EXAMINING PRIMARY SCHOOL CONTENT AND LANGUAGE INTEGRATED

LEARNING WRITING THROUGH THE SIMPLE VIEW OF WRITING

A Dissertation

by

KEITH MATHEW GRAHAM

Submitted to the Office of Graduate and Professional Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Zohreh Eslami
Committee Members,	Poh Wee Koh
	Sharon Matthews
	Wen Luo
Head of Department,	Michael de Miranda

May 2020

Major Subject: Curriculum and Instruction

ABSTRACT

This dissertation uses the Simple View of Writing to investigate Content and

Language Integrated Learning (CLIL) writing development through three research studies:

one systematic review and two empirical studies. The introduction to this dissertation

provides an overview of conceptualizations of writing, CLIL, and Taiwan's recent

proposal for developing into a bilingual nation through CLIL instruction.

The introduction is followed by the first study which examines previous research

on CLIL writing measurement and outcomes through a systematic search of the literature.

The review shows that research on CLIL writing can be divided into two categories,

research on CLIL versus non-CLIL writing outcomes and research on CLIL growth.

Within these studies, researchers use four categories of writing measurement: vocabulary

measures, rubrics, complexity/accuracy/fluency (CAF) measures, and content analysis. The

study highlights the use of singular metrics to examine writing rather than examinations

through models that account for the dimensions of the Simple View of Writing.

The second study investigates CLIL writing as a product by testing the fit of a

higher-order model based on the Simple View of Writing with narrative and expository

writing data collected from primary school CLIL students in Taiwan. Results show that a

model accounting for transcription and ideation exhibits a better fit with the data than a

single-factor model, suggesting that the Simple View of Writing may adequately explain

writing as a product for CLIL primary students.

The final study measures the effect of gender, spelling, vocabulary, and oral

language skills on writing for primary sixth grade CLIL students in Taiwan. The study

examines CLIL writing as a cognitive process driven by transcription and ideation component skills affected by gender. Using two hierarchical regression models for narrative and expository writing, results show the possibility of transcription and ideation component skills as predictors of writing. Additionally, it was found that gender may also be a predictor of writing outcomes when accounting for ideation ability. However, additional research is needed to better understand the contributions of transcription, ideation, and gender for predicting writing.

The dissertation concludes with recommendations for CLIL researchers and practitioners. Specifically, it encourages the use of the Simple View of Writing as a framework for investigating CLIL writing and planning curriculum and instruction in CLIL classrooms both in Taiwan and around the world.

ACKNOWLEDGEMENTS

This dissertation was made possible with the love and support of so many wonderful people in my life. I would like to begin by thanking my committee. I thank my chair, Dr. Eslami. You have inspired me, challenged me, and given me opportunities beyond what any Ph.D. student could hope for. Next, I would like to thank Dr. Matthews who was always there to cheer me on and encourage me, regardless of the circumstances. I would also like to thank Dr. Koh who provided feedback that lifted my research to new levels. Finally, I would like to thank Dr. Luo who inspired my love of educational statistics.

I would also like to thank other faculty members who, though not on my committee, have inspired and supported my work. I thank Dr. Dixon and Dr. Kuo for allowing me opportunities to learn and grow through the Ready-Set-Write project, from which much of my interest in writing development has been inspired. Additionally, I would like to thank Dr. Fowler, Dr. Richardson, and all of the wonderful people at the Center for Teaching Excellence for encouraging me, inspiring me, and making me feel part of a family at Texas A&M University. Finally, I would like to thank Dr. Viruru who contributed greatly to my development as a researcher and thinker and whose continued support helped me finish this doctoral journey.

Thanks also go to my wonderful friends at Texas A&M University. Thank you to Stephanie Moody for the daily text messages and support sessions throughout this program. I would also like to thank Haemin Kim who was always there to lend a helping

hand and support me in all my academic endeavors. Thank you also goes to Bizhu He and Yunkyeong Choi who have been wonderful collaborators and friends.

I also send a thanks to my Aggie Research Scholars undergraduate research team: Lauren Boraud, Dorothy McIntush, Macy Land, Omar Manzur, Mariana Melo, Ashok Meyyappan, Skylar Nelson, and Victor Villasana. You are the research team every Ph.D. student wishes they had. Thank you for your hard work, your support, and your eagerness to help me do great research.

Finally, I would like to conclude with a special thank you to my family. I thank my mother and father whose love, sacrifice, guidance, and support are the only reason I am the person I am today. I would also like to thank my wife Michelle for her patience, love, and willingness to travel to the other side of the world so I could pursue my American dream.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

Funding Sources

TABLE OF CONTENTS

LIST OF FIGURES

Page

x

LIST OF TABLES

1. INTRODUCTION

This dissertation examines English writing development in a Content and Language Integrated Learning (CLIL) program in Taiwan. With the release of the Taiwanese government's Blueprint for Developing Taiwan into a Bilingual Nation by 2030 in late 2018 (National Development Council, 2018), investigations of the intersection of English writing, CLIL, and Taiwan are of the upmost importance for Taiwanese students, teachers, citizens, and government officials. Many of these stakeholders are asking questions about what it means for Taiwan to be a bilingual nation of English and Chinese, how can this be assessed in terms of bilingual language skills, and how do the principles of CLIL help advance the country toward this goal.

In this introductory section, the intersection of writing, CLIL, and Taiwan will be explored. In the following sections, writing will first be defined and prominent theories discussed. Next, CLIL as an approach to language and content will be addressed along with how CLIL may support writing development. Following this, the Blueprint for Developing Taiwan into a Bilingual Nation by 2030 will be explored in detail, particularly as it pertains to its goals and mandates in regard to writing and CLIL. This introductory section will then conclude with a brief overview of the subsequent sections in this dissertation that look to add to the scholarship on CLIL writing development.

1.1. Writing as a Product and Process

Writing is a complex construct, comprised of various dimensions, which takes on many forms. The word *writing* can be conceptualized as both a noun, where it is seen as a product, or as a verb, where it is a process (Hyland, 2008). For educational researchers and

practitioners, understanding both the product and the process of writing are important. When writing is thought of as a product, it can be examined at many levels—the word, clause, and discourse levels (Berninger & Swanson, 1994). The writing product can also be understood as comprised of language and ideas (Berninger & Swanson, 1994; Juel, Griffith, & Gough, 1986). When we consider writing as a process, we often think of broad writing stages of planning, translating, and reviewing (Berninger & Swanson, 1994; Berninger & Winn, 2006; Flower & Hayes, 1981). However, writing as a process could also be considered from a component perspective, where writing occurs through the interaction of various component skills (Berninger & Winn, 2006; Harrison et al., 2016; Kim, Al Otaiba, Wanzek, & Gatlin, 2015; Kim, Gatlin, Al Otaiba, & Wanzek, 2017). Whether considered as a product or process, the object or act of writing can often be governed by "how texts actually work as communication" (Hyland, 2008, p. 3), or what could be referred to as genre. In the following paragraphs, the concepts of writing as a product, writing as a process, and the effect of genre will be further detailed.

1.1.1. Writing as a Product

Writing can be conceptualized in many different ways. According to the Simple View of Writing, writing is comprised of two broad elements—transcription and ideation (Berninger & Swanson, 1994; Juel et al., 1986). In other words, when we consider writing as a product, both the language elements and the ideas presented must be considered. These elements could be considered at the word, clause, and/or discourse levels. Considering first the word level, a writing product is often defined by its lexicon, or vocabulary words, and spelling. The lexicon used are tied to ideas that are being expressed by the writer and spelling is the method for which a writer translates these ideas into the

2

written form. At this level, the success of a written product is determined by how well the lexicon used expresses the ideas of the writer, and the accuracy of spelling can determine the effective transmission of these ideas through writing.

At the second level, the clause level, ideas are now further explored through the combination of words together, and the rules of language determine the degree of clarity for which these ideas are expressed. The clause allows ideas to be expanded from singular objects or abstract entities to multi-faceted entities that take on characteristics and/or engage in actions, often with other singular ideas, in a dynamic textual world. On the language side, the organization of the multiple ideas is bound by rules of morphology and syntax. The use of morphemes to create the correct form of words (e.g., noun, adjective, etc.) and the ordering of words based on syntactic rules often regulate the success of a writing product's expression of ideas.

The final level of writing as a product is the discourse level. At this level, the dynamic ideas of the previous level are combined with other dynamic ideas to form a communicative text. Ideas at this level join together in a complex web of language. Register, also known as voice, and organization within this complex language system become critical for communicative success. The accessibility of the written product is aided by cohesion between clauses and paragraphs through various language devices. This level is the final stage of writing as a product, the culmination of the previous word- and clause-level transcription and ideation. Together, the written product is realized.

1.1.2. Writing as a Process

Associated with the complexity of writing as a product comes an equally complex process of producing writing. Writing as a process can be examined from two

perspectives—a process model (Flower & Hayes, 1981) or a components model (Berninger & Winn, 2006; Harrison et al., 2016; Kim et al., 2017). Beginning with the process model, Flower and Hayes (1981) present writing as a process broadly through planning, translating, and reviewing. Planning is composed of three sub-processes: generating ideas, organizing, and goal setting. This is often referred to as pre-writing. After planning, the writer begins translating, or what might be referred to as drafting. Finally, the writer engages in evaluating and revising in the reviewing stage. Flower and Hayes (1981) suggest that each of these stages are governed by a monitor, "which determines when the writer moves from one process to the next" (p. 374).

The process model proposed by Flower and Hayes (1981) has been criticized for its limited applicability only to proficient, older writers. From these criticisms, writing as a process, governed by component skills, emerged (Berninger & Swanson, 1994). The Simple View of Writing, comprised of transcription and ideation skills, is one model proposed as the processes for developing writers (Berninger & Swanson, 1994; Juel, Griffith, & Gough, 1986). Berninger and Winn (2006) present three other models to explain the process of writing and writing development. The Learning Triangle Model shows how writing is developed through interactions between the learner, the teacher's instruction, and the instructional and curricula materials used. The second model is the Brain-Mind-Behavior Model, which documents how the brain is involved in writing through the external processes of the writer's behavior and the internal processes of the brain's neuroanatomy, or processing of the brain, and the writer's mind, which could be described as thought production. The final model is what has been termed the Not-So-Simple-View of Writing. This model combines the process of transcribing text with

4

executive functions similar to those expressed in the Flower and Hayes (1981) model described above, all governed by working memory. Writing as a process has also been influenced by research into how other linguistic component skills govern how writing is produced. Kim et al. (2017) have demonstrated the predictive ability of a components model of oral language, handwriting fluency, and spelling for writing quality, moderated by text writing fluency. This model suggests that these linguistic skills—handwriting, spelling, and oral language—are used in conjunction with each other and are moderated by text writing fluency skills to produce quality writing. Similarly, Harrison et al. (2016) found transcription ability, a composite of handwriting and spelling, were influential component skills in predicting writing achievement. Additionally, skills in vocabulary and syntactic awareness were also influential for writing achievement. Taken together, the process of writing is comprised of multiple processes emanating from instructional processes, neurological processes, and various linguistic and metacognitive processes.

1.1.3. The Influence of Genre on Writing as a Product and Process

As Hyland (2007) notes, "Writing varies with context and cannot be distilled down to a set of abstract cognitive or technical abilities" (p. 150). In other words, writing as a product or as a process cannot be defined in a general way that is applicable in all situations. Various social, academic, and professional communities engage in unique processes and produce different forms of written texts. These different types of writing are referred to as genres, or "abstract, socially recognized ways of using language" (Hyland, 2007, p. 149).

Genre has implications for the word, clause, and discourse levels of a written product as well as for the component skills of a writing process. The words a writer might

5

use in a science text differ from those in a narrative text. Even when the words used are the same, they may take on semantically different meanings. For example, referring to a "table" in a science text is often different than referring to a "table" within a narrative text. At the clause and discourse level, genres also take on very distinct characteristics. As Schleppegrell, Achugar, and Oteíza (2004) note, history texts are often constructed "through particular linguistic choices, and these choices are quite different from the language students use to talk about the events of everyday life" (p. 74).

In regard to the influence of genre on writing processes, the differences begin in the planning stage, where a scientist writing a lab report on an experiment engages in a different prewriting process than a historian writing about a past event. These differences continue in the translating and reviewing stages where the writers will conform to writing norms of their discourse community. Similar to the influence of genre on writing product, the requirements of various skills in a component skill view of the writing process would differ by genre, with writers needing different discipline-specific language and knowledge. In sum, while writing can be conceived as both a product and process, this conceptualization cannot necessarily be generalized and may be influenced by the context and discourse community.

1.2. Content and Language Integrated Learning

Having examined writing as both a complex product and process, we now turn to Content and Language Integrated Learning (CLIL), both what it is and its intersection with writing. Though there are "varied interpretations" of what CLIL means both in principle and in practice (Cenoz, Genesee, & Gorter, 2013, p. 244), most researchers tend to define CLIL as "a dual educational approach where equal attention is paid to both content and

language objectives" (Dafouz, Camacho, & Urquia, 2014, p. 224). Ball, Kelly, and Clegg (2015) have suggested that CLIL takes on many forms from a Hard CLIL which focuses mostly on content to a Soft CLIL which is more of a balance between content and language. However, Hard CLIL is likely closer to what many would now term as English Medium Instruction (EMI) (Macaro, 2018). The other side of the spectrum, where the focus is exclusively on language but within a general or specific content area, would most likely be referred to as English for Academic Purposes (EAP) or English for Specific Purposes (ESP) (Airey, 2016; Macaro, 2018). To clarify, in EAP/ESP, students are expected to already know the content and only focus on connecting foreign language terms to what they already know; in EMI, it is often expected that students already have a grasp of the foreign language used during instruction, and the focus is on learning new content. With CLIL, neither assumption is made, meaning language and content are learned simultaneously.

Though Dalton-Puffer (2011) opines, "CLIL classrooms share a great deal more with traditional language lessons than a partisan look would make one believe" (p. 195), there is a distinct difference. Hymes (2005) discusses the concept of a speech community, "defined as a community sharing rules for the conduct and interpretation of speech, and rules for the interpretation of at least one linguistic variety" (p. 6). When considering traditional language lessons, the focus is on learning the language of a general speech community, should such a thing even exist. In contrast, CLIL teaches the language of specific speech communities. Each content area, be it science, math, and so forth, has a speech community tied to it. Through the integrating of content and language, CLIL

students both learn the knowledge and the speech of a discipline-specific community through their content classes.

It is this distinction, speech communities, where the intersection of writing and CLIL meet. As discussed above, whether discussing writing as a product or writing as a process, genre influences the way ideation and transcription present themselves in writing. Genres are tied to speech communities, who decide the rules and interpretation of writing. Given the importance of genre in writing, the teaching and learning of writing must consider the speech communities from which this writing emerges. CLIL's focus on integrating content and language allows for the knowledge and language practices of a speech community to be taught and learned at one time. Thus, CLIL instruction is designed to promote the teaching and learning of the writing of science, the writing of social studies, the writing of math, and so forth, rather than simply English writing in a broad and general sense. In this way, CLIL as an instructional approach may inherently be better suited for both the learning of transcription (language) and ideation (content).

1.3. Blueprint for Developing Taiwan into a Bilingual Nation by 2030

On December 10, 2018, Taiwan's Executive Yuan released a document titled *Blueprint for Developing Taiwan into a Bilingual Nation by 2030* (National Development Council, 2018). The document outlines the government's plans for Taiwan to become a "bilingual nation in which equal importance [is] attached to Chinese and English" (National Development Council, 2018, p. 2). The rationale for calling for this equal balance of languages is likely driven by globalization forces. Statements such as "English is currently the most important common language for international communication" and "'English proficiency' has become an essential ability for opening the gateway to

globalization" suggest that Taiwan sees a need for bilingualism in order to be "more internationally competitive" (National Development Council, 2018, p. 1). Toward this goal, the document puts an emphasis on two main areas: (1) written documents in English throughout the government and (2) English proficiency of citizens, developed through CLIL. Each of these are examined further below.

