exploratory factor analysis with small sample sizes. *Multivariate Behavioral Research*, *44*(2), 147–181. https://doi.org/10.1080/00273170902794206

DeBusk-Lane, M., Lester, A., & Zumbrunn, S. (2019). *Understanding the complexity of writing self-efficacy through mixed methods.* Manuscript in preparation.

Deng, L., & Chan, W. (2017). Testing the difference between reliability coefficients alpha and omega. *Educational and Psychological Measurement*, *77*(2), 185–203. https://doi.org/10.1177/0013164416658325

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods, 5*(2), 155-174. http://dx.doi.org/10.1037/1082-989X.5.2.155

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. https://doi.org/10.1080/13562517.2014.974026

Engelhard, G., & Behizadeh, N. (2012). Exploring the alignment of writing self-efficacy with writing achievement using rasch measurement theory and qualitative methods. *Journal of Applied Measurement*, *13*(2), 132–145.

Fabrigar, L. R., & Wegener, D. T. (2011). *Exploratory Factor Analysis*. Oxford University Press. https://doi.org/10.1093/acprof:osobl/9780199734177.001.0001

Fan, X. & Sivo, S. (2005). Sensitivity of fit indexes to misspecified structural or measurement model components: Rational of two-index strategy revisited. *Structural Equation Modeling: A Multidisciplinary Journal, 12*(3).

Fan, X., & Sivo, S. A. (2007). Sensitivity of fit indices to model misspecification and model types. *Multivariate Behavioral Research, 42*, 509-529. http://dx.doi.org/10.1080/00273170701382864

Finney, S. J., & DiStefano, C. (2013). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Quantitative methods in education and the behavioral sciences: Issues, research, and teaching. Structural equation modeling: A second course* (pp. 439-492). Charlotte, NC, US: IAP Information Age Publishing.

Finney, S., & DiStefano, C. (2006). Non-normal and categorical data in structural equation modeling. In G.R. Hancock & R.O. Mueller (Eds.), *Structural equation modeling: A second course* (pp. 269–314). Greenwich: Information Age.

Fisher, R. A. (1935). *The design of experiments*. Oliver and Boyd.

Gagne, P., & Hancock, G. R. (2006). Measurement model quality, sample size, and solution propriety in confirmatory factor models. *Multivariate Behavioral Research*, *41*(1), 65–83. https://doi.org/10.1207/s15327906mbr4101_5

García, J.-N., & Fidalgo, R. (2008). Writing self-efficacy changes after cognitive strategy intervention in students with learning disabilities: The mediational role of gender in calibration. *The Spanish Journal of Psychology*, *11*(2), 414–432. https://doi.org/10.1017/S1138741600004431

Geiser, C. (2013). *Methodology in the social sciences. Data analysis with Mplus.* New York, NY, US: Guilford Press.

Gignac, G. E. (2006). A confirmatory examination of the factor structure of the multidimensional aptitude battery: Contrasting oblique, higher-order, and nested factor models. *Educational and Psychological Measurement, 66*, 136–145.

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. Wiine (Eds.), *Hand book of educational psychology* (pp. 457–477). Mahwah, NJ: Erlbaum.

Graham, S. (2018). A revised writer(s)-within-community model of writing. *Educational Psychologist*, 1–22. https://doi.org/10.1080/00461520.2018.1481406

Graham, S., Harris, K. R., Kiuhara, S. A., & 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), 000–000.

Graham, S., McKeown, D., Kiuhara, S., & Harris, K. R. (2012). A meta-analysis of writing instruction for students in the elementary grades. *Journal of Educational Psychology*, *104*(4), 879–896. https://doi.org/10.1037/a0029185

Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, *99*(3), 445–476. https://doi.org/10.1037/0022-0663.99.3.445

Grimm, K., Steele, J., Ram, N. & Nesselroade, J. (2013). Exploratory latent growth models in the structural equation modeling framework. *Structural Equation Modeling: A Multidisciplinary Journal, 20*(4), 568-591.

Guilford, J. P. (1954). *Psychometric methods* (2nd ed.). New York, NY, US: McGraw-Hill.

Gustafsson, J.-E., & Balke, G. (1993). General and specific abilities as predictors of school achievement. *Multivariate Behavioral Research, 28*(4), 407-434. http://dx.doi.org/10.1207/s15327906mbr2804_2

Hallquist, M. N., & Wiley, J. F. (2018). MplusAutomation: An r package for facilitating large-scale latent variable analyses in mplus. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–18. https://doi.org/10.1080/10705511.2017.1402334

Hancock, G. R., & Mueller, R. O. (2001). *Rethinking Construct Reliability within Latent Variable Systems.* In R. Cudeck, S. du Toit, & D. S?rbom (Eds.), Structural Equation Modeling: Present und Future—A Festschrift in Honor of Karl Joreskog (pp. 195-216). Lincolnwood, IL: Scientific Software International.

Harris, K. R., & Graham, S. (2016). Self-regulated strategy development in writing: Policy implication of an evidence-based practice. *Policy Insights from the Behavioral and Brain Sciences*, *3*(1), 88–84.

Hershberger, S. L., & Marcoulides, G. A. (2013). The problem of equivalent structural models. In G. R. Hancock & R. O. Mueller (Eds.), *Quantitative methods in education and the behavioral sciences: Issues, research, and teaching. Structural equation modeling: A second course* (pp. 3-39). Charlotte, NC, US: IAP Information Age Publishing.

Henson, R. K. (2002). From adolescent angst to adulthood: Substantive implications and measurement dilemmas in the development of teacher efficacy research. *Educational Psychologist*, *37*(3), 137–150. https://doi.org/10.1207/S15326985EP3703_1

Horn, J. & McArdle, J. (1992). A practical and theoretical guide to measurement invariance in aging research. *Experimental Aging Research,* 18. 117-144.

Hu, L. & Bentler, P. M. (1998). Fit indices in covariance structure modeling: Sensitivity to underparameterized model misspecification. *Psychological Methods, 3*(4), 424-453. http://dx.doi.org/10.1037/1082-989X.3.4.424

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55. http://dx.doi.org/10.1080/10705519909540118

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

Jones, E. (2008). Predicting performance in first-semester college basic writers: Revisiting the role of self-beliefs. *Contemporary Educational Psychology*, *33*(2), 209–238. https://doi.org/10.1016/j.cedpsych.2006.11.001

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass*, *2*(1), 302–317. https://doi.org/10.1111/j.1751-9004.2007.00054.x

Kitayama, S., & Uskul, A. K. (2011). Culture, mind, and the brain: Current evidence and future directions. *Annual Review of Psychology*, *62*(1), 419–449. https://doi.org/10.1146/annurev-psych-120709-145357

Klassen, R. M., & Usher, E. L. (2010). Self-efficacy in educational settings: Recent research and emerging directions. In T. C. Urdan & S. A. Karabenick (Eds.), *Advances in Motivation and Achievement* (Vol. 16, pp. 1–33). Emerald Group Publishing Limited. https://doi.org/10.1108/S0749-7423(2010)000016A004

Klassen, R. M., Usher, E. L., & Bong, M. (2010). Teachers' collective efficacy, job satisfaction, and job stress in cross-cultural context. *The Journal of Experimental Education*, *78*(4), 464–486. https://doi.org/10.1080/00220970903292975

Kline, R. B. (2016). *Methodology in the social sciences. Principles and practice of structural equation modeling (4th ed.).* New York, NY, US: Guilford Press.

Laursen, B., & Joff, E. (2006). Person-centered and variable-centered approaches to longitudinal data. *Merrill-Palmer Quarterly, 52*(3).

Limpo, T. (2018). Development of a short measure of writing apprehension: Validity evidence and association with writing frequency, process, and performance. *Learning and Instruction*, *58*, 115–125. https://doi.org/10.1016/j.learninstruc.2018.06.001

Limpo, T., & Alves, R. A. (2017). Relating beliefs in writing skill malleability to writing performance: The mediating role of achievement goals and self-efficacy. *Journal of Writing Research, 9*(2), 97–124. https://doi.org/10.17239/jowr-2017.09.02.01

Lord, F.M., Novick, M.R., & Birnbaum, A. (1968). *Statistical theories of mental test scores.* Oxford, England: Addison-Wesley.

Lubke, G. H., & Muthén, B. (2005). Investigating population heterogeneity with factor mixture models. *Psychological Methods*, *10*(1), 21–39. https://doi.org/10.1037/1082-989X.10.1.21

MacArthur, C. A., Philippakos, Z. A., & Ianetta, M. (2015). Self-regulated strategy instruction in college developmental writing. *Journal of Educational Psychology*, *107*(3), 855–867.

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods, 1,* 130-149.

MacCallum, R., Browne, M., & Sugawara, H. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2). 130-149.

MacCallum, R., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods*, *4*(1), 84–99.

Malouff, J. M., & Thorsteinsson, E. B. (2016). Bias in grading: A meta-analysis of experimental research findings. *Australian Journal of Education*, *60*(3), 245–256. https://doi.org/10.1177/0004944116664618

Marsh, H. W., & Hau, K.-T. (2007). Applications of latent-variable models in educational psychology: The need for methodological-substantive synergies. *Contemporary Educational Psychology*, *32*(1), 151–170. https://doi.org/10.1016/j.cedpsych.2006.10.008

Marsh, H. W., Hau, K.-T., & Grayson, D. (2005). Goodness of fit in structural equation models. In A. Maydeu-Olivares & J. J. McArdle (Eds.), *Multivariate applications book series. Contemporary psychometrics: A festschrift for Roderick P. McDonald*(pp. 275-340). Mahwah, NJ, US: Lawrence Erlbaum Associates Publishers.

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J. S., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment*, *22*(3), 471–491. https://doi.org/10.1037/a0019227

Marsh, H. W., Lüdtke, O., Nagengast, B., Morin, A. J. S., & von Davier, M. (2013). Why item parcels are (almost) never appropriate: Two wrongs do not make a right—Camouflaging misspecification with item parcels in CFA models. *Psychological Methods*, *18*(3), 257–284. https://doi.org/10.1037/a0032773

Marsh, H. W., Lüdtke, O., Trautwein, U., & Morin, A. J. S. (2009). Classical latent profile analysis of academic self-concept dimensions: Synergy of person- and variable-centered approaches to theoretical models of self-concept. *Structural Equation Modeling: A Multidisciplinary Journal*, *16*(2), 191–225. https://doi.org/10.1080/10705510902751010

Marsh, H. W., Nagengast, B., & Morin, A. J. S. (2013). Measurement invariance of big-five factors over the life span: ESEM tests of gender, age, plasticity, maturity, and La Dolce Vita effects. *Developmental Psychology*, *49*, 1194–1218.

Marsh, H. W., Pekrun, R., Parker, P. D., Murayama, K., Guo, J., Dicke, T., & Arens, A. K. (2018). The murky distinction between self-concept and self-efficacy: Beware of lurking jingle-jangle fallacies. *Journal of Educational Psychology*. https://doi.org/10.1037/edu0000281

Marsh, H. W., Scalas, L. F., & Nagengast, B. (2010). Longitudinal tests of competing factor structures for the Rosenberg Self-Esteem Scale: Traits, ephemeral artifacts, and stable response styles. *Psychological Assessment, 22,* 366–381. doi:10.1037/a0019225

Martinez, C., Nock, N., & Cass, J. (2011). Pain and pleasure in short essay writing: Factors predicting university students' writing anxiety and writing self-efficacy. *Journal of Adolescent & Adults Literacy*, *54*(5), 351–360.

Masyn, T. D. (Ed.). (2013). *The Oxford handbook of quantitative methods*. New York: Oxford University Press.

McCarthy, P., Meier, S., & Rinderer, R. (1985). Self-efficacy and writing: A different view of self-evaluation. *College Composition and Communication*, *36*(4), 465. https://doi.org/10.2307/357865

McCrae, R. R., Zonderman, A. B., Costa, P., Bond, M. H., & Paunonen, S. V. (1996). Evaluating replicability of factors in the revised neo personality inventory: Confirmatory factor analysis versus Procrustes rotation. *Journal of Personality and Social Psychology*, *70*(3), 552-566.

McMillan, J. (2016). *Fundamentals of educational research*. Pearson.

McNeish, D. (2017). Thanks coefficient alpha, we'll take it from here. *Psychological Methods*. https://doi.org/10.1037/met0000144

Mellenbergh, G. J. (1996). Measurement precision in test score and item response models. *Psychological Methods, 1*(3), 293-299.

Meredith, W., & Teresi, J. (2006). An essay on measurement and factorial invariance. *Medical Care, 44*(11). 69-77.

Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, *50*(9), 741.

Meyer, J. P., & Morin, A. J. S. (2016). A person-centered approach to commitment research: Theory, research, and methodology: Person-Centered Commitment Research. *Journal of Organizational Behavior*, *37*(4), 584–612. https://doi.org/10.1002/job.2085

Morin, A. J. S., & Marsh, H. W. (2015). Disentangling shape from level effects in person-centered analyses: An illustration based on university teachers' multidimensional profiles of effectiveness. *Structural Equation Modeling: A Multidisciplinary Journal*, *22*(1), 39–59. https://doi.org/10.1080/10705511.2014.919825

Morin, A. J. S., Arens, A. K., & Marsh, H. W. (2016). A bifactor exploratory structural equation modeling framework for the identification of distinct sources of construct-relevant psychometric multidimensionality. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(1), 116–139. https://doi.org/10.1080/10705511.2014.961800

Morin, A. J. S., Arens, A. K., Tran, A., & Caci, H. (2016a). Exploring sources of construct-relevant multidimensionality in psychiatric measurement: A tutorial and illustration using the composite scale of morningness. *International Journal of Methods in Psychiatric Research, 25*, 277–288. http://dx.doi.org/10.1002/mpr.1485.

Morin, A. J. S., Boudrias, J. S., Marsh, H. W., Madore, I., & Desrumaux, P. (2016). Further reflections on disentangling shape and level effects in person-centered analyses: An illustration exploring the dimensionality of psychological health. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(3), 438–454. https://doi.org/10.1080/10705511.2015.1116077

Morin, A. J. S., Boudrias, J.-S., Marsh, H. W., McInerney, D. M., Dagenais-Desmarais, V., Madore, I., & Litalien, D. (2017). Complementary variable- and person-centered approaches to the dimensionality of psychometric constructs: Application to psychological wellbeing at work. *Journal of Business and Psychology*, *32*(4), 395–419. https://doi.org/10.1007/s10869-016-9448-7

Morin, A. J. S., Morizot, J., Boudrias, J.-S., & Madore, I. (2011). A multifoci person-centered perspective on workplace affective commitment: A latent profile/factor mixture analysis. *Organizational Research Methods*, *14*(1), 58–90. https://doi.org/10.1177/1094428109356476

Morin, A.J.S., & Asparouhov, T. (2018). *Estimation of a hierarchical Exploratory Structural Equation Model (ESEM) using ESEM-within-CFA*. Montreal, QC: Substantive Methodological Synergy Research Laboratory.