1.3.1. Written Documents in English

Written government documents in English are a key component of this policy document. The plan pushes for "promoting all bilingual websites of central government agencies," and states that any agencies whose websites are not yet in both English and Chinese "should be fully bilingualized" (National Development Council, 2018, p. 9). As one would expect from a national government, the agencies in Taiwan span across all disciplines, each with their unique lexicon and writing style. The importance of being able to write in English for specific academic and professional purposes are shown to be fully realized here through this ambitious plan for bilingual websites in Chinese and English.

The need for written English within the blueprint does not stop with websites. The plan further calls for "bilingualizing the information of important government business, such as the bilingualization of policy statements, public announcements, news releases, study reports, statistical reports, statistical data inquiry systems, and statistical indicators…laws, regulations, and administration rules" (National Development Council, 2018, p. 11). Like the government websites, this list of documents could potentially span across a variety of disciplines. But beyond this, it should be noted that each of these documents are a genre, each requiring a specific style and form of writing.

From reading through the various written documents needed to fulfill Taiwan's bilingual goals, both on websites and in other documentation, it becomes evident that writing proficiency in general English will not suffice for achieving these goals. Those tasked with creating the English documents will need to have a grasp of disciplinary knowledge, the language used to express this disciplinary knowledge appropriately in English, and knowledge of the genre-specific style and form of the document. In other words, specialized knowledge in discipline-specific English writing is required.

1.3.2. Developing English Proficiency of Citizens

Perhaps in recognition that general English proficiency will not be adequate for rising to a globally-competitive status, the *Blueprint for Developing Taiwan into a Bilingual Nation by 2030* promotes professional and academic English proficiency as the goal. This is illustrated through a call for "the bilingualization of all kinds of licenses and credentials" (p. 9) and "cultivating English specialists" (National Development Council, 2018, p. 11). Additionally, the document calls for "bilingualizing the examination and certification of National Skill Test [sic]" (National Development Council, 2018, p. 11). From these moves, it seems that monolingualism in Chinese will no longer be enough to enter into the government workforce; bilingualism in Chinese and English will be required.

However, prior to the examination and certification of bilingual specialists, citizens will need a way to develop academic and professional bilingualism. Toward this end, the Taiwanese government is pushing for a change in English education policy away from traditional forms of English education focusing on general English toward "implementing in full scale the bilingualization of Taiwan's educational system, cultivating bilingual talents and international perspective" (National Development Council, 2018, p. 11). The

policy spans widely beginning in preschool through "developing feasible modes for integrating English into preschool caretaking activities" (National Development Council, 2018, p. 11) to high schools where there are plans of "building a comprehensively Englishized learning environment in science park experimental high schools" (National Development Council, 2018, p. 11). At all levels, the government is pushing a policy of "Teaching English in English" (National Development Council, 2018, p. 13), and specifically in regard to developing academic and professional English, "Promoting CLIL (Content and Language Integrated Learning) in designated primary and secondary learning domains or subjects, and promoting ESP (English for Specific Purposes) teaching and learning for vocational schools" (National Development Council, 2018, p. 13). It is clear that the goal of this educational policy is to cultivate not just general English abilities but academic and professional English proficiency starting from a young age throughout all stages of education.

The Blueprint for Developing Taiwan into a Bilingual Nation brings the intersection of English writing, CLIL, and Taiwan to the forefront. It describes ambitious goals of "enhancing the people's English…writing communication abilities" (National Development Council, 2018, p. 4). But these goals of enhancing writing abilities are not general; they are discipline specific. They call for an English proficiency that allows for English specialists to write on topics that span professions in various genres. To cultivate citizens capable of these highly professional and specialized writing tasks, Taiwan has proposed the use of CLIL as a form of English instruction. However, questions remain about the viability of CLIL to develop English writing (Graham, Choi, Davoodi, Razmeh,

& Dixon, 2018). It is both the intersection of English writing, CLIL, and Taiwan and the lack of research in CLIL writing that drives the research reported in this dissertation.

1.4. Overview of Studies

In the following three sections, three studies on writing and CLIL will be presented. Each of these studies are grounded in the Simple View of Writing, where writing is seen as both transcription (language) and ideation (content). This theory is used as a guide for interpretation and design as it seems to mirror the integration of language and content inherent in CLIL. By using the Simple View of Writing throughout, this dissertation will examine its usefulness and applicability to writing both as a product and as a process in CLIL.

Section 2 utilizes a systematic review strategy to examine the current research on CLIL writing in relation to the Simple View of Writing. The study looks at CLIL writing research in terms of writing measurement and CLIL writing outcomes. Specifically, current research is examined for how transcription and ideation in writing has been measured and the success of CLIL for developing student transcription and ideation abilities.

Section 3 examines writing as a product in CLIL. Using narrative and expository writing samples collected from students in primary grades two through six enrolled in a CLIL program in Taiwan, a confirmatory factor analysis model based on the Simple View of Writing will be tested. The model tested is a higher-order factor model, with CLIL narrative and expository writing as covarying second-order factors, each with two latent variables, transcription and ideation, as first-order factors. The first-order factors have two observed variables for transcription and three for ideation, respectively. This model will be

tested against a single-factor model with one latent variable for CLIL writing and five observed variables from word, clause, and discourse levels. The goal of this study is to examine whether a model based on the Simple View of Writing can adequately explain CLIL writing outcomes.

Section 4 uses the Simple View of Writing to investigate writing from a different angle—writing as a process. This study takes a component skills view of writing as a process and examines how skills of transcription (spelling) and ideation (vocabulary and oral language) act as predictors of writing in CLIL. Additionally, the role of gender as a predictor of writing in CLIL will be explored. Narrative and expository writing from grade six CLIL students in Taiwan will be predicted using a hierarchical regression model with gender added in step one, spelling (transcription) at step two, vocabulary (word-level ideation) at step three, and oral language (discourse-level ideation) at step four. By understanding how transcription and ideation skills are involved in the processes of writing, CLIL instructors may be able to take a more targeted approach to writing instruction.

Having presented a review of current literature on CLIL writing along with empirical studies on CLIL writing as a product and as a process, implications and future directions for the use of the Simple View of Writing as a theory for CLIL writing will be discussed in Section 5. Both pedagogical and research applications will be considered.

1.5. References

Airey, J. (2016). EAP, EMI, or CLIL. In K. Hyland & P. Shaw (Eds.), *The Routledge handbook of English for academic purposes*. New York: Routledge.

Ball, P., Kelly, K., & Clegg, J. (2015). *Putting CLIL into practice*. New York: Oxford

      University Press.

Berninger, V. W., & Swanson, H. L. (1994). Modifying Hayes and Flower's model of

      skilled writing to explain beginning and developing writing. *Advances in Cognition*

      and Educational Practice, 2, 57-81.

Berninger, V. W., & Winn, W. (2006). Implications of advancements in brain research and

      technology for writing development, writing instruction, and educational evolution.

      In S. Graham, J. Fitzgerald, & C. A. MacArthur (Eds.), *Handbook of writing*

      research (pp. 96-114). New York: The Guilford Press.

Cenoz, J., Genesee, F., & Gorter, D. (2013). Critical analysis of CLIL: Taking stock and

      looking forward. *Applied Linguistics, 35*(3), 243-262.

      https://doi.org/10.1093/applin/amt011

Dafouz, E., Camacho, M., & Urquia, E. (2014). 'Surely they can't do as well': A

      comparison of business students' academic performance in English-medium and

      Spanish-as-first-language-medium programmes. *Language and Education, 18*(3),

      223-236. https://doi.org/10.1080/09500782.2013.808661

Dalton-Puffer, C. (2011). Content-and-Language Integrated Learning: From practice to

      principles? *Annual Review of Applied Linguistics, 31*, 182-204.

      https://doi.org/10.1017/S0267190511000092

Flower, L., & Hayes, J. R. (1981). A cognitive process theory of writing. *College*

      Composition and Communication, 32(4), 365-387. https://doi.org/10.2307/356600

Graham, K. M., Choi, Y., Davoodi, A., Razmeh, S., & Dixon, L. Q. (2018). Language and

      content outcomes of CLIL and EMI: A systematic review. *Latin American Journal*

of Content & Language Integrated Learning, 11(1), 19-37.

https://doi.org/10.5294/laclil.2018.11.1.2

Harrison, G. L., Goegan, L. D., Jalbert, R., McManus, K., Sinclair, K., & Spurling, J.

(2016). Predictors of spelling and writing skills in first- and second-language

learners. *Reading and Writing, 29*, 69-89. https://doi.org/10.1007/s11145-015-

9580-1

Hyland, K. (2007). Genre pedagogy: Language, literacy and L2 writing instruction.

Journal of Second Language Writing, 16(3), 148-164.

https://doi.org/10.1016/j.jslw.2007.07.005

Hyland, K. (2008). Writing theories and writing pedagogies. *Indonesian JELT, 4*(2), 1-20.

Hymes, D. (2005). Models of interaction of language and social life: Toward a descriptive

theory. In S. F. Kiesling & C. B. Paulston (Eds.*), Intercultural discourse and

communication: The essential readings* (pp. 4-16). Malden, MA: Blackwell

Publishing.

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal

study of children in first and second grade. *Journal of Educational Psychology,

78(*4), 243-255. https://doi.org/10.1037/0022-0663.78.4.243

Kim, Y.-S., Al Otaiba, S., Wanzek, J., & Gatlin, B. (2015). Toward an understanding of

dimensions, predictors, and the gender gap in written composition. *Journal of

Educational Psychology, 107*(1), 79-95.

Kim, Y.-S., Gatlin, B., Al Otaiba, S., & Wanzek, J. (2017). Theorization and an empirical

investigation of the component-based and developmental text writing fluency

construct. *Journal of Learning Disabilities, 51*(4), 320-335.

https://doi.org/10.1177/0022219417712016

Macaro, E. (2018). *English medium instruction*. Oxford, U.K.: Oxford University Press.

National Development Council. (2018). *Blueprint for developing Taiwan into a bilingual nation by 2030*. Retrieved from

https://www.ndc.gov.tw/en/Content_List.aspx?n=D933E5569A87A91C&upn=963 3B537E92778BB.

Schleppegrell, M. J., Achugar, M., & Oteíza, T. (2004). The grammar of history: Enhancing content-based instruction through a functional focus on language. *TESOL Quarterly, 38*(1), 67-93. https://doi.org/10.2307/3588259

2. CONTENT AND LANGUAGE INTEGRATED LEARNING WRITING MEASURES AND OUTCOMES: A SYSTEMATIC REVIEW

In recent decades, there has been a conceptual shift away from teaching English as a foreign language (EFL) where students *learn about* a foreign language toward teaching students to be able to *use* the language in a variety of contexts. In today's digital age, more and more of these contexts where EFL learners engage—whether personal, academic, or professional—involve the written form of English. While traditional forms of EFL teaching may be adequate for addressing written tasks in social contexts, such as an email to a friend, their curricula often do not adequately prepare students for the language used in academic and professional contexts. It is for this reason, according to Brinton and Snow (2017), that many language teaching programs have begun to integrate content with language. This type of language teaching in EFL settings is often referred to as Content and Language Integrated Learning (CLIL) (Brinton & Snow, 2017; Macaro, 2018).

The spread of CLIL around the world has been broad and rapid (Dearden, 2015). Many students are motivated to learn English with the hope to use their English abilities for academic and professional reasons, and access to such opportunities is often contingent upon mastery of academic English (Graham & Eslami, 2019). In the case of writing, early mastery of writing, particularly academic writing, may predict later success in academic and professional venues (Camacho & Alves, 2017). The promise of CLIL to provide these opportunities has inspired many governments to enact educational policies to promote these forms of language teaching (see Macaro, 2018, for a broad overview of policies around the world), while at the same time many question the current evidence of its

effectiveness (Bruton, 2011; Graham, Choi, Davoodi, Razmeh, & Dixon, 2018; Macaro, Curle, Pun, An, & Dearden, 2017).

Given the importance of writing in today's world coupled with past suggestions that CLIL may not be adequately developing students' writing abilities (Dalton-Puffer, 2008), it is important that scholarly work investigate writing development in CLIL settings. This review will examine how writing development in CLIL has been measured by researchers and report on current empirical evidence of CLIL's effectiveness in writing development. While previous reviews have considered writing as only part of the language side of CLIL (e.g., Graham et al., 2018), this review will be grounded in the idea that effective writing in CLIL should be measured in terms of both language and content. In the following section, this theoretical perspective will be further explained. The systematic search strategy of this review will then be detailed, and findings of the review will be reported. The paper will conclude with a discussion of these findings grounded in the theoretical framework of CLIL writing along with suggestions for future research on CLIL writing.

2.1. Theoretical Framework

Scholarly literature in this review will be examined from a theoretical perspective grounded in two premises: (1) CLIL is the integration of content and language and, thus, any examination of its effectiveness must consider both dimensions together and (2) that the integration of content and language in writing is represented by the Simple View of Writing (Berninger & Swanson, 1994; Juel, Griffith, & Gough, 1986), which posits that successful writing is comprised of two broad dimensions—ideation (content) and

transcription (language). In the following section, CLIL and the Simple View of Writing will be defined and how these two concepts complement each other will be explained.

2.1.1. Defining Content and Language Integrated Learning

Content and Language Integrated Learning (CLIL), as its name suggests, is the integration of content and language in the classroom. However, the exact nature of this integration has been the topic of much debate, and implementations of CLIL have varied greatly (Cenoz, Genesee, & Gorter, 2013). An increasing number of scholars suggest that integration in CLIL means an equal balance of content and language. For example, Dafouz, Camacho, and Urquia (2014) suggest CLIL is "a dual educational approach where *equal attention* [emphasis added] is paid to both content and language objectives" (p. 224). Macaro (2018) also positions CLIL as a balance of content and language but also suggests that the objectives of CLIL in terms of content remain conceptually unclear. It is perhaps this lack of clarity that has encouraged some scholars to give more emphasis to content and, subsequently, reduce the emphasis on language when defining CLIL. For instance, Ruiz de Zarobe and Jiménez Catalán (2009) suggest that language in CLIL is simply used to facilitate the learning of content. Ball, Kelly, and Clegg (2015) have asserted that CLIL can take on many forms from a "hard" CLIL which focuses on content exclusively to a "soft" CLIL which is more of a balance between language and content. Though the terminology debate is likely far from over, it seems most scholars now reserve the term CLIL for learning contexts where language and content receive equal attention and use the term English Medium Instruction (EMI) to describe settings where content is the main focus (Macaro, 2018).

Not inherent in the name, but equally important to define, is that CLIL (also EMI) has been generally reserved for "countries or jurisdictions where the first language of the majority of the population is not English" (Macaro, 2018, p. 37) or contexts where "for most participants in the setting, English is a second language (L2)" (Pecorari & Malmström, 2018, p. 499). It is this distinction that separates CLIL and EMI from other terms such as Content Based Language Teaching (CBLT), Content Based Instruction (CBI), or immersion, which are often used to describe the learning of the English language along with content in environments, such as North America, where English is the language of the majority and used in an official capacity throughout society (Macaro, 2018). This distinction is important because the nature of teaching and the stakes involved in CLIL/EMI versus CBI/CBLT are often different, where in the latter failure to learn English may have severe academic and professional consequences, whereas this is not necessarily the case in the former where mastery of academic and professional language in the first language of the country is sufficient and English only facilitates possible additional academic/professional benefits.

Taken together, successful implementations of CLIL should allow students from non-target language contexts to successfully engage in various venues, or content domains, using the target language (English in the case of this review). This requires learners to be able to use the language in ways that are appropriate in each respective content discipline, which have their own lexicon and discourse norms. Given that successful communication in each discipline requires adherence to its unique language norms, the content and language outcomes of CLIL cannot be viewed as bifurcated dimensions, and success in the various venues only occurs with true integration.