Muthén, B. (2002). Using mplus monte carlo simulations in practice: A note on assessing estimation quality and power in latent variable models. *Webnote 1*.

Muthén, L.K. and Muthén, B.O. (1998-2017). *Mplus user's guide.* Eighth Edition. Los Angeles, CA: Muthén & Muthén

National Commission on Writing. (2004, September). *Writing: A ticket to work. . . or a ticket out: A survey of business leaders.* Available www.collegeboard.com

National Commission on Writing. (2005, July). *Writing: A powerful message from State government.* Retrieved from www.collegeboard.com

Nylund-Gibson, K., & Choi, A. Y. (2018). Ten frequently asked questions about latent class analysis. *Translational Issues in Psychological Science*, *4*(4), 440–461. https://doi.org/10.1037/tps0000176

Nylund-Gibson, K., & Masyn, K. E. (2016). Covariates and mixture modeling: Results of a simulation study exploring the impact of misspecified effects on class enumeration. *Structural Equation Modeling: A Multidisciplinary Journal*, *23*(6), 782–797. https://doi.org/10.1080/10705511.2016.1221313

Nylund, K. L., Asparouhov, T., & Muthén, B. O. (2007). Deciding on the number of classes in latent class analysis and growth mixture modeling: A monte carlo simulation study. *Structural Equation Modeling: A Multidisciplinary Journal*, *14*(4), 535–569. https://doi.org/10.1080/10705510701575396

Pajares, F., & Valiante, G. (2006). Self-efficacy beliefs and motivation in writing development. In C. MacArthur, S. Graham, & J. Fitzgerald (Eds.), *Handbook of writing research* (pp. 158-170). New York, NY: Guilford Press.

Pajares, F., Hartley, J., & Valiante, G. (2001). Response format in writing self-efficacy assessment: greater discrimination increases prediction. *Measurement and Evaluation in Couseling and Development*, *33*, 214–221.

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

Pajares, Frank, & Johnson, M. J. (1996). Self-efficacy beliefs and the writing performance of entering high school students. *Psychology in the Schools*, *33*, 163–175.

Pajares, Frank, & Miller, M. D. (1995). Mathematics self-efficacy and mathematics performances: The need for specificity of assessment. *Journal of Counseling Psychology*, *42*(2), 190–198.

Pajares, Frank, & Schunk, D. H. (2002). Self and self-belief in psychology and education: A historical perspective. In *Improving academic achievement* (p. 20).

Pajares, Frank, & Usher, E. L. (2008). Self-efficacy, motivation, and achievement in school from the perspective of reciprocal determinism. In *Advances in motivation and achievement* (Vol. 15, pp. 391–423). Bingley, United Kingdom: Emerald Group Publishing Limited.

Pajares, Frank, & Valiante, G. (1997). Influence of self-efficacy on elementary students' writing. *The Journal of Educational Research*, *90*(6), 353–360.

Pajares, Frank, & 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, Frank, & Valiante, G. (2001). Gender differences in writing motivation and achievement of middle school students: A function of gender orientation? *Contemporary Educational Psychology*, *26*(3), 366–381. https://doi.org/10.1006/ceps.2000.1069

Pajares, Frank, Britner, S. L., & Valiante, G. (2000). Relation between achievement goals and self-beliefs of middle school students in writing and science. *Contemporary Educational Psychology*, *25*(4), 406–422. https://doi.org/10.1006/ceps.1999.1027

Pajares, Frank, 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, Frank, Miller, M. D., & Johnson, M. J. (1999). Gender differences in writing self-beliefs of elementary school students. *Journal of Educational Psychology*, *91*(1), 50.

Pajares, Frank, Valiante, G., & Cheong, Y. F. (2007). Writing self-efficacy and its relation to gender, writing motivation and writing competence: A developmental perspective. In S. Hidi & P. Boscolo (Eds.), *Writing and motivation*. Elsevier.

Pajares, Frank. (1996). Self-efficacy beliefs in academic settings. *Review of Educational Research*, *66*(4), 543. https://doi.org/10.2307/1170653

Pajares, Frank. (2002). Gender and perceived self-efficacy in self-regulated learning. *Theory Into Practice*, *41*(2), 116–125.

Pajares, Frank. (2003). Self-efficacy beliefs, motivation, and achievement in writing: A review of the literature. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, *19*(2), 139–158. https://doi.org/10.1080/10573560308222

Pajares, Frank. (2007). Empirical properties of a scale to assess writing self-efficacy in school contexts. *Measurement and Evaluation in Counseling and Development*, *39*(4), 239.

Perera, H. N., Calkins, C., & Part, R. (2019). Teacher self-efficacy profiles: Determinants, outcomes, and generalizability across teaching level. *Contemporary Educational Psychology*, *58*, 186–203. https://doi.org/10.1016/j.cedpsych.2019.02.006

Perera, H. N., Wiens, P. D., McIlveen, P., Calkins, C., & McLennan, B. (2019). A unifying perspective on the dimensionality of teacher self-efficacy: Evidence from Australian and US teachers. [submitted for publication].

Petras, H., & Masyn, K. (2010). General growth mixture analysis with antecedents and consequences of change. In A. Piquero, & D. Weisburd (Eds.), *Handbook of Quantitative Criminology*. New York: Springer-Verlag.

Petras, H.; Masyn, K. General growth mixture models. In: Piquero, A.; Weisburd, D., editors. *Handbook of quantitative criminology*. Springer; New York: 2010. p. 69-100.

Prat-Sala, M., & Redford, P. (2010). The interplay between motivation, self-efficacy, and approaches to studying. *British Journal of Educational Psychology*, *80*(2), 283–305. https://doi.org/10.1348/000709909X480563

R Core Team (2019). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.URL http://www.R-project.org/.

Ramos-Villagrasa, P. J., Sánchez-Iglesias, I., Grande-de-Prado, M., Oliván-Blázquez, B., Martín-Peña, J., & Cancer-Lizaga, P. (2018). Spanish version of “self-efficacy for writing scale” (SEWS). *Anales de Psicología*, *34*(1), 86. https://doi.org/10.6018/analesps.34.1.264931

Rasch, G. (1960). Probabilistic models for some intelligence and attainments tests. Copenhagen: Danish Institute for Educational Research.

Reeve, S., Kitchen, E., Sudweeks, R., Bell, J., & Bradshaw, W. (2011). Development of an instrument for measuring self-efficacy in cell biology. *Journal of Applied Measurement, 12*(3). 242-260.

Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, *47*(5), 667–696. https://doi.org/10.1080/00273171.2012.715555

Reise, S. P., Bonifay, W. E., & Haviland, M. G. (2013). Scoring and modeling psychological measures in the presence of multidimensionality. *Journal of Personality Assessment*, *95*(2), 129–140. https://doi.org/10.1080/00223891.2012.725437

Reise, S. P., Moore, T. M., & Haviland, M. G. (2010). Bifactor models and rotations: Exploring the extent to which multidimensional data yield univocal scale scores. *Journal of Personality Assessment*, *92*(6), 544–559. https://doi.org/10.1080/00223891.2010.496477

Reise, S., Moore, T., & Maydeu-Olivares, A. (2011). Target rotations and assessing the impact of model violations on the parameters of unidimensional item response theory models. *Educational and Psychological Measurement, 71*(4), 684-711. http://dx.doi.org/10.1177/0013164410378690

Reise, S. P., Scheines, R., Widaman, K. F., & Haviland, M. G. (2013). Multidimensionality and structural coefficient bias in structural equation modeling: A bifactor perspective. *Educational and Psychological Mea- surement, 73,* 5–26. http://dx.doi.org/10.1177/0013164412449831

Rindskopf, D., & Rose, T. (1988). Some theory and applications of confirmatory second-order factor analysis. *Multivariate Behavioral Research*, *23*(1), 51–67. https://doi.org/10.1207/s15327906mbr2301_3

Sanders-Reio, J., Alexander, P. A., Reio, 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. https://doi.org/10.1016/j.learninstruc.2014.02.001

Sass, D. A., Schmitt, T. A., & Marsh, H. W. (2014). Evaluating model fit with ordered categorical data within a measurement invariance framework: A comparison of estimators. *Structural Equation Modeling: A Multidisciplinary Journal*, *21*(2), 167–180. https://doi.org/10.1080/10705511.2014.882658

Schmidt, K., & Alexander, J. (2012). The empirical development of an instrument to measure writerly self-eficacy in writing centers. *The Journal of Writing Assessment*, *5*(2). Retrieved from http://www.journalofwritingassessment.org/article.php?article=62

Schunk, D. H. (2003). Self-efficacy for reading and writing: Influence of modeling, goal setting, and self-evaluation. *Reading & Writing Quarterly*, *19*(2), 159–172. https://doi.org/10.1080/10573560308219

Schunk, D. H., & Usher, E. L. (2012). Social cognitive theory and motivation. In R. M. Ryan (Ed.), *Oxford library of psychology. The Oxford handbook of human motivation* (pp. 13-27). New York, NY, US: Oxford University Press.

Schunk, D., & DiBenedetto, M. (2016). Self-efficacy theory in education In K. R. Wentzel & D. B. Miele (Eds.), *Handbook of motivation at school.* Abingdon: Routledge

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*(3), 386–398.

Shell, D. F., Murphy, C., & Bruning, R. H. (1989). Self-efficacy and outcome expectancy mechanisms in reading and writing achievement. *Journal of Educational Psychology*, *81*(1), 91–100.

Smith, E. V., Wakely, M. B., de Kruif, R. E. L., & Swartz, C. W. (2003). Optimizing rating scales for self-efficacy (and other) research. *Educational and Psychological Measurement*, *63*(3), 369–391. https://doi.org/10.1177/0013164403063003002

Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*(2), 173-180.

Steinley, D., & McDonald, R. P. (2007). Examining factor score distributions to determine the nature of latent spaces. *Multivariate Behavioral Research*, *42*(1), 133–156. https://doi.org/10.1080/00273170701341217

Stewart, G., Seifert, 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).

Ten Berge, J. M., & Socˇan, G. (2004). The greatest lower bound to the reliability of a test and the hypothesis of unidimensionality. *Psychometrika, 69,* 613–625. http://dx.doi.org/10.1007/BF02289858

Török, R., Tóth-Király, I., Bőthe, B., & Orosz, G. (2017). Analyzing models of career decision self-efficacy: First-order, hierarchical, and bifactor models of the career decision self-efficacy scale. *Current Psychology*, *36*(4), 764–773. https://doi.org/10.1007/s12144-016-9464-9

Troia, G. A., Harbaugh, A. G., Shankland, R. K., Wolbers, K. A., & Lawrence, A. M. (2013). Relationships between writing motivation, writing activity, and writing performance:

effects of grade, sex, and ability. *Reading and Writing*, *26*(1), 17–44. https://doi.org/10.1007/s11145-012-9379-2

Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1-10. http://dx.doi.org/10.1007/BF02291170

Usher, E. L. (2015). Personal capability beliefs. In *Handbook of Educational Psychology* (p. 15).

Usher, E. L., & Pajares, F. (2008). Self-efficacy for self-regulated learning: A validation study. *Educational and Psychological Measurement*, *68*(3), 443–463. https://doi.org/10.1177/0013164407308475

Usher, E. L., & Schunk, D. H. (2018). Social cognitive theoretical perspective of self-regulation. In D. H. Schunk & J. A. Greene (Eds.), *Educational psychology handbook series. Handbook of self-regulation of learning and performance* (pp. 19-35). New York, NY, US: Routledge/Taylor & Francis Group.

Usher, E. L., & Weidner, B. (2018). Sociocultural influences on self-efficacy development. In *Big theories revisited* (pp. 141–164). Information Age Publishing, Inc.

Valentine, J. C., DuBois, D. L., & Cooper, H. (2004). The relation between self-beliefs and academic achievement: A meta-analytic review. *Educational Psychologist*, *39*(2), 111–133. https://doi.org/10.1207/s15326985ep3902_3

Vandenberg, R. J., & Lance, C. E. (2000). A review and synthesis of the measurement invariance literature: Suggestions, practices, and recommendations for organizational research. *Organizational Research Methods*, *3*(1), 4–70.

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.

Vermunt, J. K. (2010). Latent class modeling with covariates: Two improved three-step approaches. *Political Analysis*, *18*(04), 450–469. https://doi.org/10.1093/pan/mpq025

Villalón, R., Mateos, M., & Cuevas, I. (2015). High school boys' and girls' writing conceptions and writing self-efficacy beliefs: what is their role in writing performance? *Educational Psychology*, *35*(6), 653–674. https://doi.org/10.1080/01443410.2013.836157

White, K. R. (1982). The relation between socioeconomic status and academic achievement. *Psychological B*, *91*(3), 461–481.

Woodrow, L. (2011). College English writing affect: Self-efficacy and anxiety. *System*, *39*(4), 510–522.

Yates, A. (1987). *Multivariate exploratory data analysis: A perspective on exploratory factor analysis*. Albany: State University of New York Press.

Yilmaz Soylu, M., 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*. https://doi.org/10.3389/fpsyg.2017.01406

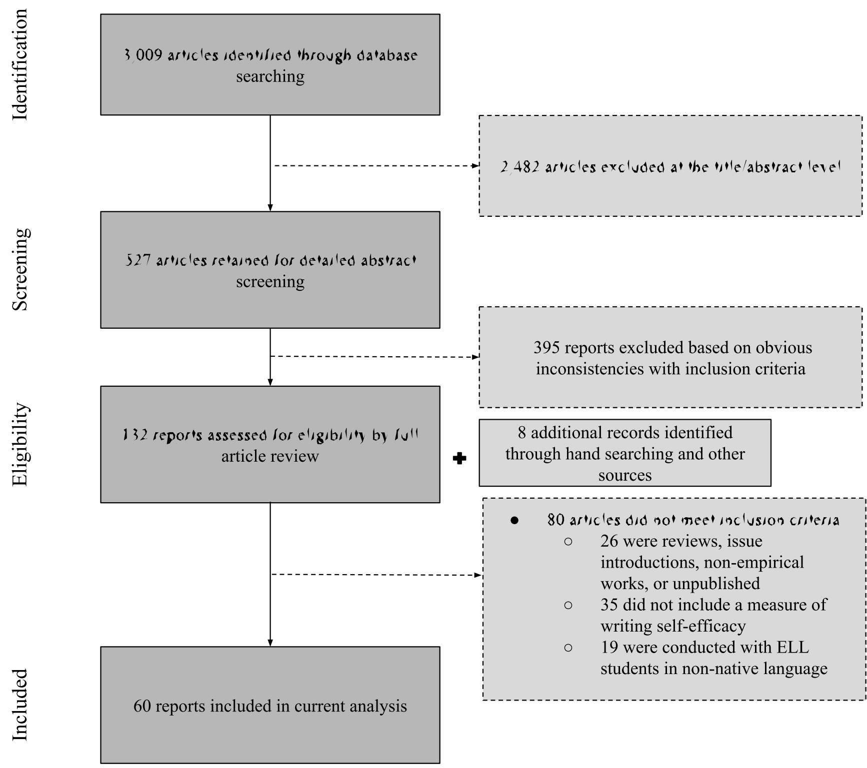
Zhang, Z., & Yuan, K.-H. (2016). Robust Coefficients Alpha and Omega and Confidence Intervals With Outlying Observations and Missing Data: Methods and Software. *Educational and Psychological Measurement*, *76*(3), 387–411. https://doi.org/10.1177/0013164415594658

Zimmerman, B. J., & Bandura, A. (1994). Impact of self-regulatory influences on writing course attainment. *American Educational Research Journal*, *31*(4), 845. https://doi.org/10.2307/1163397

Zumbrunn, S., Broda, M., Variar, D., & Conklin, S. (2019). *Examining the multidimensional role of self-efficacy for writing on student writing self-regulation and grades in elementary and high school.* Manuscript submitted for publication.

Zumbrunn, S., Marrs, S., & Mewborn, C. (2016). Toward a better understanding of student perceptions of writing feedback: a mixed methods study. *Reading and Writing*, *29*(2), 349–370. https://doi.org/10.1007/s11145-015-9599-3

Appendix A - PRISMA Flow Diagram



Appendix B - Systematic Literature Review Descriptive Statistics

Descriptive Statistics of Writing Self-Efficacy Studies 2008-2018

Design/Method	Frequency	% of All Studies
Quantitative	52	86.7%
Mixed Methods	8	13.3%
Longitudinal	16	26.7%
Experimental	17	28.3%
Sample Descriptives		
Learning Disabled Students	7	11.7%
International	28	46.7%
School Level		
Adult	2	3.3%
Graduate	1	1.7%
Undergraduates	27	45.0%
High-School	1	1.7%
Middle-School	7	11.7%
Elementary	10	16.7%
MS & HS	6	10.0%
Elem & MS	5	8.3%
Sample size*		
< 200	34	56.7%
200-500	12	20.0%
500-1000	11	18.3%
> 1000	2	3.3%

Note. One article did not report a sample size.

Appendix C - Measures

Adapted Self-Efficacy for Writing Scale (SEWS; Ekholm et al., 2015; Zumbrunn et al., 2016)

We now would like you to think about writing in your English/Language Arts Class. For each statement, please choose the world that best describes you.

1. I can write complete sentences.
2. I can think of many words to describe my ideas.
3. I can punctuate my sentences correctly.
4. I can concentrate on my writing for a long time.
5. I can spell my words correctly.
6. I can think of many ideas for my writing.
7. I can put my ideas into writing.
8. I can avoid distraction when I write.
9. I can keep writing even when it is difficult.

Response Scale: 4 point scale: 1 = *Almost never*, 2 = *Sometimes*, 3 = *Often*, 4 = *Almost Always*

Writing Self-Efficacy Scale (Pajares, 2007)

1. Correctly spell all words in a one page story or composition.
2. Correctly punctuate a one page story of composition.
3. Correctly use all parts of speech in a written composition.
4. Writing simple sentences with good grammar.
5. Correctly use singulars and plurals, verb tenses, prefixes, and suffixes.
6. Writing a strong paragraph that has a good topic sentence or main idea.
7. Structure paragraphs to support ideas in the topic sentences.
8. End paragraphs with proper conclusions.
9. Write a well-organized and well-sequenced paper that has a good introduction, body, and conclusion.
10. Get ideas across in a clear manner by staying focused without getting off topic.

Response Scale: 0 (*No Confidence At All*) -100 (*Completely Confident*)

Writing Apprehension Scale (WAS-12; Limpo, 2018)

I look forward to writing down my ideas.
I would enjoy submitting my writing to magazines for evaluation and publication.
I like to write down my ideas.
I enjoy writing.
Writing is a lot of fun.
I like seeing my thoughts on paper.

I'm nervous about writing.
I expect to do poorly in composition classes even before I enter them.
When I hand in a composition, I know I'm going to do poorly.
It's easy for me to write good compositions.
I don't think I write as well as most other people.
I'm not good at writing.

Response Scale: 1 (*completely disagree*) to 5 (*completely agree*).

Appendix D - Standardized Writing Test Blueprints

Due to the size of these files, hyperlinks are provided below. They may either be completely pasted into a web-browser or clicked on. Each test blueprint can also be sent electronically upon request.

Grade 8:
http://www.doe.virginia.gov/testing/sol/blueprints/english_blueprints/2010/2010_blueprint_gr8_writing.pdf

Grade 9: Blueprint was not available.

Grade 10:
http://www.doe.virginia.gov/testing/sol/blueprints/english_blueprints/2010/2010_blueprint_eoc_writing.pdf

Appendix E - Performance Based Assessment

Virginia Quality Criteria Review Tool for Performance Assessments
Revised: January 18, 2018

This document details a set of criteria for the development of performance assessments that measure the application of content knowledge and skills. The criteria are designed to support comparability in rigor and quality across the state.

Criterion 1: Standards/Intended Learning Outcomes

The rubric for the quality rating is as follows: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence.

#	Description	Quality Rating	Evidence or Rationale
1A	Virginia Standards of Learning selected for the performance assessment are clearly listed in a task template, developmentally appropriate for target students, and aligned to the grade-level scope and sequence or grade-level curriculum. Performance assessment components, resources/materials, and student products are aligned to the listed SOLs.		
1B	The performance assessment goes beyond simple recall, elicits evidence of complex student thinking, and requires application of disciplinary or cross-disciplinary concepts, practices, and/or transferable skills, such as application, analysis, evaluation, synthesis, or original creation.		
1C	The performance assessment provides an opportunity for students to develop and demonstrate (even if not explicitly assessed): • Deeper learning competencies, defined as mastering rigorous academic content; learning how to think critically and solve problems; working collaboratively; communicating effectively; directing one's own learning; and developing an academic mindset.		

#	Description	Quality Rating	Evidence or Rationale
	The performance assessment may also provide opportunities for students to develop and demonstrate: • Life-Ready competencies defined by the Profile of a Virginia Graduate as content knowledge, career planning, workplace skills, and community and civic responsibility; • Technology-related competencies; • Integration of intended learning outcomes from two or more subjects.		

Criterion 2: Authenticity

The rubric for the quality rating is as follows: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence.

#	Description	Quality Rating	Evidence or Rationale
2	The performance assessment is authentic along the dimensions: • The performance assessment's topic, context (scenario), materials/resources, products, and purpose/audience (i.e., what students are asked to do and for whom) are relevant to the real-world, students' community, students' interests, future careers, or other meaningful context. • The performance assessment asks students to do work authentic to the discipline (i.e., what adult practitioners of the discipline do), such as science inquiry; math problem-solving; analyzing and critiquing a text; analyzing and evaluating historical sources.		

Criterion 3: Language Use for Expressing Reasoning

The rubric for the quality rating is as follows: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence.