2.1.2. The Simple View of Writing

The Simple View of Writing defines writing through two components: transcription and ideation (Berninger & Swanson, 1994; Juel et al., 1986). Transcription can be defined as the orthographic elements of writing, or how language is put into written form. At the word level this could refer to the way words are formed and spelled, and at the sentence level this may extend to the syntactical and grammatical aspects of a language.

Ideation, or "the generation of ideas," is often represented through "vocabulary and domain knowledge" (Ritchey & Coker, 2013, pp. 91-92). Ideas often begin with vocabulary at the word level. Vocabulary in a language connects objects or abstract ideas to words, allowing for communication to occur. At the sentence and discourse level, these single ideas represented by vocabulary words are connected with other vocabulary words, thus communicating domain knowledge. Further, at the discourse level the presentation of this domain knowledge is often organized by the genre rules, which designate the way of communication within a discipline.

Far from what its name suggests, the Simple View of Writing shows that writing is a complex skill, and the model helps to conceptualize the complex system from which writing is created. Though this theory is only comprised of two broad dimensions— transcription and ideation—when these two dimensions are broken down at the level of word, sentence, and discourse, a complex hierarchical system emerges. From the perspective of the Simple View of Writing, successful writing can only occur when the various levels of transcription and ideation are mastered, and any evaluation of writing must consider both dimensions equally.

2.1.3. Connecting CLIL and the Simple View of Writing

Juxtaposed together, CLIL and the Simple View of Writing seem to fit well conceptually. CLIL goes beyond traditional EFL teaching that exclusively focuses on language and gives an equal importance to how languages intersect with content. The integration of language and content brings to the forefront the idea that various language-use contexts exist and that learners should be prepared to use various forms of language in various content domains. Likewise, the Simple View of Writing also presents a dual focus. On the one hand, the construct of transcription addresses the language side of writing. It encompasses the various linguistic aspects of writing at multiple levels. On the other hand, it also recognizes that writing goes beyond its linguistic components and recognizes that domain knowledge, in the form of ideas, is equally important in successful writing. When considering the evaluation of CLIL writing, the Simple View of Writing offers a framework which accounts for the integration of language and content through transcription and ideation, respectively.

2.2. The Current Review

Given the conceptual fit between CLIL and the Simple View of Writing, this review seeks to examine the state of research on CLIL writing through the lens of the Simple View of Writing. This review looks to explore the following research questions:

1. How has previous CLIL research measured writing?
2. What does current research show in terms of CLIL writing development outcomes?

The answers to these research questions will be viewed in light of the Simple View of Writing. More specifically, do the measures and outcomes of CLIL writing reported in research account for both dimensions—transcription and ideation.

2.3. Methods

In order to better understand the state of research on English writing outcomes in CLIL, this study utilized systematic search procedures to compile a comprehensive collection of research examining CLIL writing. The motivation for using systematic review procedures was to "locate, appraise, and synthesize the best available evidence relating to a specific research question" (Boland, Cherry, & Dickson, 2014, p. 3). This is achieved by clearly defining the process undertaken during the search to ensure a comprehensive search was conducted and minimize bias within the results that may occur from selectively choosing certain studies (Campbell Collaboration, 2017). In this section the study's inclusion criteria, search, screening, and coding of studies will be detailed.

2.3.1. Review Criteria

A critical requirement of the systematic review process is that criteria for inclusion must be set *a priori* (Uman, 2011). Prior to the search, the following criteria were set for studies to be included in the review:

1. The study must examine English writing beyond reporting a single holistic score.

2. The study's participants must be learning in an EFL context, in other words, in a location where English is not the first language of the majority of the population.

3. The study's participants must be learning English using a CLIL or EMI curriculum.

4. The study must be empirical, reported in English, and available in a written scholarly format through a journal, academic book chapter, conference proceeding, or dissertation/thesis.

First, studies included in this review were required to examine writing beyond a singular, holistic score. There is much literature that examines CLIL language outcomes broadly by

reporting single scores for language skills measured by standardized tests (e.g., Lorenzo, Casal, & Moore, 2010; Yang, 2015). Given the complexity of writing as a construct, studies which only report a singular writing score add little to our understanding of how the various dimensions of writing may develop and, thus, were not included. To be included, a study must examine one or various dimensions of English writing at the word, sentence, or discourse level. Additionally, this criterion specifies a focus on English writing, rather than writing in other languages. The additional factors that would need to be considered with differences in orthography between languages are beyond the scope of this review.

Second, the study must be set in an EFL context. Though the dichotomy of what constitutes an EFL context and what does not is far from straightforward and somewhat controversial, particularly when considering inner circle versus outer circle English contexts (Kachru, 1985), there are undeniable differences between the goals and stakes involved in learning English as a second language (ESL) in the United States or the United Kingdom, where lack of acquisition could have serious future consequences for a learner academically and economically, as opposed to EFL learning in, for example, Taiwan, where the consequences of not acquiring English are minimal in comparison. Further, this key distinction, ESL versus EFL, is one used by many to differentiate content-based instruction types around the world, with CLIL and EMI being exclusively in EFL contexts (Brinton & Snow, 2017; Lightbown, 2014; Macaro, 2018; Pecorari & Malmström, 2018).

Third, though there are certainly distinct differences between CLIL and EMI, most prominent being the degree of focus on language outcomes (Macaro, 2018), both were allowed inclusion in this article, rather than exclusively including CLIL programs. The

motivation for this is two-fold. First, as noted previously in this paper, there has been much

confusion around the terminology of CLIL and EMI (Airey, 2016; Macaro, 2018), causing

various researchers to either incorrectly classify programs or simply see the terminology as

interchangeable. Second, it cannot be ignored that the degree in which language outcomes

take a role in CLIL classrooms vary greatly (Brinton & Snow, 2017; Macaro, 2018). This

variation has inspired terms such as hard CLIL versus soft CLIL to better define the degree

of language focus (Ball et al., 2015). Acknowledging this spectrum and historical

ambiguity within the literature, both CLIL and EMI were included in this review.

Finally, the study must be an empirical investigation of CLIL writing reported in

English. In order to be as inclusive as possible, all types of scholarly publications (i.e.,

journals, book chapters, dissertations/theses, and conference proceedings) were included.

However, only studies written in English were included in this review due to the linguistic

limitations of the author. Admittedly, this is a major limitation, particularly given that the

review reports on international English learning. It is acknowledged that there are likely

various reports on CLIL writing written in languages other than English that may not have

been included. It is hoped that as a result of the systematic search and inclusiveness of all

publication types that any foreign language articles missed would only serve to enhance,

rather than counter, the findings of the review, though future reviews of non-English

studies would be needed to confirm this.

2.3.2. Search Process

The systematic search process for this review involved several steps and strategies.

First, three electronic databases were initially used to retrieve scholarly articles: ERIC for

education articles, PsycINFO for the field of psychology, and Linguistics and Language

Behavior Abstracts for studies in linguistics. The search involved two clusters of words. The first cluster comprises of various terminology or acronyms that researchers may have used in describing CLIL settings connected with the Boolean term OR: "content and language integrated learning" OR CLIL OR immersion OR "content based language teaching" OR CBLT OR "content based instruction" OR CBI OR "English medium instruction" OR EMI. The parameter for this cluster was set to search abstracts. The second cluster was used to further limit the search to research on writing. The term "writ*" was used in the second field, with the asterisk allowing for the inclusion of writing, write, and written. This field's parameters were set to only search for these terms in the title to allow only for studies directly related to writing to be retrieved. Additionally, the search parameter for "peer-reviewed" was selected. This initial database search is, admittedly, very restrictive in terms of the parameters set and was designed to first find an initial set of articles that would act as a base for subsequent searches detailed below.

Two other sources were included to compensate for the "peer-reviewed" parameter limitation. A fourth database, ProQuest Dissertations & Theses Global, was used to search for dissertations and theses that were not included in the search results of the initial three databases. The search in ProQuest Dissertations & Theses Global followed a similar search strategy with two clusters. Abstracts were searched using "content and language integrated learning" OR CLIL OR "English medium instruction" OR EMI, and titles were searched for "writ*." To further compensate for the restrictive parameters of the initial search, the reference sections of several comprehensive literature reviews on CLIL and EMI were used to retrieve articles specifically related to writing (Goris, Denessen, & Verhoeven, 2019; Graham et al., 2018; Macaro et al., 2017; Tedick & Wesely, 2015).

Once a base set of studies were retrieved using the strategies above, several

iterations of forward and backward searches were conducted. Forward searches were

conducted using Google Scholar's "Cited By" feature, which returns a list of articles that

have cited the searched article. Titles of every article were put into Google Scholar's

search field. Then, the "Cited By" link for the article was clicked, and the returned titles

were searched for relevant studies. For the backward search, the titles in the reference

sections of all included articles were examined for studies meeting the criteria of this

review. This process was conducted several times as new articles were found and

concluded once no additional studies were returned. The search for this study was

conducted September/October 2019.

2.3.3. Search and Screening Results

Figure 2.1 displays the search and screening results of the systematic search. For

the initial database search, ERIC returned 24 records, PsycINFO returned 57 records, and

LLBA returned 66. The search conducted through ProQuest Dissertations & Theses Global

returned an additional three, for a total of 150 records. These records were uploaded into

Rayyan, a free online software designed to assist with the screening process of systematic

reviews (Ouzzani, Hammady, Fedorowicz, & Elmagarmid, 2016). After an automated and

manual screening for duplicate records, 40 records were identified as duplicates and

removed, leaving 110 records for screening.

Using the inclusion criteria as detailed above, titles and abstracts were screened and

marked for inclusion/exclusion. For the first criteria 45 articles were excluded for not being

research on English writing, and 15 were excluded for being conducted outside of an EFL

context. An additional 15 articles were excluded for not having participants using a CLIL

or EMI curriculum. Finally, 23 articles were excluded for not being empirical and three for

being written in a language other than English. In total, 101 articles were marked for

exclusion with nine studies remaining. A search of past reviews on CLIL and EMI added

an additional four articles, for a total of 13.

Next, the iterations of forward and backward searches were conducted using the 13

articles that met the criteria for inclusion in this review. Three iterations of this process

were conducted before no further articles were found.
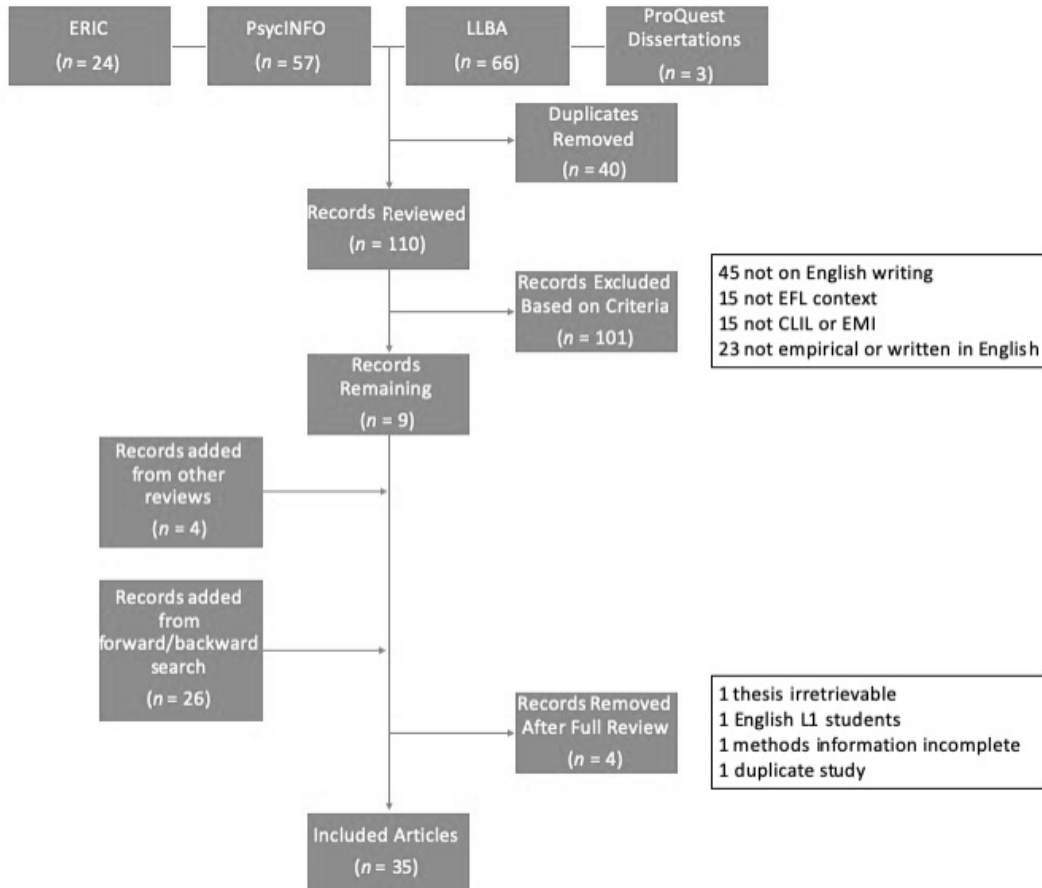
Figure 2.1 Systematic review search process and screening results.

The first iteration resulted in ten articles from the backward search and three from the

forward search. The second iteration returned six from the backward search, though one

master's thesis was irretrievable bringing the total to five, and five from the forward

search. The third, and final, iteration returned one article for both the backward and

forward search, respectively. After completing searches of databases, existing reviews, and

backward/forward searches, a total of 38 articles were marked for inclusion. However,

upon further inspection one thesis had a significant population of native English speakers,

one journal article seemed to report on the same study as another included in the review,

and one journal article lacked appropriate population and procedural information in the

methods section. Thus, these were removed, bringing the final total to 35 articles.

2.4. Findings

This section will begin with a brief overview of included articles in terms of

publication type, country, education level, and writing types. Following this, findings for

each of the research questions addressing writing measurement and CLIL writing

outcomes will be presented.

2.4.1. Overview of Articles

Table 2.1 presents an overview of articles including author (year), publication type,

country, education level, writing types, writing measurements, and a brief summary of

results.

Table 2.1 Overview of Studies Included in Systematic Review

Author (Year)	Publication	Country	Level	Writing	Measurement	Results
Ackerl (2007)	J	Vienna	S	NR	Error Analysis	No difference in number of errors between CLIL and non-CLIL, but CLIL errors are more advanced.
Adamson & Coulson (2015)	J	Japan	U	Exp	Citation Count	Number of L1 citations decreased over time despite encouragement to translanguage.
Agustín Llach & Jiménez Catalán (2007)	J	Spain	P	Nar	Vocabulary	CLIL uses more unique words than non-CLIL, but lexical reiteration is generally low across both groups.
Agustín-Llach (2016)	J	Spain	P	Nar	Vocabulary	No significant difference in CLIL and non-CLIL lexical borrowings, lexical creations, and lexical profile.
Agustín-Llach (2017)	J	Spain	P	Nar	Vocabulary	No significant difference in CLIL and non-CLIL lexical production.
Bennett (2017)	D/T	Japan	U	Exp	Metaphor Use	No difference in overall metaphor usage between CLIL experimental and CLIL control but differences found in explicitly taught metaphors.
Celaya (2008)	BC	Spain	P/S	Nar	Vocabulary	Lower percentage of lexical borrowings for CLIL compared with non-CLIL but similar percentages for lexical inventions.
Chansri & Wasanasomsithi (2016)	J	Thailand	U	Exp	Rubric	CLIL group showed a significant improvement in writing between a pre- and post-test.
Coetzee-Lachmann (2007)	D/T	Germany	S	Exp	Rubric	CLIL writing did not show content mastery, and language appropriateness more of a problem than structure and linking.