#	Description	Quality Rating	Evidence or Rationale
3A	The performance assessment supports language use and development by providing multiple means of accessing and using developmentally appropriate academic and disciplinary language for the students to express their reasoning.		
3B	The performance assessment should require students to use one or more forms of language to communicate their reasoning. The performance assessment may provide access to functional, academic, and disciplinary language in various forms of language media (text, video, audio, oral) OR provide opportunity to practice the use of language through multiple means of expression and language production (text, language media production, oral language, or conversation with peers).		

Criterion 4: Success Criteria for Students

The rubric for the quality rating is as follows: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence.

#	Description	Quality Rating	Evidence or Rationale
4A	The performance assessment includes a rubric or other appropriate scoring tools (e.g., checklist, analytic rubric) with scoring dimensions that are tightly aligned to performance expectations of the intended learning outcomes targeted within the performance assessment. Criteria should include language objectives, if applicable.		
4B	The scoring tool is written clearly and concisely, with audience-friendly language, as appropriate. Language of the scoring tool should describe how a response demonstrates performance expectations so that the tool		

#	Description	Quality Rating	Evidence or Rationale
	may be used to provide feedback to students about their work and how it can be improved.		
4C	The scoring tool or feedback methodology should be used across performance assessments within the course so that results on the performance assessment can be used to communicate a consistent set of expectations to students, monitor students' academic growth over time, inform instructional decisions, and communicate student proficiency to others (e.g., parents/guardians).		

Criterion 5: Student Directions, Prompt, and Resources/Materials

The rubric for the quality rating is as follows: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence.

#	Description	Quality Rating	Evidence or Rationale
5A	The student-facing task prompt, directions, and resources/materials are aligned to the intended learning outcomes, task purpose, and the performance expectations being assessed (i.e., the student product will provide evidence of the performance expectations).		
5B	The student-facing task prompt, directions, and resources/materials are clear, complete, written in accessible language appropriate to the grade level, and organized for students in an accessible format.		
5C	The task prompt/directions, topic, context (scenario), and materials/resources are sensitive to the community and free of bias.		

Criterion 6: Accessibility

The rubric for the quality rating is as follows: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence.

#	Description	Quality Rating	Evidence or Rationale
6A	The performance assessment is designed to accommodate the participation of all students. Directions for teachers for the performance assessment identify appropriate supports or alternatives to facilitate accessibility while maintaining the validity and reliability of the assessment.		
6B	The performance assessment is accessible and allows for differentiating the ways that students demonstrate their knowledge such as through the application of principles of Universal Design for Learning (UDL). Refer to the National Center on UDL at the Center for Applied Special Technology (CAST).		

Criterion 7: Feasibility

The rubric for the quality rating is as follows: 0-No Evidence; 1-Limited Evidence; 2-Partial Evidence; 3-Full Evidence.

#	Description	Quality Rating	Evidence or Rationale
7A	Student-facing prompts, directions, resources/materials, and scoring tools are included. Resources and materials required by the performance assessment are realistic and easily accessible to teachers.		
7B	Duration of implementation of the performance assessment is indicated and is realistic for the complexity of the assessment and the scope of performance expectations being assessed.		
7C	If the performance assessment is implemented over multiple lessons, a schedule indicating how the performance assessment is implemented across the lessons is included. Information about students' prior learning and how the performance assessment fits within a learning sequence is included.		

Appendix F - Correlation Matrix for All Variables

(Variances are on the diagonal)

	FEMALE	ETHNIC	ELL	GIFTED	GRADE	SE1	SE2	SE3	SE4	SE5	SE6
FEMALE											
ETHNIC	-0.055	1.364									
ELL	0.021	-0.273									
GIFTED	-0.098	0.152	-0.236								
GRADE	0.002	-0.076	0.093	-0.073							
SE1	0.122	0.07	-0.407	0.436	0.078						
SE2	-0.024	0.09	-0.37	0.317	0.004	0.579					
SE3	0.109	0.065	-0.294	0.302	0.012	0.592	0.399				
SE4	0.057	0.029	-0.118	0.067	-0.006	0.392	0.457	0.281			
SE5	0.03	0.032	-0.264	0.324	0.029	0.413	0.346	0.437	0.22		
SE6	-0.011	0.041	-0.232	0.09	-0.05	0.331	0.576	0.279	0.55	0.161	
SE7	0.049	0.001	-0.249	0.145	-0.093	0.468	0.601	0.371	0.563	0.235	0.709
SE8	0.015	-0.023	-0.023	0.054	-0.031	0.307	0.39	0.262	0.641	0.233	0.445
SE9	-0.044	0.026	-0.207	0.147	-0.061	0.37	0.475	0.303	0.603	0.211	0.547
WSES1	0.01	0.053	-0.231	0.326	0.04	0.428	0.357	0.437	0.275	0.69	0.207
WSES5	0.013	0.045	-0.309	0.434	-0.011	0.527	0.426	0.523	0.37	0.487	0.33
WSES8	0.035	0.062	-0.219	0.177	-0.027	0.469	0.43	0.4	0.42	0.331	0.413
WSES2	0.075	0.089	-0.241	0.328	-0.021	0.478	0.372	0.639	0.319	0.394	0.303
WSES6	0.035	0.072	-0.262	0.23	-0.021	0.48	0.492	0.371	0.467	0.305	0.495
WSES3	0.042	0.072	-0.273	0.285	-0.042	0.496	0.467	0.481	0.416	0.424	0.387
WSES10	-0.001	0.034	-0.235	0.176	-0.006	0.46	0.471	0.366	0.507	0.345	0.455
WSES7	0.043	0.059	-0.27	0.273	0.002	0.497	0.468	0.404	0.459	0.374	0.428
WSES4	0.124	0.073	-0.304	0.328	0.025	0.537	0.408	0.475	0.323	0.468	0.295
WSES9	0.046	0.057	-0.292	0.26	-0.011	0.509	0.477	0.422	0.489	0.365	0.46
WAS1	0.179	0.013	-0.056	-0.07	-0.01	0.209	0.275	0.161	0.391	0.102	0.445
WAS3	0.267	-0.008	-0.032	-0.041	0.009	0.207	0.244	0.164	0.35	0.083	0.383
WAS5	0.189	0.073	-0.154	0.047	-0.052	0.215	0.287	0.169	0.433	0.104	0.456
WAS11	0.078	-0.038	0.206	-0.155	0.002	-0.194	-0.347	-0.213	-0.237	-0.138	-0.294
WAS6	0.21	-0.011	-0.134	-0.023	0.03	0.196	0.273	0.178	0.363	0.128	0.363
WAS7	0.139	-0.04	0.206	-0.163	0.029	-0.188	-0.274	-0.168	-0.201	-0.115	-0.23
WAS8	-0.046	-0.033	0.185	-0.133	-0.031	-0.253	-0.269	-0.211	-0.245	-0.178	-0.26
WAS4	0.218	0.112	-0.167	0.037	-0.025	0.233	0.285	0.189	0.409	0.075	0.458
WAS9	-0.026	-0.035	0.179	-0.131	-0.059	-0.243	-0.291	-0.23	-0.263	-0.18	-0.279
WAS2	0.113	0.028	-0.054	0.017	-0.062	0.059	0.195	0.071	0.267	0.015	0.302
WAS10	0.009	0.084	-0.19	0.153	-0.019	0.349	0.457	0.273	0.454	0.2	0.527
WAS12	-0.045	-0.061	0.154	-0.109	-0.017	-0.269	-0.343	-0.228	-0.363	-0.168	-0.405