Table 2.1 Continued

Author (Year)	Publication	Country	Level	Writing	Measurement	Results
Gené-Gil et al. (2015)	J	Spain	S	Nar	CAF	Significant increase across three years on all measures for CLIL but only lexical complexity and accuracy for non-CLIL. Non-CLIL progress more in lexical complexity than CLIL, but CLIL higher overall.
Gené-Gil et al. (2015)	BC	Spain	S	Nar	Rubric & CAF	CLIL showed development in most, but not all, measures of complexity, accuracy, and fluency, whereas non-CLIL only in accuracy. CLIL showed higher achievement in non-CLIL on CLIL ratings of organization, language use, and total score.
Järvinen (2010)	BC	Finland	S	Exp	SFL	International school students showed greater grasp of subject-specific linguistic convention and devices than CLIL group.
Jexenflicker & Dalton-Puffer (2010)	BC	Austria	S	Nar	Rubric	CLIL stronger than non-CLIL for task fulfillment, vocabulary, and grammar. Both groups weak on organization, though still statistically different.
Kjellén Simes (2008)	D/T	Sweden	S	Nar	Vocabulary	EMI group had higher levels of low frequency vocabulary and motivated tense shifts than control. Low performing EMI students showed more development than low performing control students.
Kong (2010)	J	Hong Kong	S	Exp	Content Analysis	Much of the student writing was copied from other texts and showed little understanding of genre features.

Table 2.1 Continued

Author (Year)	Publication	Country	Level	Writing	Measurement	Results
Lahuerta (2017)	J	Spain	S	Per	Error Analysis	CLIL showed greater improvement in errors. Non-CLIL had higher proportion of errors but no differences were found when broken down by error types.
Lahuerta Martínez (2017)	J	Spain	S	Per	Rubric & Complexity	CLIL significantly higher in all measures over non-CLIL group.
Llinares & Whittaker (2007)	J	Spain	S	Exp	SFL	Writing shows CLIL students beginning to acquire some features of the discipline but elaboration rarely used.
Llinares & Whittaker (2010)	BC	Spain	S	Exp	SFL	CLIL writing similar to English L1 writing in process types but show limited ability in circumstances and present more personal views.
Manzano Vázquez (2014)	J	Spain	S	Nar	Vocabulary	No statistical difference found between CLIL and non-CLIL in lexical transfer errors after removal of outliers.
Miret Pallarols (2009)	D/T	Spain	S	Nar	Rubric	CLIL outperformed non-CLIL in all measures.
Navés (2011)	BC	Spain	P/S	NR	CAF	CLIL outperformed non-CLIL in all measures except accuracy.
Olsson (2015)	J	Sweden	S	Per/Exp	Vocabulary	Differences over time shown between CLIL and non-CLIL for Academic Word List and Academic Vocabulary List word usage but regression analysis shows similar development rates.
Olsson & Sylvén (2017)	J	Sweden	S	Per/Exp	Vocabulary	CLIL students did not use more academic vocabulary than non-CLIL students.

Table 2.1 Continued

Author (Year)	Publication	Country	Level	Writing	Measurement	Results
Pérez & Basse (2015)	C	Spain	P	Nar	Error Analysis	No statistical difference in errors between CLIL and non-CLIL groups.
Perez & Ramiro (2015)	J	Spain	U	Exp	Content Analysis	Differences between low-rated texts and high-rated texts were found at the text and discourse level. Texts were found to be similar in morphosyntactic and lexicosemantic performance.
Pessoa et al. (2014)	J	Qatar	U	Per/Exp	SFL	Increasing trends were found for markers of academic writing, though inconsistent and non-linear between time points.
Ramiro & Pérez (2015)	J	Spain	U	Exp	Rubric	EMI student writing was found to be above average on the rubric in all aspects except organization. Students displayed weaknesses in text format, register, and structure and discourse issues.
Roquet & Pérez-Vidal (2017)	J	Spain	S	Nar	Rubric & CAF	No statistical difference found between CLIL and non-CLIL except for accuracy, which showed CLIL higher.
Ruiz de Zarobe (2010)	BC	Spain	S	Nar	Rubric	Overall, CLIL groups performed significantly better than non-CLIL in content and vocabulary. The intensive CLIL group scored higher in all measures.
Tai (2015)	J	Taiwan	U	Exp	CAF	Showed CLIL improvement in most measurements of accuracy and fluency but not complexity.

Table 2.1 Continued

Author (Year)	Publication	Country	Level	Writing	Measurement	Results
Vidal & Jarvis (2018)	J	Spain	U	Per	Vocabulary	No differences between Year 1 and 3 EMI students for lexical diversity, but Year 3 had higher ratings on CEFR rubric.
Whittaker & Llinares (2009)	BC	Spain	S	Exp	SFL	CLIL students seem to show progress toward disciplinary writing but distance in certain areas in need of development remain.
Whittaker et al. (2011)	J	Spain	S	Exp	SFL	CLIL students showed development away from listing topics toward reporting more detail and depth.
Xanthou (2017)	D/T	Greece	P	Exp	Vocabulary	Few differences found in vocabulary use between three levels of CLIL students.

2.4.1.1. Publication Types

The majority of articles included in this review are published journal articles ($n =$ 21). Eight studies were published as book chapters, and five were either doctoral dissertations or master's theses. One study is a published conference proceeding paper.

2.4.1.2. Country

Given that the origins of CLIL are attributed to Europe (Marsh, 2009), it is of little surprise that 29 of the 35 studies were conducted in European countries. The bulk of these studies ($n = 21$) are from Spain, three from Sweden, two from Austria, and one each from Finland, Germany, and Greece. The six remaining articles all came from Asia, with one from the Middle East (Qatar), one from Southeast Asia (Thailand), and four from East Asia. Two of the East Asian studies came from Japan, and one each came from Hong Kong and Taiwan. No studies were found for inclusion from the continents of South America or Africa.

2.4.1.3. Education Level

Twenty studies were conducted with secondary students at the junior high or high school level. Studies at the university level were the second most prominent with eight studies. Primary school studies were the least common with five studies containing populations from this level of education. Two additional studies used a combination of populations from upper primary and secondary schools.

2.4.1.4. Writing Types

There has been fairly equal attention given to the investigation of narrative ($n = 13$) and expository ($n = 14$) genres within CLIL writing research. On the one hand, this may be somewhat unexpected given the emphasis of content within CLIL which may require

expository writing, yet on the other hand, it is of no surprise when considering that most studies are comparative studies of CLIL and non-CLIL groups, where narrative may be a common genre used in both groups. Only three studies investigated persuasive/ argumentative writing genres, and three studies used a combination of expository and persuasive writing samples. Two studies did not report the genre of writing used in the studies.

2.4.2. RQ1: How Has Previous CLIL Research Measured Writing?

In this section, the various methods of measuring writing in studies of CLIL writing will be discussed. Various types of writing measurement methods have been used including measures of vocabulary, rubrics, complexity/accuracy/fluency (CAF) measures, and content analysis.

2.4.2.1. Measures of Vocabulary

Ten studies used various methods to analyze the vocabulary used within CLIL writing, though in a variety of different ways. Many studies examined lexical diversity of writing using various measures of type/token ratios (Agustín Llach & Jiménez Catalán, 2007; Olsson & Sylvén, 2017; Vidal & Jarvis, 2018; Xanthou, 2017). The type/token ratio (TTR), in its basic form, is a ratio of the amount of unique words (types) to total words (tokens). However, this metric alone is sensitive to word count and cannot be used accurately to compare writing of different lengths. To compensate for this limitation, researchers in the included articles have used TTR variations such as the Moving Average Type/Token Ratio (MATTR) (Olsson & Sylvén, 2017; Vidal & Jarvis, 2018; Xanthou, 2017), where a TTR is calculated and averaged for each specified word length, for example every set of 100 words (1-100, 2-101, etc.). In addition to MATTR, Vidal and Jarvis

(2018) used two other variations of TTR: the Measure of Textual Lexical Diversity (MTLD), where "MTLD is calculated as a text's average maximum number of running words that maintain a TTR of at least 0.72" (p. 8), and the MTLD Wrap Around (MTLD-W), which is similar to MTLD except that it wraps around to the beginning of a text until TTR drops below 0.72 instead of stopping calculations at the final word of the text. Other researchers have simply calculated TTR for the first 100 words (Agustín Llach & Jiménez Catalán, 2007).

Another popular measure of vocabulary is the measure of lexical profiles according to vocabulary level lists (Agustín-Llach, 2016, 2017; Kjellén Simes, 2008; Olsson, 2015; Olsson & Sylvén, 2017; Xanthou, 2017). A variety of lists have been used as a reference for profiling vocabulary levels. Profilers from the website Lextutor.ca, both general and those designed for young learners, have been frequently used. These tools allow for the tabulation of the most common words based on a variety of vocabulary lists such as the VP-Kids list, which profiles vocabulary based on ten levels of 250 word used by young language learners (Agustín-Llach, 2016, 2017; Xanthou, 2017). Other lists include word frequency from the British National Corpus (Xanthou, 2017) and the Academic Word List (Olsson, 2015). Other computer applications have been used to tabulate vocabulary levels from other lists including Collins Cobuild Frequency Bands (Kjellén Simes, 2008), the Academic Vocabulary List (Olsson, 2015; Olsson & Sylvén, 2017), and the CEFR vocabulary list (Xanthou, 2017). Olsson (2015) compared results of using the Academic Word List versus the Academic Vocabulary List and found the Academic Vocabulary List to be more sensitive to the changes in CLIL student writing.

Finally, various studies have examined writing in terms of specific types of vocabulary including lexical reiteration (Agustín Llach & Jiménez Catalán, 2007) and lexical first language (L1) transfer errors through lexical borrowings and lexical inventions/creations (Agustín-Llach, 2016; Celaya, 2008; Manzano Vázquez, 2014). Agustín Llach and Jiménez Catalán (2007) examined lexical reiteration as a way to measure lexical cohesiveness, where cohesion is created in writing with the repetition of words in various forms such as word repetition, hyponyms, synonyms, general nouns, meronyms, and antonyms. Other studies have looked specifically at the negative cross-linguistic influence of the L1 on English writing through lexical transfer errors. Researchers have examined English writing for the presence of lexical borrowings, where a writer uses an L1 word within their writing, or lexical inventions/creation, where the writer adapts L1 words using rules from the L2. Studies have generally looked for a decrease in lexical transfer errors as evidence of L2 writing growth.

2.4.2.2. Rubrics

Writing rubrics are grading tools that measure multiple dimensions of writing. Nine studies included in this review utilized rubrics (Chansri & Wasanasomsithi, 2016; Coetzee-Lachmann, 2007; Gené-Gil, Juan-Garau, & Salazar-Noguera, 2015a; Jexenflicker & Dalton-Puffer, 2010; Lahuerta Martínez, 2017; Miret Pallarols, 2009; Roquet & Pérez-Vidal, 2017; Ruiz de Zarobe, 2010; Sagrario Salaberri Ramiro & Del Mar Sánchez Pérez, 2015). The majority of these studies used a 5-point rubric called the ESL Composition Profile (Jacobs, Zinkgraf, Wormuth, Hartfield, & Hughey, 1981) comprised of scores for content, organization, vocabulary, language use, and mechanics. Content measures the degree of development of the topic and whether the task was fulfilled. Organization

concerns the structure and cohesion of the writing. Vocabulary examines the kinds of words used and their appropriateness within the writing. Language use accounts for morpho-syntactic issues, and mechanics includes issues of spelling, capitalization, and punctuation. Most other studies not using the 5-point rubric above used a 4-point rubric with similar categories: task fulfillment (content), organization, grammar, and vocabulary.

Of the studies using rubrics, only Coetzee-Lachmann (2007) used a rubric that went beyond these categories. One of the main goals of this dissertation was to create "analytic scales that can be used across different tasks and subjects to rate learners' language use in their *subject-specific* [emphasis added] writing" (Coetzee-Lachmann, 2007, p. 3). This is a different approach than other studies in that it aimed to measure subject-specific aspects, rather than just general writing characteristics. The rubric developed in the study, broadly speaking, measures (a) the completeness and correctness of meaning constructed and (b) the appropriateness of the textual realization of meaning. The "completeness and correctness of meaning constructed" part of the rubric directly addresses the content side of CLIL writing, examining writing for discipline-specific content (e.g., "Is there an identification of a particular type of climate?" [Coetzee-Lachmann, 2007, p. 268]). The "appropriateness of the textual realization of meaning" part of the rubric, on the other hand, looks at many of the aspects from the general rubrics above such as vocabulary and grammar but also looks for specific terminology and formal language conventions associated with the genre.

2.4.2.3. Complexity/Accuracy/Fluency

Complexity, accuracy, and fluency (CAF) have, for many years, been the metrics used for measuring second language development (Larsen-Freeman, 2009). Nine studies in

this review used one or all of these CAF metrics to measure CLIL writing (Ackerl, 2007; Celestén Pérez & Basse, 2015; Gené-Gil et al., 2015a; Gené-Gil, Juan-Garau, & Salazar-Noguera, 2015b; Lahuerta, 2017; Lahuerta Martínez, 2017; Navés, 2011; Roquet & Pérez-Vidal, 2017; Tai, 2015).

Complexity in these studies often examines syntactic or lexical complexity (Table 2.2). Syntactic complexity has been measured in various ways through length or number of t-units (defined as an independent clause and its associated dependent clauses, if any), independent clauses, dependent clauses, simple sentences, compound sentences, complex sentences, and compound-complex sentences. Ratios are often calculated using one or various syntactic measures by dividing these by total t-units, total clauses, or total sentences. One study also measured syntactic complexity at the word level by calculating the average words per sentence (Lahuerta Martínez, 2017). The measurement of lexical complexity is equally diverse and is generally calculated with word types, a unique count of words, and tokens (the total number of words). Various measures used for lexical complexity include counts of noun types, verb types, adjective types, and adverb types, as well as variations on the type/token ratio (D-Value, Guiraud's Index).

Table 2.2 Measures of Complexity

Complexity Measures	Category	Calculation	Studies
Mean Sentence Length	Syntactic	Average Words per Sentence	Lahuerta Martínez (2017)
Number of Subordinate Clauses	Syntactic	Count of Subordinate Clauses	Navés (2011)
Number of Coordinated Clauses	Syntactic	Count of Coordinated Clauses	Navés (2011)
Simple Sentence Ratio	Syntactic	Simple Sentences/Total Sentences	Lahuerta Martínez (2017)
Compound Sentence Ratio	Syntactic	Compound Sentences/Total Sentences	Lahuerta Martínez (2017)
Complex Sentence Ratio	Syntactic	Complex Sentences/Total Sentences	Lahuerta Martínez (2017)
Compound-Complex Sentence Ratio	Syntactic	Compound-Complex Sentences/Total Sentences	Lahuerta Martínez (2017)
Coordination Index	Syntactic	Independent Clauses/ Total Clauses	Gené-Gil et al. (2015a), Gené-Gil et al. (2015b), Roquet and Pérez-Vidal (2017)
Coordination Ratio	Syntactic	T-Units/ Total Sentences	Gené-Gil et al. (2015a), Gené-Gil et al. (2015b) , Lahuerta Martínez (2017)
Clauses per T-Unit	Syntactic	Clauses/ Total T-Units	Tai (2015)
Subordinate per Clause	Syntactic	Subordinate Clauses/Total Clauses	Navés (2011)
Dependent Clauses per T-Unit	Syntactic	Dependent Clauses/ Total T-Units	Tai (2015)
Subordination Ratio	Syntactic	Clauses/Total Sentences	Gené-Gil et al. (2015b) , Lahuerta Martínez (2017), Navés (2011)
Dependent Clause Ratio	Syntactic	Dependent Clauses/ Total Sentences	Lahuerta Martínez (2017), Navés (2011)

Table 2.2 Continued

Complexity Measures	Category	Calculation	Studies
Mean Length of Noun Phrase	Syntactic	Average Words per Noun Phrase	Lahuerta Martínez (2017)
Noun Types	Lexical	Count of Distinct Noun Types	Navés (2011)
Verb Types	Lexical	Count of Distinct Verb Types	Navés (2011)
Adjective Types	Lexical	Count of Distinct Adjective Types	Navés (2011)
Adverb Types	Lexical	Count of Distinct Adverb Types	Navés (2011)
D-Value	Lexical	Alternative Type/Token Ratio Calculation	Gené-Gil et al. (2015b)
Guiraud's Index	Lexical	Alternative Type/Token Ratio Calculation	Roquet and Pérez-Vidal (2017), Navés (2011)

Like complexity, many of the metrics used to measure accuracy involve various ratios using t-units, clauses, and sentences (Table 2.3). These ratios are either created through error-free instances or instances with errors of a unit divided by a total. Metrics used in the studies included in this review are the error ratio, error t-unit ratio, error-free t-unit ratio, error-free clause ratio, and error-free sentence ratio. Additionally, three studies conducted error analysis to examine error types (Ackerl, 2007; Celestén Pérez & Basse, 2015; Lahuerta, 2017). Ackerl (2007) and Celestén Pérez and Basse (2015) use three

classifications of errors: substance errors (spelling and punctuation), text errors (lexical and grammatical), and discourse errors (cohesion and appropriateness). Though Lahuerta (2017) does not use these categories, specifically, the error types examined—syntactic, morphological, lexical, lexico-grammatical, spelling, and punctuation—do seem to be accounted for in the substance, text, and discourse error categories described above.