SE7	SE8	SE9	WSES1	WSES5	WSES8	WSES2	WSES6	WSES3	WSES10	WSES7	WSES4
0.489											
0.58	0.612										
0.302	0.27	0.279	5.532								
0.386	0.315	0.356	0.632	6.15							
0.45	0.372	0.414	0.46	0.55	5.328						
0.347	0.28	0.339	0.574	0.651	0.589	5.427					
0.535	0.399	0.462	0.474	0.529	0.645	0.55	5.024				
0.454	0.378	0.424	0.558	0.695	0.645	0.633	0.675	5.151			
0.52	0.482	0.494	0.478	0.548	0.602	0.54	0.672	0.635	5.521		
0.507	0.422	0.455	0.507	0.601	0.667	0.574	0.732	0.69	0.745	4.795	
0.358	0.28	0.294	0.603	0.595	0.498	0.621	0.578	0.604	0.523	0.61	4.715
0.524	0.41	0.465	0.5	0.591	0.675	0.603	0.757	0.683	0.719	0.758	0.635
0.379	0.299	0.301	0.203	0.21	0.283	0.223	0.363	0.268	0.323	0.314	0.25
0.381	0.283	0.29	0.172	0.186	0.275	0.22	0.338	0.264	0.31	0.3	0.226
0.427	0.301	0.364	0.18	0.23	0.23	0.221	0.314	0.287	0.319	0.301	0.217
-0.314	-0.22	-0.241	-0.161	-0.2	-0.237	-0.188	-0.309	-0.224	-0.255	-0.271	-0.183
0.372	0.316	0.311	0.212	0.218	0.292	0.25	0.359	0.281	0.367	0.344	0.25
-0.274	-0.204	-0.222	-0.134	-0.153	-0.187	-0.145	-0.297	-0.203	-0.226	-0.218	-0.175
-0.284	-0.244	-0.224	-0.209	-0.238	-0.279	-0.25	-0.352	-0.251	-0.297	-0.311	-0.281
0.405	0.282	0.337	0.149	0.207	0.235	0.226	0.311	0.273	0.303	0.302	0.231
-0.333	-0.233	-0.26	-0.238	-0.236	-0.299	-0.254	-0.372	-0.273	-0.295	-0.32	-0.285
0.288	0.2	0.245	0.093	0.103	0.163	0.113	0.235	0.174	0.219	0.185	0.077
0.523	0.383	0.458	0.313	0.381	0.409	0.375	0.525	0.427	0.48	0.489	0.358
-0.427	-0.285	-0.323	-0.189	-0.263	-0.31	-0.26	-0.409	-0.322	-0.362	-0.396	-0.296

WSES 9	WAS 1	WAS 3	WAS 5	WAS1 1	WAS 6	WAS 7	WAS 8	WAS 4	WAS 9	WAS 2	WAS1 0	WAS1 2
4.919												
0.335	1.485											
0.302	0.754	1.623										
0.328	0.601	0.546	1.794									
-0.307	-0.056	-0.057	-0.163	1.662								
0.352	0.607	0.635	0.541	-0.075	1.435							
-0.285	-0.012	0.028	-0.059	0.518	0.025	1.759						
-0.349	-0.118	-0.158	-0.14	0.443	-0.131	0.513	1.501					
0.329	0.566	0.55	0.842	-0.173	0.531	-0.063	-0.171	1.871				
-0.359	-0.111	-0.141	-0.097	0.471	-0.111	0.495	0.696	-0.134	1.196			
0.238	0.39	0.363	0.478	-0.14	0.357	-0.069	-0.101	0.472	-0.058	1.638		
0.537	0.373	0.349	0.449	-0.355	0.367	-0.314	-0.309	0.454	-0.324	0.445	1.145	
-0.419	-0.253	-0.244	-0.391	0.549	-0.241	0.443	0.533	-0.411	0.56	-0.208	-0.463	1.621

Appendix G – Online Supplemental Link

All Mplus input and output files, R code, and extra visualizations can be found on the Github repository established for this dissertation.

You may find this repository here: https://github.com/debusklaneml/hatch

You may also find enumeration visualizations here: https://debusklaneml.github.io

Appendix H – LPA Enumeration (CFA/bESEM) and Split-Sample Cross-Validation

LPA Enumeration Fit Indices for CFA and bESEM Calibration Data

Model	*N*	Parameters	Loglikelihood	cf	AIC	CAIC	BIC	aBIC	Entropy	aLMR	aLMR *p-value*
1-Calibrate CFA LPA	734	6	-2309.580	0.934	4631.161	4631.277	4658.752	4639.700			
2-Calibrate CFA LPA	734	10	-1869.684	1.333	3759.367	3759.671	3805.353	3773.599	0.796	847.677	0.000
3-Calibrate CFA LPA	734	14	-1621.817	1.366	3271.634	3272.218	3336.013	3291.558	0.852	477.637	0.000
4-Calibrate CFA LPA	**734**	**18**	**-1493.890**	**1.374**	**3023.780**	**3024.737**	**3106.553**	**3049.397**	**0.851**	**246.515**	**0.012**
5-Calibrate CFA LPA	734	22	-1430.506	2.683	2905.011	2906.434	3006.178	2936.321	0.866	122.141	0.744
6-Calibrate CFA LPA	734	26	-1374.602	2.545	2801.205	2803.191	2920.766	2838.207	0.870	107.725	0.512
7-Calibrate CFA LPA	734	30	-1332.163	1.131	2724.326	2726.972	2862.281	2767.021	0.869	81.781	0.016
4-Validate (fixed) CFA LPA	732	0	-1458.138		2916.276	2916.276	2916.276	2916.276	0.848		
4-Validate (free) CFA LPA	732	18	-1442.008	1.468	2920.015	2920.974	3002.739	2945.583	0.855		
4-Calibrate (fixed) CFA LPA	734	0	-1511.973		3023.945	3023.945	3023.945	3023.945	0.854		
1-Calibrate ESEM LPA	734	6	-2789.029	0.937	5590.058	5590.174	5617.649	5598.597			
2-Calibrate ESEM LPA	734	10	-2455.822	1.288	4931.644	4931.948	4977.629	4945.875	0.744	642.087	0.000
3-Calibrate ESEM LPA	734	14	-2273.916	1.317	4575.832	4576.416	4640.211	4595.757	0.821	350.531	0.001
4-Calibrate ESEM LPA	**734**	**18**	**-2186.542**	**1.208**	**4409.084**	**4410.041**	**4491.857**	**4434.701**	**0.815**	**168.369**	**0.004**
5-Calibrate ESEM LPA	734	22	-2150.404	1.606	4344.808	4346.231	4445.975	4376.118	0.810	69.637	0.502
6-Calibrate ESEM LPA	734	26	-2122.752	1.091	4297.503	4299.489	4417.065	4334.506	0.836	53.286	0.001
7-Calibrate ESEM LPA	734	30	-2107.389	1.064	4274.777	4277.423	4412.733	4317.473	0.819	29.604	0.014
4-Validative (fixed) ESEM LPA	732	0	-2204.554		4409.107	4409.107	4409.107	4409.107	0.804		
4-Validative (free) ESEM LPA	732	18	-2184.242	1.241	4404.484	4405.443	4487.208	4430.052	0.784		
4-Calibrate (fixed) ESEM LPA	734	0	-2206.448		4412.896	4412.896	4412.896	4412.896	0.785		
1-Calibrate Alt bESEM LPA	734	8	-3037.983	0.981	6091.966	6092.165	6128.754	6103.352			
2-Calibrate Alt bESEM LPA	734	13	-2996.172	1.056	6018.345	6018.851	6078.125	6036.846	0.483	81.161	0.000
3-Calibrate Alt bESEM LPA	**734**	**18**	**-2938.864**	**1.226**	**5913.728**	**5914.685**	**5996.501**	**5939.345**	**0.643**	**111.245**	**0.006**
4-Calibrate Alt bESEM LPA	734	23	-2916.516	1.434	5879.032	5880.587	5984.798	5911.765	0.624	43.381	0.433

5-Calibrate Alt bESEM LPA	734	28	-2891.042	1.197	5838.084	5840.388	5966.842	5877.933	0.715	49.450	0.042
6-Calibrate Alt bESEM LPA	734	33	-2859.280	1.180	5784.561	5787.767	5936.312	5831.526	0.777	61.654	0.043
7-Calibrate Alt bESEM LPA	734	38	-2807.559	1.293	5691.118	5695.383	5865.862	5745.199	0.840	47.243	0.106
3-Validate Alt (free) bESEM LPA	732	18	-2918.269	1.103	5872.538	5873.497	5955.262	5898.106	0.575		
3-Validate Alt (fixed) bESEM LPA	732	0	-2945.291		5890.581	5890.581	5890.581	5890.581	0.639		
3-Calibrate Alt (fixed) bESEM LPA	734	0	-2962.970		5925.939	5925.939	5925.939	5925.939	0.553		
4-Validate Alt (fixed) bESEM LPA	732	0	-2922.352		5844.704	5844.704	5844.704	5844.704	0.616		
4-Validate Alt (free) bESEM LPA	732	23	-2890.026	1.095	5826.052	5827.611322	5931.755	5858.722	0.626		
4-Calibrate Alt (fixed) bESEM LPA	734	0	-2947.739		5895.478	5895.478	5895.478	5895.478	0.610		
5-Validate Alt (free) bESEM LPA	732	28	-2854.074	1.162	5764.147	5766.457	5892.829	5803.920	0.744		
5-Validate Alt (fixed) bESEM LPA	732	0	-2912.297		5824.594	5824.594	5824.594	5824.594	0.711		
5-Calibrate Alt (fixed) bESEM LPA	734	0	-2941.111		5882.223	5882.223	5882.223	5882.223	0.724		
6-Validate Alt (fixed) bESEM LPA	732	0	-2884.852		5769.705	5769.705	5769.705	5769.705	0.778		
6-Validate Alt (free) bESEM LPA	732	33	-2794.400	1.250	5654.800	5658.015	5806.461	5701.675	0.845		
6-Calibrate Alt (fixed) bESEM LPA	734	0	-2869.077		5738.155	5738.155	5738.155	5738.155	0.839		
bESEM Multi-Group Configural	1466	37	-6873.285	1.305	13820.571	13822.540	14016.311	13898.774	0.761		
bESEM Multi-Group Structural	1466	25	-6885.670	1.466	13821.341	13822.244	13953.598	13874.181	0.742		
bESEM Multi-Group Dispersional	1466	21	-6885.998	1.579	13813.995	13814.635	13925.091	13858.381	0.743		
bESEM Multi-Group Distributional	1466	19	-6887.129	1.601	13812.259	13812.785	13912.774	13852.417	0.744		
Final bESEM LPA - Full Sample	**1466**	**18**	**-5870.977**	**1.649**	**11777.954**	**11778.43**	**11873.179**	**11815.999**	**0.583**		

Appendix I – Chi-square Loglikelihood Ratio Tests

CFA								
4-Profile Validation Fixed w/ Calibration svalues Compared to Validative Freely Estimated								
L0	L1	c0	c1	p0	p1	cd	TRd	*p*-value
-1458.1380	-1442.0080	0.0000	1.4675	0.0000	18.0000	1.4675	21.9830	0.2327
4-Profile Calibration Fixed w/ Validative svalues Compard to Calibrated Freely Estimated								
L0	L1	c0	c1	p0	p1	cd	TRd	*p*-value
-1511.9730	-1493.8900	0.0000	1.3737	0.0000	18.0000	1.3737	26.3274	0.0925
ESEM								
4-Profile Validation Fixed w/ Calibration svalues Compared to Validative Freely Estimated								
L0	L1	c0	c1	p0	p1	cd	TRd	*p*-value
-2204.5540	-2184.2420	0.0000	1.2407	0.0000	18.0000	1.2407	32.7428	0.0179
4-Profile Calibration Fixed w/ Validative svalues Compard to Calibrated Freely Estimated								
L0	L1	c0	c1	p0	p1	cd	TRd	*p*-value
-2206.4480	-2186.5420	0.0000	1.2083	0.0000	18.0000	1.2083	32.9488	0.0169
Bifactor ESEM								
3-Profile Validation Fixed w/ Calibration svalues Compared to Validative Freely Estimated								
L0	L1	c0	c1	p0	p1	cd	TRd	*p*-value
-2945.2910	-2918.2690	0.0000	1.1028	0.0000	18.0000	1.1028	49.0062	0.0001
3-Profile Calibration Fixed w/ Validative svalues Compard to Calibrated Freely Estimated								
L0	L1	c0	c1	p0	p1	cd	TRd	*p*-value
-2962.9700	-2938.8640	0.0000	1.2262	0.0000	18.0000	1.2262	39.3182	0.0026
4-Profile Validation Fixed w/ Calibration svalues Compared to Validative Freely Estimated								
L0	L1	c0	c1	p0	p1	cd	TRd	*p*-value
-2922.3520	-2890.0260	0.0000	1.0951	0.0000	23.0000	1.0951	59.0375	0.0001
4-Profile Calibration Fixed w/ Validative svalues Compard to Calibrated Freely Estimated								
L0	L1	c0	c1	p0	p1	cd	TRd	*p*-value
-2947.7390	-2916.5160	0.0000	1.4335	0.0000	23.0000	1.4335	43.5619	0.0059
5-Profile Validation Fixed w/ Calibration svalues Compared to Validative Freely Estimated								
L0	L1	c0	c1	p0	p1	cd	TRd	*p*-value
-2947.7390	-2854.0740	0.0000	1.1624	0.0000	28.0000	1.1624	161.1579	0.0000
5-Profile Calibration Fixed w/ Validative svalues Compard to Calibrated Freely Estimated								
L0	L1	c0	c1	p0	p1	cd	TRd	*p*-value
-2941.1110	-2891.0420	0.0000	1.1967	0.0000	28.0000	1.1967	83.6784	0.0000
6-Profile Validation Fixed w/ Calibration svalues Compared to Validative Freely Estimated								
L0	L1	c0	c1	p0	p1	cd	TRd	*p*-value
-2884.8520	-2794.4000	0.0000	1.2496	0.0000	33.0000	1.2496	144.7695	0.0000
6-Profile Calibration Fixed w/ Validative svalues Compard to Calibrated Freely Estimated								
L0	L1	c0	c1	p0	p1	cd	TRd	*p*-value
-2869.0770	-2859.2800	0.0000	1.1796	0.0000	33.0000	1.1796	16.6107	0.9922

Note. L0 and L1 = Logliklihood values; c0 and c1 = MLR scaling correction factor; p0 and p1 = Parameters; cd = scaling correction; TRd = Chi-square difference test