Table 2.3 Measures of Accuracy

Accuracy Measures	Calculation	Studies
Error Ratio	Number of Errors/Total Words	Lahuerta (2017), Roquet and Pérez-Vidal (2017)
Error-Free T-Units Ratio	Number of Error-Free T-Units/Total T-Units	Tai (2015), Gené-Gil et al. (2015a), Gené-Gil et al. (2015b)
Error T-Unit Ratio	Number of T-Units with Errors/ Total T-Units	Gené-Gil et al. (2015a), Gené-Gil et al. (2015b)
Error-Free Clauses Ratio	Number of Error-Free Clauses/Total Clauses	Tai (2015)
Error-Free Sentence Ratio	Number of Sentences /Total Sentences	Lahuerta (2017), Navés (2011)

Finally, measures of fluency are generally calculated through counts or through a ratio of counts per minute (Table 2.4). Fluency measures used in the included studies of this review are total words, total t-units, total clauses, total sentences, words per t-unit, words per minutes, and t-units per minute.

Table 2.4 Measures of Fluency

Fluency Measures	Calculation	Studies
Total Words	Count of Total Words	Roquet and Pérez-Vidal (2017), Tai (2015)
Total T-Units	Count of Total T-Units	Tai (2015)
Total Clauses	Count of Total Clauses	Navés (2011)
Total Sentences	Count of Total Sentences	Navés (2011)
Ratio of Words per T-Unit	Total Words/Total T-Units	Gené-Gil et al. (2015a), Gené-Gil et al. (2015b), Tai (2015)
Ratio of Words per Minute	Total Words/Total Minutes of Writing	Gené-Gil et al. (2015a), Gené-Gil et al. (2015b)
Ratio of T-Units per Minute	Total T-Units/Total Minutes of Writing	Gené-Gil et al. (2015a), Gené-Gil et al. (2015b)

2.4.2.4. Content Analysis

Ten studies in this review performed various forms of content analysis, examining specific features of the writing. Six studies examined the content of writing from a perspective grounded in Systemic Functional Linguistics (SFL) (Järvinen, 2010; Llinares & Whittaker, 2007; Llinares & Whittaker, 2010; Pessoa, Miller, & Kaufer, 2014; Whittaker & Llinares, 2009; Whittaker, Llinares, & McCabe, 2011). SFL focuses on the way language performs functionally in various contexts, and the studies in this review specifically looked at how systematic functional grammars are attained in CLIL texts of various academic disciplines. The majority of SFL studies in this review examine the use of language in terms of the ideational function, which is the use of language to represent, and the interpersonal function, or the use of language to exchange (Llinares & Whittaker, 2007; Llinares & Whittaker, 2010; Whittaker & Llinares, 2009; Whittaker et al., 2011).

Järvinen (2010) also examines the development of writing in terms of the ideational function through grammatical metaphor, but this study also addresses the textual function, or language as message, by examining syntactic intricacy and thematic organization. Though SFL analysis is generally conducted through human coding, Pessoa et al. (2014) used the corpus-based tool DocuScope to perform a computer-coded SFL analysis of institutional register, academic register, reasoning, elaboration, and reporting.

Kong (2010) and Del Mar Sanchez Perez and Sagrario Salaberri Ramiro (2015) conducted qualitative content analyses of various dimensions of writing. Kong (2010) examined the way CLIL writers approach content, audience, purpose, and genre. Content examined whether writing demonstrated a knowledge of content, with similar aims to the rubric used by Coetzee-Lachmann (2007). Writing was also analyzed for the audience and purpose in which writing was geared toward, looking specifically for the various types of audience and purpose CLIL may, or may not, support and encourage. Finally, the authors examined whether various writing exhibited properties characteristic of the genres written. Rather than looking at broad aspects of writing, Del Mar Sanchez Perez and Sagrario Salaberri Ramiro (2015) analyzed writing at its various levels—lexiosemantic level, morpho-syntactic level, discourse level, and textual level. The authors used a checklist to guide their analysis for each level, similar to a rubric but without the quantitative score. The lexiosemantic level examined the type of vocabulary used at the word level while the morpho-syntactic level focused at the clause/sentence level and the grammatical choices within the writing. At the discourse level, the authors looked specifically for elements of cohesion and coherence that support the flow of the writing. Finally, the textual level, or

genre macrostructure, looked specifically for the sections required by the genre, which in the case of this study was a lab report.

Finally, two studies focused exclusively on examining specific features within writing. Adamson and Coulson (2015) were interested in the citations used in academic writing. Specifically, the researchers examined the variation in first language and second language citations over time in order to gauge students use of translanguaging in constructing academic papers for class. Bennett (2017) also investigated the presence of a specific feature of writing—metaphor use. Acknowledging the importance of metaphors in various disciplinary writing, the author examined the writing for presence of and increase in metaphors used by CLIL writers.

2.4.3. RQ2: What Does Current Research Show in Terms of CLIL Writing Development Outcomes?

Having looked at the way writing has been measured in CLIL, this section will now address findings on the influence of CLIL on English writing. The studies in this review take one of two approaches, either (a) comparing CLIL with a non-CLIL control group or (b) examining writing production or growth of CLIL writers. Both of these types of studies provide us with information on the efficacy of CLIL for developing writing. Studies with non-CLIL control groups serve to explore the question of whether CLIL produces superior results compared with traditional forms of language teaching. Studies looking exclusively at the writing of CLIL students, without a non-CLIL control, serve to provide information on various aspects of CLIL writing and highlight various shortcomings in need of remedy, either through teacher professional development or curriculum redesign. A synthesis of results for these two types of studies is provided in the sections that follow.

2.4.3.1. CLIL Versus Non-CLIL

One major question asked by researchers, policy makers, and practitioners is whether or not CLIL will facilitate better student outcomes than traditional language teaching. Just over half of the studies ($n = 19$) explore this question, specifically examining writing outcomes. In most of the studies in this review, CLIL students either perform equally or better than non-CLIL students in a variety of metrics, though with a few exceptions.

Five studies used scores from rubrics to compare CLIL and non-CLIL groups. The rubrics used in the reviewed studies examined aspects of content (task fulfillment), organization, vocabulary, language use (grammar), and mechanics. In terms of content and task fulfillment, studies seem to be split on whether CLIL is superior in this metric with some finding CLIL groups scoring significantly higher (Jexenflicker & Dalton-Puffer, 2010; Miret Pallarols, 2009; Ruiz de Zarobe, 2010), and two studies finding no significant differences (Gené-Gil et al., 2015b; Roquet & Pérez-Vidal, 2017). There is a similar split in study results for organization as well, with three studies finding significant difference in favor of CLIL (Gené-Gil et al., 2015b; Jexenflicker & Dalton-Puffer, 2010; Miret Pallarols, 2009), and two studies finding no significant differences (Roquet & Pérez-Vidal, 2017; Ruiz de Zarobe, 2010). However, though Jexenflicker and Dalton-Puffer (2010) found CLIL scoring higher on organization than non-CLIL groups, the authors noted that the writing of both groups was generally weak in terms of paragraphing and structure. With the exception of one study (Roquet & Pérez-Vidal, 2017), the studies in this review have found that CLIL groups score higher on the vocabulary dimension of the rubric than non-CLIL groups. Finally, there are mixed results for language use/grammar, with three

studies favoring CLIL (Gené-Gil et al., 2015b; Jexenflicker & Dalton-Puffer, 2010; Miret

Pallarols, 2009). In terms of mechanics, there were no studies reporting significant

differences.

Five studies used one or more CAF metrics to examine differences between CLIL

and non-CLIL groups. Much like studies using rubrics, studies using CAF have found

mixed results. For syntactic complexity, two studies have found CLIL groups to score

higher than non-CLIL (Lahuerta Martínez, 2017; Navés, 2011) and three found no

differences between CLIL and non-CLIL groups (Gené-Gil et al., 2015a, 2015b; Roquet &

Pérez-Vidal, 2017). In terms of lexical complexity, none of the studies found significant

differences between the groups. Somewhat in contrast with the results found in studies

using rubrics, non-CLIL groups have been found in some studies to perform better than

CLIL groups in the accuracy dimension (Navés, 2011). Though Gené-Gil et al. (2015a)

found no significant differences between groups for accuracy, the study found that non-

CLIL groups made greater gains over time than CLIL groups. Only one study found CLIL

to write significantly more accurate texts than non-CLIL (Roquet & Pérez-Vidal, 2017).

Finally, only one study found CLIL to outperform non-CLIL for fluency (Navés, 2011),

with most studies finding no significant differences between the two groups (Gené-Gil et

al., 2015a, 2015b; Roquet & Pérez-Vidal, 2017).

Studies using error analysis to compare groups have generally found no difference

between groups in terms of errors (Ackerl, 2007; Celestén Pérez & Basse, 2015; Lahuerta,

2017), contrasting with the CAF studies which found some non-CLIL groups to write more

accurate texts. Studies specifically examining lexical transfer errors have also generally

found mixed results with one study suggesting no difference in L1 transfer errors (Agustín-

Llach, 2016) and two studies finding less transfer errors for the CLIL group (Celaya, 2008; Manzano Vázquez, 2014). However, it should be noted that one of these studies did not run any statistical significance tests (Celaya, 2008). The discrepancy between studies using error analysis and those using CAF measures may be explained by the methods of measurement. As Ackerl (2007) notes, CLIL does not necessarily make fewer mistakes than non-CLIL, but the types of errors made are different. The author suggests that CLIL students write more complex texts which result in different types of errors made. On the other hand, Lahuerta (2017) notes that while non-CLIL may have a higher proportion of errors, the distribution of error types tends to be similar between groups.

Finally, six studies examined CLIL and non-CLIL groups in terms of lexical production (Agustín Llach & Jiménez Catalán, 2007; Agustín-Llach, 2016, 2017; Kjellén Simes, 2008; Olsson, 2015; Olsson & Sylvén, 2017). As is the case with other metrics reviewed above, the studies examining general and academic vocabulary in texts have reported mixed results. For general vocabulary, Agustín Llach and Jiménez Catalán (2007) found that CLIL texts had a higher lexical diversity than non-CLIL texts, but Agustín-Llach (2016, 2017) found little difference when examining vocabulary production by frequency bands and type/token ratios. Given the focus on content subjects in CLIL, one might expect that academic vocabulary would be higher within CLIL texts. Kjellén Simes (2008) examined low frequency vocabulary use, which the author links to academic vocabulary, and found greater use of low frequency vocabulary for CLIL. However, two studies have shown similar use of academic vocabulary within texts between CLIL and non-CLIL groups when using academic word lists as the metric (Olsson, 2015; Olsson & Sylvén, 2017).

49

2.4.3.2. CLIL Writing Development and Growth

The remaining sixteen studies examined CLIL writing development and growth within CLIL without a non-CLIL control group. Studies in this category are in one of two broad groups—(1) those comparing CLIL over time or at different proficiency levels and (2) those examining CLIL writing for genre-specific features. Various studies have shown that engagement in CLIL instruction may result in growth in various areas of writing over time. A study of CLIL university students by Tai (2015) found statistically significant growth for accuracy and fluency, but not complexity, over one semester. Adamson and Coulson (2015) showed that the use of Japanese L1 citations within English writing decreased over time, despite the instructors' encouragement of translanguaging, which if interpreted from a different perspective than that of the authors may indicate increased proficiency with the incorporation of English L2 texts within academic writing. Chansri and Wasanasomsithi (2016) also found growth in writing as a result of CLIL instruction on a 5-point rubric for content, organization, vocabulary, language use, and mechanics between the beginning and the end of the semester. Using a cross-sectional sample of first- and third-year university students, Vidal and Jarvis (2018) found higher holistic ratings for the essays of third-year students based on the Common European Framework of Reference for Languages (CEFR) but no significant differences in lexical diversity between the two groups. The lack of difference in lexical production between proficiency groups was also found by Xanthou (2017), who suggests that CLIL may provide opportunities that equalizes differences between learners. Sagrario Salaberri Ramiro and Del Mar Sánchez Pérez (2015) found similar results when examining the writing of different proficiency levels of CLIL learners at the word level but found differences between groups at the

sentence and discourse levels, particularly in the areas of text format, register, and structure. This is similar to the findings of Del Mar Sanchez Perez and Sagrario Salaberri Ramiro (2015) who encourage CLIL instruction to go beyond the sentence level and address text and discourse level issues. When such levels are addressed, CLIL students have been shown over time to produce writing using the correct register for academic writing (Whittaker et al., 2011), though as Pessoa et al. (2014) note, the development of academic writing features is not always linear and often requires time.

The second group of studies analyzes CLIL writing samples for specific writing features. Coetzee-Lachmann (2007) found that CLIL geography students showed a lack of content mastery and were unable to use language appropriately as required by the genre. The author suggests that this shows the need for systematic writing instruction within CLIL content areas. Similarly, Järvinen (2010) reports a lack of awareness of genre-specific linguistic conventions and devices within CLIL writing. Kong (2010) suggests that the lack of characteristic content, audience, purpose, and genre features found in CLIL writing may be attributed to teachers' lack of understanding about writing and, subsequently, lack of instruction on writing in CLIL classrooms. Though various studies have shown CLIL students to produce writing with some genre-appropriate characteristics, many suggest that for successful content writing to be achieved in CLIL classrooms, teachers will need explicit training in the linguistic features of writing within their disciplines (Llinares & Whittaker, 2007; Llinares & Whittaker, 2010; Whittaker & Llinares, 2009). When teachers have language awareness and provide explicit instruction in the language features of academic writing, CLIL students have been shown to produce more genre-specific language features, as illustrated in Bennett (2017) with metaphor use.

51

2.5. Discussion

This systematic review has examined research on writing in CLIL with a specific focus on measurement and outcomes. The findings show that a variety of approaches to writing measurement have been used in CLIL research including measures of vocabulary, writing rubrics, complexity/accuracy/fluency measures, and content analysis. Using these methods, research findings are mixed on the effectiveness of CLIL for teaching writing. In comparison with non-CLIL groups, CLIL writing outcomes tend to be equal or higher, with the exception of a few metrics. Further, CLIL instruction has also been shown to increase scores on some writing metrics over time. However, when examined for genre- and discipline-specific writing features, evidence suggests there are various areas where the writing of CLIL students fails to exhibit content writing mastery. Below, these findings will be further discussed from the framework of the Simple View of Writing.

The methods used in the reviewed studies address both dimensions of the Simple View of Writing—transcription and ideation. However, transcription seems to have received the bulk of the attention whereas ideation seems under-explored. Transcription, which can be roughly equated to the language side of CLIL, has been addressed through a variety of measures at the word and sentence levels. Measures of CAF have explored CLIL students' abilities to transcribe at the word, clause, and sentence level in a variety of ways. Specifically, accuracy looks at the ability to transcribe without error, and fluency can roughly gauge the degree of effort a student expends on transcribing. The dimensions of language use and mechanics found on rubrics within the included studies also serve to address transcription at the word and sentence level. In terms of the discourse level, the organization dimension of the rubrics and content analysis, particularly as used in studies

conducted from a systematic functional linguistic perspective, provide measures of discourse-level transcription.

In contrast, the dimension of ideation has been given far less attention in measurement methods. Under the theoretical framework from which this study is based, ideation connects to the content objectives of CLIL. While ideation at the word, sentence, and discourse levels may account for general ideation, such as through CAF measures of complexity, CLIL ideation must also address content-specific ideation. The word level of ideation is often examined through vocabulary use. Though many studies have explored vocabulary through rubrics or vocabulary-specific measurements, only a small portion consider general academic vocabulary (Olsson, 2015; Olsson & Sylvén, 2017), and no studies seem to have examined words from discipline-specific lists. The measurement of the sentence and discourse levels of ideation within the studies also seem to measure general, rather than discipline-specific, ideation. In other words, these measures do not specifically address ideas associated with the subject matter and often only measure ideas in a general way through the content/task fulfillment dimension of the ESL Composition Profile rubric, which was designed to be used across genres of writing. Only Coetzee-Lachmann (2007), Kong (2010), and Del Mar Sanchez Perez and Sagrario Salaberri Ramiro (2015), to varying degrees, have systematically examined writing at the sentence and discourse level for discipline-specific ideas. When viewed through the Simple View of Writing, the lack of measurement in terms of ideation reveals a major gap in the investigation of CLIL writing.

When findings on CLIL writing outcomes are viewed through the Simple View of Writing, some concerns may arise about whether CLIL is currently meeting its purported

dual-focused goal. A look at results in terms of transcription generally show that CLIL provides mostly equal, but at times superior, outcomes with traditional EFL instruction. Only in a few instances were non-CLIL groups found to exceed CLIL groups in transcription accuracy (Gené-Gil et al., 2015a; Navés, 2011). Though the results are far from meeting Macaro's (2018) standard of CLIL/EMI needing to exceed traditional EFL instruction in terms of language outcomes, when considering other benefits provided by CLIL (Coyle, Hood, & Marsh, 2010), the fact that CLIL is generally not underperforming traditional EFL outcomes may be considered positive.

However, the few studies that have addressed the ideation side of writing have found less than desirable results. In general, CLIL writing has not been found to follow genre- and discipline-specific norms as one would hope nor has writing shown that CLIL students display an adequate amount of knowledge in regard to content. That being said, as mentioned above, only a few studies have actually attempted to measure the dimension of ideation past the word level, and the various studies that have looked at word-level ideation have only done so in general terms. Therefore, current research on this area is seen to be lacking and does not necessarily allow for generalizations to be made.

In sum, when viewed from the lens of the Simple View of Writing, continued work is needed in the measurement and assessment of outcomes in CLIL writing. While there seems to be adequate measurement of transcription and evaluation results for transcription show CLIL to be at least on par with traditional EFL instruction, the area of ideation remains understudied. The few studies addressing ideation have suggested that CLIL may be falling short of its goals for content learning. Therefore, future attention toward ideation in writing, both by researchers and practitioners, is needed.

2.6. Limitations

This study has a few limitations that should be considered when interpreting the results. First, as discussed earlier, this study was limited to studies written in English and, therefore, may not include studies in other languages that may have served to change the interpretation. It is encouraged that researchers from a variety of linguistic backgrounds conduct review studies inclusive of research in other languages in order to verify the interpretations of this study. Second, given the amount of quantitative studies with control groups presented in this review, a quantitative synthesis of studies using meta-analysis techniques is needed to better understand the effects of CLIL versus non-CLIL instruction.

2.7. Conclusion

In this systematic review, the Simple View of Writing has been used as a theoretical framework from which to assess writing in CLIL. Through this lens, studies on CLIL writing were examined from an integrated view of language and content. This framework has allowed for the identification of various issues in CLIL writing, particularly as it pertains to the dimension of ideation within CLIL. It is hoped that future research on CLIL writing will consider both dimensions of the Simple View of Writing—transcription and ideation—equally when measuring and assessing outcomes. Without considering both dimensions, the field will be unable to adequately assess the integration of language in content within CLIL.

2.8. References

Ackerl, C. (2007). Lexico-grammar in the essays of CLIL and non-CLIL students: Error

   analysis of written production. *Vienna English Working Papers, 16*(3), 6-11.

   Retrieved from https://anglistik.univie.ac.at/research/views/archive/

Adamson, J., & Coulson, D. (2015). Translanguaging in English academic writing

    preparation. *International Journal of Pedagogies and Learning, 10*(1), 24-37.

    https://doi.org/10.1080/22040552.2015.1084674

Agustín Llach, M. P., & Jiménez Catalán, R. M. (2007). Lexical reiteration in EFL young

    learners' essays: Does it relate to the type of instruction? *International Journal of*

    English Studies, 7(2), 85-104. Retrieved from

    https://revistas.um.es/ijes/article/view/49001

Agustín-Llach, M. P. (2016). Age and type of instruction (CLIC vs. Traditional EFL) in

    lexical development. *International Journal of English Studies, 16*(1), 75-96.

Agustín-Llach, M. P. (2017). The effects of the CLIL approach in young foreign language

    learners' lexical profiles. *International Journal of Bilingual Education and*

    Bilingualism, 20(5), 557-573. https://doi.org/10.6018/ijes/2016/1/220691

Airey, J. (2016). EAP, EMI, or CLIL. In K. Hyland & P. Shaw (Eds.), *The Routledge*

    Handbook of English for Academic Purposes. New York: Routledge.

Ball, P., Kelly, K., & Clegg, J. (2015). *Putting CLIL into practice*. New York: Oxford

    University Press.

Bennett, P. J. (2017). *The effects of raising learners' awareness of metaphorical*

    vocabulary on written production in the content-based classroom. (Unpublished

    doctoral dissertation), University of Birmingham, Birmingham, U.K.

Berninger, V. W., & Swanson, H. L. (1994). Modifying Hayes and Flower's model of

    skilled writing to explain beginning and developing writing. *Advances in Cognition*

    and Educational Practice, 2, 57-81.

Boland, A., Cherry, M. G., & Dickson, R. (2014). Carrying out a systematic review as a master's thesis. In A. Boland, M. G. Cherry, & R. Dickson (Eds.), *Doing a Systematic Review: A Student's Guide*. London: Sage Publications Ltd.

Brinton, D. M., & Snow, M. A. (2017). The evolving architecture of CBI. In M. A. Snow & D. M. Brinton (Eds.), *The content-based classroom* (pp. 2-20). Ann Arbor: University of Michigan Press.

Bruton, A. (2011). Are the differences between CLIL and non-CLIL groups in Andalusia due to CLIL? A reply to Lorenzo, Casal and Moore (2010). *Applied Linguistics, 32*(2), 236-241. https://doi.org/10.1093/applin/amr007

Camacho, A., & Alves, R. A. (2017). Fostering parental involvement in writing: Development and testing of the program Cultivating Writing. *Reading and Writing: An Interdisciplinary Journal, 30*(2), 253-277. https://doi.org/10.1007/s11145-016-9672-6

Campbell Collaboration. (2017). What is a systematic review? Retrieved from https://www.campbellcollaboration.org/research-resources/writing-a-campbell-systematic-review/systemic-review.html

Celaya, M. L. (2008). "I study natus in English": Lexical transfer in CLIL and regular learners. In R. Monroy & A. Sánchez (Eds.), *25 years of applied linguistics in Spain: Milestones and challenges* (pp. 43-49). Murica: Servicio de Publicaciones de la Universidad de Murcia.

Celestén Pérez, M., & Basse, R. (2015). *Analysing Errors of CLIL and non-CLIL Primary School Students in their Written and Oral Productions: A Comparative Study.* Paper presented at the 32nd International Conference of the Spanish Association of

Applied Linguistics (AESLA): Language Industries and Social Change, Madrid,

Spain. https://doi.org/10.1016/j.sbspro.2015.02.023

Cenoz, J., Genesee, F., & Gorter, D. (2013). Critical analysis of CLIL: Taking stock and

looking forward. *Applied Linguistics, 35*(3), 243-262.

https://doi.org/10.1093/applin/amt011

Chansri, C., & Wasanasomsithi, P. (2016). Implementing CLIL in higher education in

Thailand: The extent to which CLIL improves agricultural students' writing ability,

agricultural content, and cultural knowledge. *PASAA: Journal of Language*

Teaching and Learning in Thailand, 51, 15-38. https://doi.org/

10.14456/pasaa.2016.1

Coetzee-Lachmann, D. (2007). *Assessment of subject-specific task performance of*

bilingual geography learners: Analysing aspects of subject-specific written

discourse. (Unpublished doctoral dissertation), Osnabrück University, Osnabrück,

Germany.

Coyle, D., Hood, P., & Marsh, D. (2010). *CLIL: Content and language integrated*

learning. Cambridge: Cambridge University Press.

Dafouz, E., Camacho, M., & Urquia, E. (2014). 'Surely they can't do as well': A

comparison of business students' academic performance in English-medium and

Spanish-as-first-language-medium programmes. *Language and Education, 18*(3),

223-236. https://doi.org/10.1080/09500782.2013.808661

Dalton-Puffer, C. (2008). Outcomes and processes in Content and Language Integrated

Learning (CLIL): Current research from Europe. In W. Delanoy & L. Volkmann

(Eds.), *Future perspectives for English language teaching* (pp. 139-158).

Heidelberg: Carl Winter.

Dearden, J. (2015). *English as a medium of instruction —a growing global phenomenon*.

Retrieved from British Council: https://ora.ox.ac.uk/objects/uuid:4f72cdf8-b2eb-

4d41-a785-4a283bf6caaa

Del Mar Sanchez Perez, M., & Sagrario Salaberri Ramiro, M. (2015). Writing professional

genres in a second language: Results from a Spanish university CLIL context.

American Journal of Educational Research, 3(5), 576-580.

https://doi.org/10.12691/education-3-5-7

Gené-Gil, M., Juan-Garau, M., & Salazar-Noguera, J. (2015a). Development of EFL

writing over three years in secondary education: CLIL and non-CLIL settings.

Language Learning Journal : Journal of the Association for Language Learning,

43(3), 286-303. https://doi.org/10.1080/09571736.2015.1053278

Gené-Gil, M., Juan-Garau, M., & Salazar-Noguera, J. (2015b). Writing development under

CLIL provision. In M. Juan-Garau & J. Salazar-Noguera (Eds.), *Content-based*

language learning in multilingual educational environments (pp. 139-161). Cham:

Springer. https://doi.org/10.1007/978-3-319-11496-5_9

Goris, J. A., Denessen, E., & Verhoeven, L. (2019). Effects of content and language

integrated learning in Europe: A systematic review of longitudinal experimental

studies. *European Educational Research Journal*, 1-24.

https://doi.org/10.1177/1474904119872426

Graham, K. M., Choi, Y., Davoodi, A., Razmeh, S., & Dixon, L. Q. (2018). Language and

content outcomes of CLIL and EMI: A systematic review. *Latin American Journal*

of Content & Language Integrated Learning, 11(1), 19-37.

https://doi.org/10.5294/laclil.2018.11.1.2

Graham, K. M., & Eslami, Z. R. (2019). Attitudes toward EMI in East Asia and the Gulf:

A systematic review. *Language Problems and Language Planning, 43*(1), 8-31.

https://doi.org/10.1075/lplp.00030.gra

Jacobs, H. L., Zinkgraf, S. A., Wormuth, D. R., Hartfield, V. F., & Hughey, J. B. (1981).

Testing ESL composition. Newbury, MA: Rowley

Järvinen, H. M. (2010). Language as a meaning making resource in learning and teaching

content: Analysing historical writing in content and language integrated learning. In

C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language*

learning in CLIL classrooms (pp. 145-168). Amsterdam: John Benjamins.

https://doi.org/10.1075/aals.7.08jar

Jexenflicker, S., & Dalton-Puffer, C. (2010). The CLIL differential: Comparing the writing

of CLIL and non-CLIL students in higher colleges of technology. In C. Dalton-

Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL*

classrooms (pp. 169-189). Amsterdam: John Benjamins.

https://doi.org/10.1075/aals.7.09jex

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal

study of children in first and second grade. *Journal of Educational Psychology,*

78(4), 243-255. https://doi.org/10.1037/0022-0663.78.4.243

Kachru, B. B. (1985). Standards, codification and sociolinguistic realm: The English

language in the outer circle. In R. Quick & H. G. Widdowson (Eds.), *English in the*

world (pp. 11-30). Cambridge: Cambridge University Press.

Kjellén Simes, M. (2008). *Room for improvement?: A comparative study of Swedish learners' free written production in English in the foreign language classroom and in immersion education.* (Unpublished doctoral dissertation), Karlstad University, Karlstad, Sweden.

Kong, S. (2010). Writing in late immersion biology and history classes in Hong Kong. *Writing & Pedagogy, 2*(1), 13-38. https://doi.org/10.1558/wap.v2i1.13

Lahuerta, A. (2017). Analysis of accuracy in the writing of EFL students enrolled on CLIL and non-CLIL programmes: The impact of grade and gender. *The Language Learning Journal*, 1-12. https://doi.org/10.1080/09571736.2017.1303745

Lahuerta Martínez, A. C. (2017). Syntactic complexity in secondary-level english writing: Differences among writers enrolled on bilingual and non-bilingual programmes. *Porta Linguarum, 28*, 67-80.

Larsen-Freeman, D. (2009). Adjusting Expectations: The Study of Complexity, Accuracy, and Fluency in Second Language Acquisition. *Applied Linguistics, 30*(4), 579-589. https://doi.org/10.1093/applin/amp043

Lightbown, P. M. (2014). *Focus on content-based language teaching*. Oxford, U.K.: Oxford.

Llinares, A., & Whittaker, R. (2007). Talking and writing in a foreign language in CLIL contexts: A linguistic analysis of secondary school learners of geography and history. *Volumen Monográfico*, 83-91.

Llinares, A., & Whittaker, R. (2010). Writing and speaking in the history class: A comparative analysis of CLIL and first language contexts. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL*

classrooms (pp. 125-143). Amsterdam: John Benjamins.

https://doi.org/10.1075/aals.7.14dal

Lorenzo, F., Casal, S., & Moore, P. (2010). The effects of Content and Language

Integrated Learning in European education: Key findings from the Andalusian

Bilingual Sections Evaluation Project. *Applied Linguistics, 31*(3), 418-442.

https://doi.org/10.1093/applin/amp041

Macaro, E. (2018). *English Medium Instruction*. Oxford, U.K.: Oxford University Press.

Macaro, E., Curle, S., Pun, J., An, J., & Dearden, J. (2017). A systematic review of English

medium instruction in higher education. *Language Teaching, 51*(01), 36-76.

https://doi.org/10.1017/s0261444817000350

Manzano Vázquez, B. (2014). Lexical transfer in the written production of a CLIL Group

and a non-CLIL group. *International Journal of English Studies, 14*(2), 57-76.

https://doi.org/10.6018/j.166251

Marsh, D. (2009). Foreword. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán (Eds.),

Content and language integrated learning: Evidence from research in Europe.

Bristol Multilingual Matters. https://doi.org/10.21832/9781847691675-001

Miret Pallarols, M. (2009). *CLIL effects on writing performance and overall proficiency of*

secondary school students. (Unpublished master's thesis), University of Barcelona,

Barcelona, Spain.

Navés, T. (2011). How promising are the results of integrating content and language for

EFL writing and overall EFL proficiency? In Y. Ruiz de Zarobe, J. M. Sierra, & F.

Gallardo del Puerto (Eds.), *Content and foreign language integrated learning:*

Contributions to multilingualism in European contexts (pp. 155-186). Bern: Peter

Lang. https://doi.org/10.3726/978-3-0351-0171-3

Olsson, E. (2015). Progress in English academic vocabulary use in writing among CLIL

and non-CLIL students in Sweden. *Moderna Sprak, 109*(2), 51-74.

Olsson, E., & Sylvén, L. K. (2017). Validity in High-and Low-Stakes Tests: A Comparison

of Academic Vocabulary and some Lexical Features in CLIL and non-CLIL

Students' Written Texts. *Oslo Studies in Language, 9*(3), 127–146.

https://doi.org/10.17011/apples/urn.201512234129

Ouzzani, M., Hammady, H., Fedorowicz, Z., & Elmagarmid, A. (2016). Rayyan—a web

and mobile app for systematic reviews. *Systematic Reviews, 5*(1), 2-10.

https://doi.org/10.1186/s13643-016-0384-4

Pecorari, D., & Malmström, H. (2018). At the crossroads of TESOL and English medium

instruction. *TESOL Quarterly, 52*(3), 497-515. https://doi.org/10.1002/tesq.470

Pessoa, S., Miller, R. T., & Kaufer, D. (2014). Students' challenges and development in the

transition to academic writing at an English-medium university in Qatar.

International Review of Applied Linguistics in Language Teaching, 52(2), 127-156.

https://doi.org/10.1515/iral-2014-0006

Ritchey, K. D., & Coker, D. L. (2013). An investigation of the validity and utility of two

curriculum-based measurement writing tasks. *Reading & Writing Quarterly, 29*(1),

89-119. https://doi.org/10.1080/10573569.2013.741957

Roquet, H., & Pérez-Vidal, C. (2017). Do productive skills improve in content and

language integrated learning contexts? The case of writing. *Applied Linguistics,*

38(4), 489-511. https://doi.org/10.1093/applin/amv050

Ruiz de Zarobe, Y. (2010). Written production and CLIL: An empirical study. In C.

     Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning*

     in CLIL classrooms (pp. 191-209). Amsterdam: John Benjamins.

     https://doi.org/10.1075/aals.7.10rui

Ruiz de Zarobe, Y., & Jiménez Catalán, R. M. (2009). *Content and language integrated*

     learning: Evidence from research in Europe. Bristol: Multilingual Matters.

     https://doi.org/10.21832/9781847691675

Sagrario Salaberri Ramiro, M., & Del Mar Sánchez Pérez, M. (2015). Analyzing writing in

     english-medium instruction at university. *Linguarum Arena, 6*, 45-58.

Tai, H.-Y. (2015). Writing development in syntactic complexity, accuracy and fluency in a

     content and language integrated learning class. *International Journal of Language*

     and Linguistics, 2(3), 149-156. Retrieved from http://ijllnet.com/journal/index/2131

Tedick, D. J., & Wesely, P. M. (2015). A review of research on content-based

     foreign/second language education in US K-12 contexts. *Language, Culture and*

     Curriculum, 28(1), 25-40. https://doi.org/10.1080/07908318.2014.1000923

Uman, L. S. (2011). Systematic reviews and meta-analyses. *Journal of the Canadian*

     Academy of Child and Adolescent Psychiatry, 20(1), 57-59. Retrieved from

     https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3024725/

Vidal, K., & Jarvis, S. (2018). Effects of English-medium instruction on Spanish students'

     proficiency and lexical diversity in English. *Language Teaching Research*, 1-20.

     https://doi.org/10.1177/1362168818817945

Whittaker, R., & Llinares, A. (2009). CLIL in social science classrooms: Analysis of

     spoken and written productions. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán

(Eds.), *Content and language integrated learning: Evidence from research in Europe* (pp. 215-234). Bristol, U.K.: Multilingual Matters. https://doi.org/10.21832/9781847691675-015

Whittaker, R., Llinares, A., & McCabe, A. (2011). Written discourse development in CLIL at secondary school. *Language Teaching Research, 15*(3), 343-362. https://doi.org/10.1177/1362168811401154

Xanthou, D.-C. (2017). *The effect of CLIL on the writing and vocabulary skills of 3rd graders in the 3rd Experimental Primary school of Thessaloniki.* (Unpublished master's thesis), Aristotle University of Thessaloniki, Thessaloniki, Greece. https://doi.org/10.22540/jfsf-02-001

Yang, W. (2015). Content and Language Integrated Learning next in Asia: Evidence of learners' achievement in CLIL education from a Taiwan tertiary degree programme. *International Journal of Bilingual Education and Bilingualism, 18*(4), 361-382. https://doi.org/10.1080/13670050.2014.904840

3. MODELING THE SIMPLE VIEW OF WRITING IN CLIL WRITING

Content and Language Integrated Learning (CLIL) is a dual-focused form of language teaching which gives equal attention to both language and content outcomes (Macaro, 2018). CLIL, and its variant English Medium Instruction (EMI), has its origins in Europe, where it is widely used (Eurydice, 2006), but its spread has also been documented throughout many parts of the world (Dearden, 2015). Taiwan, specifically, is now implementing CLIL throughout its education system as a result of its recent policy proposal the *Blueprint for Developing Taiwan into a Bilingual Nation by 2030* (National Development Council, 2018).

Along with the rapid increase of CLIL around the world has been an increase in research interest in CLIL (Macaro, 2018). Specifically, scholars have been interested in whether CLIL is effective for language and content development, compared with other forms of teaching (Graham, Choi, Davoodi, Razmeh, & Dixon, 2018). In terms of language development, writing development in CLIL has received particular attention by various researchers (see Section 2), given that writing is often used to communicate academic knowledge. However, as noted by Polio and Park (2016), research on second language writing development is often "not motivated by a specific theory" (p. 290); rather, it is generally driven by individual metrics of writing and at times without regard for how these metrics fit together in a comprehensive model of writing.

This study looks to test whether the Simple View of Writing, a theory of writing that describes writing as a product through two dimensions of transcription and ideation, is adequate as a theoretical framework for CLIL research examining writing development. A

66

Confirmatory Factor Analysis (CFA) model based on the Simple View of Writing will be tested using writing data from a primary school CLIL program in Taiwan. In the following sections, theoretical orientations and research on CLIL writing will be presented. Then, details on the methods for data collection and analysis used in the study will be provided, and the results detailing the fit of the CFA models will be reported. The paper will conclude with implications and directions for future research on CLIL writing.

3.1. Theoretical Framework and Literature Review

Examinations of writing can take two approaches—writing as a product or writing as a process (Hyland, 2008). With the need to better understand the effectiveness of CLIL as an educational approach, many studies on CLIL writing take the former perspective, with concern given to the quality of writing produced by CLIL students. Cumming (2016) suggests that there are four major theories that have been influential in second language writing research—contrastive analysis, cognitive theory, genre theory, and sociocultural theories—noting that many of these theoretical traditions have histories of being borrowed and adapted from first language (L1) research. While the field of second language writing has these established theories to draw from, many times language development studies are not driven by any specific theory and usually use psychometrically-validated writing measurements to document growth (Polio & Park, 2016).

Research on writing development in CLIL generally follows this trend. While a handful of CLIL writing development studies are based in a genre theory inspired by Systemic Functional Linguistics (Järvinen, 2010; Ana Llinares & Whittaker, 2007; A. Llinares & Whittaker, 2010; Pessoa, Miller, & Kaufer, 2014; Whittaker & Llinares, 2009; Whittaker, Llinares, & McCabe, 2011), most could be described as not being driven by any

specific theory. Considering the "lack of uniform or clear theoretical perspective" in writing research (Cumming, 2016, p. 300), the lack of a guiding theory of what constitutes writing as a product makes it difficult to answer questions about CLIL's effectiveness.

3.1.1. The Simple View of Writing

One possible theory, drawn from L1 writing research, that may be able to create cohesion between the various writing measurements used in past CLIL writing research is the Simple View of Writing (Berninger & Swanson, 1994; Juel, Griffith, & Gough, 1986). From the view of this theory, the writing product can be seen as comprised of two dimensions—transcription and ideation. Transcription can be equated to the accuracy aspects of writing. In English writing, letters join to spell words, words join to form clauses, and clauses join to create written discourse. Each of these levels are governed by rules of the English language. Ideation, on the other hand, is the ideas of the writing product, that could be linked to vocabulary, task fulfillment, and complexity measures of writing. Like transcription, ideation has various levels at the word, clause, and discourse level. Words, or vocabulary/lexicon, are chosen based on the ideas to be expressed. Then, these ideas are expanded by joining with other words to form clauses which then interlink with other clauses to express written discourse. Though the Simple View of Writing provides a platform to begin thinking about writing as a product, it is anything but simple given that each level could be expanded to form exponentially more sub-levels far beyond the word, clause, and discourse level and each of these levels could be represented by numerous types of writing measurements. It is for this reason that the learning, teaching, and research of writing is a very complex matter, one that likely cannot be explained through a singular metric.

68

The Simple View of Writing as a theory for conceptualizing dimensions of writing development may fit particularly well with CLIL. Given that CLIL is "a dual educational approach where equal attention is paid to both content and language objectives" (Dafouz, Camacho, & Urquia, 2014, p. 224), a conceptual link between CLIL and the Simple View of Writing could be made. Transcription can be associated with the language aspect of CLIL while ideation could be linked to the content side. This connection provides a theoretical foundation of CLIL writing that is able to adapt to different content areas while still maintaining a consistent conceptualization of writing as a product.

3.1.2. CLIL Writing Research

Though not explicitly conceptualized as such, studies on CLIL writing have used measures connected to the Simple View of Writing's dimensions of transcription and ideation. Transcription has received much attention, usually through measures of accuracy. At the word level, several studies have examined the accuracy of spelling within writing (Ackerl, 2007; Celestén Pérez & Basse, 2015; Lahuerta, 2017). These studies generally suggest that CLIL students do not necessarily make fewer spelling errors than non-CLIL students, but the types of errors are often different. Various studies have also examined morphosyntactic errors at the clause level, usually through a ratio of clause with errors to total clauses (Gené-Gil, Juan-Garau, & Salazar-Noguera, 2015a; Lahuerta, 2017; Navés, 2011; Tai, 2015). As is the case with word-level transcription measured through spelling, studies generally show little to no difference between CLIL and non-CLIL transcription accuracy at the clause level.

Ideation within writing has also received attention in CLIL writing research, predominantly through measures of lexical diversity, a word-level ideation measure. The

studies examining lexical diversity typically use some variation of the type/token ratio

(TTR) as a representation of lexical diversity (Agustín Llach & Jiménez Catalán, 2007;

Olsson & Sylvén, 2017; Vidal & Jarvis, 2018; Xanthou, 2017). In TTR, the type refers to

the number of unique words within a piece of writing and token is the total number of

words. Overall, these studies have found no significant differences in lexical diversity in

writing between CLIL and non-CLIL groups. At the clause level, ideation can be measured

by the amount of complete ideas within a text, or in other words, the number of clauses in

a text (Navés, 2011; Tai, 2015). Navés (2011) showed that CLIL students outperformed

non-CLIL students in clause production, and Tai (2015) showed that clause production

increased over time as a result of CLIL instruction. Finally, discourse ideation has

generally been measured through the use of rubrics that assess content or task fulfillment

within a written product (Chansri & Wasanasomsithi, 2016; Miret Pallarols, 2009; Ruiz de

Zarobe, 2010). Content is generally defined by these rubrics as "development and

comprehension of the topic as well as the adequacy of the content of the text" (Ruiz de

Zarobe, 2010, p. 198). Studies comparing content for CLIL and non-CLIL writing have

shown CLIL as outperforming non-CLIL groups (Miret Pallarols, 2009; Ruiz de Zarobe,

2010). Additionally, CLIL has been shown to be influential in developing discourse-level

ideation over time (Chansri & Wasanasomsithi, 2016).

3.2. Current Study

An overall examination of CLIL writing research shows mixed results for writing

development. While both transcription and ideation have been measured in CLIL writing

studies, the measurement has often been done through singular metrics. In other words,

researchers have rarely looked at writing achievement as a composite of both transcription

and ideation within one study. Instead, the metrics within studies have been viewed as a metric in and of itself without regard to an underlying theory of writing. While some studies having shown CLIL to be superior to non-CLIL in some areas and similar in others, these measures have not yet been examined as part of a larger conceptualization of writing achievement. Acknowledging the complex nature of writing as a product and the need to evaluate the effectiveness of CLIL as an educational approach, a theory-driven approach which accounts for the various aspects of writing may be needed in CLIL research. To address this, the current study looks to examine and validate a CLIL writing model comprised of various metrics used in previous CLIL writing studies that relate to the dimensions of the Simple View of Writing: transcription and ideation. The following research questions will be examined through the comparison of two Confirmatory Factor Analysis (CFA) models:

1. Does a single-factor model for CLIL narrative and expository writing adequately fit elementary CLIL writing?

2. To what degree does a higher-order factor model with CLIL narrative and expository writing as second-order factors and transcription and ideation as first-order factors explain elementary CLIL writing?

3. How does the above higher-order factor model of CLIL writing compare with a single-factor CLIL model without transcription and ideation?

3.3. Methods

3.3.1. Participants

This study collected writing samples from 212 primary school students enrolled in a CLIL program at a private primary school in northern Taiwan. The primary school

teaches courses both in Chinese, following Taiwan's Ministry of Education requirements, and in English based on the principles of CLIL. In regard to the CLIL program, students attend 17 forty-minute periods per week of English CLIL courses (Table 3.1)

Table 3.1 Allotment of CLIL Courses

Course	Number of Periods
English/language arts	10
Science/social studies	3
Math	2
Physical education	1
Enrichment course	1
Total:	17

CLIL courses include English/language arts (ELA), science/social studies, math, physical education, and an enrichment course. ELA and science/social studies roughly follow the United States' curriculum standards and use textbooks imported from the U.S. Math is the only subject that mirrors the curriculum goals of the Chinese classes and uses a textbook published in Singapore, along with supplemental material. Both physical education and the enrichment course have an open curriculum, allowing individual teachers to decide athletic and hands-on (e.g., arts and crafts, project-based activities) activities for each class.

This study collected data at the beginning of the academic year from primary school students entering grades two through six. Students in the CLIL program are divided into English classes based on school-created English proficiency tests. Each grade has three levels of classes: lower, middle, and upper. One upper-level proficiency and one middle-level proficiency class was randomly selected from each grade. However, due to

unforeseen circumstances within the school, data from the randomly selected upper-level

proficiency grade five class was unable to be collected, meaning only middle-level

proficiency grade five writing samples are included in the data set. With the exception of

grade six, lower-level proficiency classes were excluded from random selection. School

administrators expressed concern that lower-level proficiency classes may not be able to

complete the writing assessment due to their low English proficiency and academic

achievement levels, ultimately resulting in undue anxiety and stress. Thus, the

administration requested that lower-level proficiency classes in grades two through five be

excluded from data collection. However, administrators allowed the randomly selected

lower-level proficiency sixth grade class to take the writing assessment, believing that

concerns of anxiety and stress would not be an issue for the students at this age. Table 3.2

reports tabulations of students by grade included in the study.

Table 3.2 Number of Participants by Grade Level

Grade Level	Number of Participants
Second	53
Third	44
Fourth	41
Fifth	18
Sixth	56
Total:	212

3.3.2. Writing Assessment

Each student was administered two writing assessments: one with an expository

prompt and the other with a narrative prompt. These two genres were chosen based on the

developmental level of the students in addition to consideration of the types of writing students engage in through the CLIL program. Both writing samples were administered using similar procedures but were collected at separate times. Writing assessments were administered by the respective classroom teachers, who were trained by the lead researcher. During training, the teachers were provided a script for test administration and procedures were discussed. During each writing assessment, students were given three minutes to plan the writing through drawing, outlining, or any means that was most helpful to them. Following this planning period, students were given five minutes to write about the prompt. A minute before the conclusion of the assessment, students were instructed that one minute was remaining, and they were asked to complete the writing task. The choice of restricting the writing time to five minutes was made based on age and language-level considerations. McMaster et al. (2017) recommend five minutes for young learner writing assessments. While students in grade six may have been able to write continuously beyond the five-minute period, students in second grade, particularly those of lower language proficiency, likely could not. Thus, it was decided to limit the time to five minutes, which would both allow for an adequate sample of writing to be produced for analysis and, at the same time, allow comparability across grade levels.

During the administration of the writing assessment, students were provided prompts, both orally and in writing, for each of the assessments. Prompts were reviewed by the school's teachers and administrators before the assessment was given and consensus was reached that both prompts would be appropriate for the students in the CLIL program. The narrative writing assessment asked students to write a story about a child who broke the teacher's glasses (Camacho & Alves, 2017). The expository writing assessment asked

students to write about something they learned about in science class during the previous year. Whittaker and Llinares (2009) emphasize the importance of using writing assessments with "tasks for which the students were prepared, on topics they enjoyed, to make sure that there would be sufficient production" (p. 219). With this in mind, the open-ended nature of the expository prompt was intended to allow students to choose a topic in science they were comfortable with and enjoyed rather than restricting them to a specific topic where they may lack the appropriate knowledge or motivation to complete the task. This also allowed for the same prompt to be used across grades where different science topics are taught in each grade level.

3.3.3. Writing Measures

Each writing sample for both genres was assessed using measures of transcription and ideation. Prior to scoring, all writing samples were transcribed as written by the participants (i.e., with mistakes) into a computer-readable form. The purpose of this transcription was to allow for computerized analyses to be conducted. The writing samples were originally typed into text-file format (.txt) by trained research assistants. These files were then transferred into CLAN (MacWhinney, 2000), a software designed for linguistic analysis of speech and writing, using the TXT2CLAN command in CLAN. This software was chosen as it has been used to measure writing in previous CLIL studies (Artieda, Roquet, & Nicolás-Conesa, 2017; Gené-Gil, Juan-Garau, & Salazar-Noguera, 2015a; Gené-Gil et al., 2015b). This command takes the .txt files and automatically adds all of the appropriate headings needed for CLAN files to run.

Coding of the writing samples within CLAN was conducted in various stages. First, the writing samples were divided into separate lines by t-units, or the shortest grammatical

sentence, delimited by speaker tags with capitalization and punctuation removed, an

exception being the concluding period as required by CLAN. Misspelled words were then

coded using the following tag after the misspelled word: [: *correctly-spelled word*] [*]. The

first bracket notifies CLAN of the attempted word, and the second marks the error for error

frequency counts. Following this coding, the MOR command in CLAN was run for all

files, which automatically conducts a morphological analysis of the text and adds

morphological coding. This step is needed for subsequent analysis programs to run in

CLAN. Finally, the EVAL command in CLAN was run, which produces a data table used

for calculating various writing scoring metrics.

3.3.3.1. Spelling

Spelling represents a measure of word-level transcription. The spelling score is the

number of correctly spelled words divided by total words. Words misspelled were marked

within CLAN using [*]. Words were considered misspelled if the word contained any type

of orthographical error (e.g., omission, substitution, transposition, etc.). Grammatical or

semantic errors, if committed with a correctly spelled word, were not considered as

spelling errors in this score. Based on the [*] coding within CLAN, the EVAL data table

provides the number of word errors. The number of correctly spelled words was then

calculated in Microsoft Excel by subtracting the number of spelling errors from total words

and then dividing by total words.

3.3.3.2. Correct Word Sequences

The correct word sequences score is a morphosyntactic transcription measure. The

score represents the accuracy of adjacent words within a writing sample. Only

morphosyntactic accuracy was assessed in this score, with spelling, capitalization, and

punctuation not considered. Scoring begins with examination of the first word alone. If this word is a correct first word for the writing sample, considering context, one point is awarded. The first and second words are then evaluated for the appropriateness of their adjacency. For example, if the two words are "three cat," no points are awarded because the second word is not in its appropriate plural form. Each word sequence is evaluated until the last written word. As with the first word, if the last word is an accurate final word, one final point is awarded. This approach to writing scoring has been evaluated in previous research and deemed an appropriate method of writing assessment scoring for elementary students (McMaster et al., 2017). The final correct word sequences score is calculated as number of correct word sequences divided by total words plus one. Writing samples were coded for correct word sequences by the lead researcher by adding "CWS" where correct word sequences appeared within a separate CLAN file created specifically for calculating correct word sequences. The separate file was used in order to prevent CLAN from calculating the "CWS" markings in other scores. The final score was tabulated using the FREQ command in CLAN.

3.3.3.3. Lexical Diversity

Lexical diversity is a measure of unique English vocabulary words used within a text and is considered a word-level ideation measure. This has been traditionally done with a type/token ratio (TTR). However, given the varied length of texts included in this study, the traditional type/token ratio could not be used in light of its sensitivity to word count (MacWhinney, 2019). One solution for this has been proposed by Carroll (1964) where TTR is calculated as types divided by the square root of tokens multiplied by two. This study used this calculation for lexical diversity using the type and token tabulations from

77

the CLAN EVAL data table. All attempted words, regardless of misspelling, were considered. CLAN is able to tabulate misspelled words by using the [: *correctly-spelled word*] code, which tells the program to disregard the preceding misspelled word in its tabulation.

3.3.3.4. Total T-Units

Total t-units measures clause-level ideation. The basis for this is that a single t-unit represents a complete idea given that a t-unit is defined as a unit of text that can stand alone as a sentence. As explained above, each text was divided into t-units, demarcated by the speaker tag in CLAN. This marks t-units as utterances (Utts) in the CLAN file, and the number of t-units can be found in the Total_Utts column of the EVAL table.

3.3.3.5. Holistic Rubric for Ideas

The holistic rubric for ideas measures discourse-level ideation. This was scored by human raters using an adapted version of the ideas rubric from Education Northwest (2013). This rubric has been successfully adapted and used in previous research studies of young learner writing (Kim, Al Otaiba, & Wanzek, 2015; Kim, Al Otaiba, Wanzek, & Gatlin, 2015; Kim, Gatlin, Al Otaiba, & Wanzek, 2017; Kim & Schatschneider, 2017), thus suggesting its appropriateness for the age group being studied. The rubric has a range of six points. A score of one indicates a student who attempted to express ideas through only pictures or through incomprehensible symbols, thus not expressing textual ideas through text. A two-point score indicates that the student has developed print sense, producing letters of the English alphabet or words but not producing a complete idea. A score of three indicates the student has begun to produce a full idea in the form of one complete t-unit or repeats the same idea through multiple t-units. A four-point score indicates a piece of writing that has

78

gone beyond one t-unit, but the ideas are not focused and lack cohesion, going from one topic to the next or simply listing events. A writing sample scored as a five indicates a text that is focused on one main idea and presents details that develop the idea in a cohesive manner. Finally, a six-point score exhibits the characteristics of the five-point score, but the details are well-developed and mature. In a narrative, this would mean the use of sensory details and figurative language. In an expository text, this would be achieved thorough descriptions and examples that exemplify the topic well.

The lead researcher and a Ph.D. student met together to calibrate on the rubric. During this initial session, the lead researcher explained the rubric to the graduate student and showed writing examples of each score level. The lead researcher and graduate student then practiced scoring samples independently and the given scores were compared and discussed. Once both raters felt they had reached an adequate understanding of how to apply the rubric during the practice session, a stratified random sample by grade and class of 20% (n = 48/genre) was independently scored for both genres. Given the ordinal nature of the rubric, a quadratic weighted kappa was calculated to determine inter-rater reliability using STATA 16 (StataCorp, 2019). The weighted kappa for the narrative sample is κ = .82, and for the expository sample the weighted kappa is κ = .83, suggesting a high inter-rater reliability. Discrepancies for the 20% sample were discussed and resolved between the two raters. Then, the remaining 80% was scored independently by the lead researcher.

3.3.4. Data Analysis

Descriptive statistics and Confirmatory Factor Analysis (CFA) were examined using MPlus 8.4 (Muthén & Muthén, 2017). Given the non-normality of data in the sample, a robust maximum likelihood estimator using maximum likelihood parameter

estimates with standard errors and the Satorra-Bentler chi-square was used (MLM in MPlus). The initial base model is comprised of two covarying latent variables: CLIL Narrative and CLIL Expository. Each of these latent variables have five indicators, respectively: spelling, correct word sequences, lexical diversity, t-units, and the ideas rubric. The second model testing the Simple View of Writing is a higher-order model with CLIL Narrative and CLIL Expository as covarying second-order latent variables, each with first-order latent variables for transcription and ideation. The transcription latent variables have two indicators: spelling and correct word sequences. The ideation latent variables have three indicators: lexical diversity, t-units, and the ideas rubric. By examining two models in this study, the base model and Simple View of Writing model, comparisons can be made between the two models to assess if true differences in model fit exist and determine whether or not the model driven by the Simple View of Writing theory can more adequately explain the data than the base model.

The models were assessed and compared for model fit based on chi-square statistics as well as the following global fit indices: the root mean square error of approximation (RMSEA), the standardized root mean square residual (SRMR), and the comparative fit index (CFI). For CFA models, a model exhibiting a good fit with the data would return a chi-square test with a p-value above .05, RMSEA values below .06, SRMR values below .08, and a CFI values above .95 (Hu & Bentler, 1999). To test statistical differences between nested models, the Satorra-Bentler Scaled Chi-Square Difference Test for testing model differences was used (Satorra & Bentler, 2010). A difference test returning a p-value below .05 indicates significant differences between the models. To test statistical differences between non-nested models, Akaike information criterion (AIC) and

80

Bayesian information criterion (BIC) statistics were calculated. A lower AIC and BIC

statistic indicates a better model fit.

3.4. Results

3.4.1. Descriptive Statistics

Table 3.3 presents descriptive statistics for each of the writing measurements by

genre. For all measurements, participants scored higher on the narrative writing sample

than on the expository sample.

Table 3.3 Descriptive Statistics for Writing Measures

	Narrative						Expository					
	M	*SD*	*Sk*	*Ku*	*Min*	*Max*	*M*	*SD*	*Sk*	*Ku*	*Min*	*Max*
1	0.94	0.01	-3.00	13.61	0.33	1.00	0.90	0.03	-3.49	14.69	0.00	1.00
2	0.60	0.03	-0.67	0.91	0	0.97	0.58	0.05	-1.04	0.91	0.00	0.95
3	2.76	0.45	-0.15	-0.05	0.82	4.60	2.71	0.60	-0.96	1.67	0.00	4.52
4	5.60	9.15	0.66	0.26	1.00	16.00	4.14	6.15	0.64	0.01	0.00	12.00
5	3.97	0.74	-0.29	0.79	1.00	6.00	2.37	0.72	2.25	4.24	1.00	6.00

Note. 1 = Percentage Spelling Accuracy; 2 = Percentage Correct Word Sequences; 3 = Lexical Diversity; 4 = Total T-Units; 5 = Ideas Rubric.

For spelling, participants on average spelled a high percentage of words correctly in both

genres, scoring 94% on narrative and 90% on expository. For correct word sequences, the

participants showed less accuracy than with spelling, averaging 60% on narrative and 58%

on expository. The lexical diversity in both genres was fairly close, with participants

having an average score of 2.76 and 2.71 on narrative and expository, respectively. The

average number of t-units was 5.60 in narrative and 4.14 in expository. The high standard

deviations for the t-unit score should also be considered when interpreting these means

(narrative = 9.15; expository = 6.15) as they show a great variation within the participants

that is far higher than seen in other measurements, likely reflecting greater variation

between grade levels. Finally, the average score for the ideas rubric was 3.97 for narrative

and 2.37 for expository. Referencing the rubric criteria for interpretation, this shows that

participants were on average able to produce a narrative with a clear big idea and simple

explanation. In contrast, the participants on average were beginning to present ideas in the

expository text but with no development. The ideas rubric scores seemingly show the

largest gap for participants between the two genres of all the measures.

     A correlation matrix for each of the variables in the model is shown in Table 3.4.

Table 3.4 Correlation Matrix of Writing Measures

	N1	N2	N3	N4	N5	E1	E2	E3	E4	E5
N1	1.00									
N2	.355	1.00								
N4	.295	.487	1.00							
N5	.220	.326	.644	1.00						
N6	.252	.412	.722	.740	1.00					
E1	.272	.350	.345	.335	.356	1.00				
E2	.360	.605	.490	.418	.487	.469	1.00			
E4	.307	.536	.622	.529	.579	.553	.629	1.00		
E5	.250	.406	.538	.596	.562	.335	.483	.694	1.00	
E6	.047	.191	.153	.194	.179	.129	.190	.213	.310	1.00

Note. N1 = Narrative Percentage Spelling Accuracy; N2 = Narrative Percentage Correct Word Sequences; N3 = Narrative Lexical Diversity; N4 = Narrative Total T-Units; N5 = Narrative Ideas Rubric; E1 = Expository Percentage Spelling Accuracy; E2 = Expository Percentage Correct Word Sequences; E3 = Expository Lexical Diversity; E4 = Expository Total T-Units; E5 = Expository Ideas Rubric.

An examination of the correlation matrix reveals that the writing measurements in this

study are moderately to highly correlated with each other except for a few exceptions.

Weak correlations are found between narrative spelling and most other variables except

with correct word sequences. The same holds true for expository spelling, though it is moderately correlated with other variables in the expository genre, with the ideas rubric as an exception. Interestingly, the expository ideas rubric revealed a weak correlation with all other variables except expository t-units, though this was not the case for the narrative ideas rubric which has strong correlations with other metrics.

3.4.2. Confirmatory Factor Analysis

Four CFA models with fit statistics are reported in Table 3.5. The initial model run was a base model with two latent variables, one for narrative and one for expository, each with five measured variables. The Chi-Square Test of Model Fit indicates the model lacks a perfect fit with the data, $\chi_2(34) = 112.697$, $p < .001$. The Root Mean Square Error of Approximation (RMSEA) for this model is .104 and the Standardized Root Mean Square Residual (SRMR) is .064. The Comparative Fit Index (CFI) for the model is .897. Together, these statistics suggest a poor model fit.

Table 3.5 Fit Statistics for CFA Models

Statistic	Model			
	Base 1	Base 2	SVW 1	SVM 2
χ^2_M	112.697	87.304	57.155	44.977
df_M	34	33	29	28
p	< .001	< .001	.001	.02
RMSEA [90%CI]	.104 [.083, .126]	.088 [.066, .111]	.068 [.041, .093]	.053 [.021, .081]
CFI	.897	.929	.963	.978
SRMR	.064	.057	.038	.034
AIC	2423.663	2396.990	2370.736	2358.552
BIC	2527.717	2504.400	2491.573	2482.745

Initially, the analysis proceeded to testing the Simple View of Writing model. However, the psi matrix of this model was found to be not positive definitive. It was hypothesized that this may be a result of covariance between error terms for measurements used across genres. Modification indices for the base model were examined for possible error term covariances. Results suggest that adding a covariance between narrative and expository correct word sequences, which would result in a 23.637 drop in the chi-square value. This change was found to correct the psi matrix issue in the higher-order Simple View of Writing model, and results of this modification are reported below.

Considering the modification indices result from above, a second base model was run with a covariance between the narrative and expository correct word sequences error terms. This model's Chi-Square Test of Model Fit also showed a lack of perfect fit with the data, $\chi_2(33) = 87.304$, $p < .001$. The model's RMSEA is .088, the SRMR is .057, and the CFI is .929, which together show a poor model fit.

Next, the Simple View of Writing model (SVW Model 1) was tested. This model is a higher-order factor model with two second-order factors, CLIL Narrative and CLIL Expository. Each second-order factor has two single-order factors representing transcription and ideation for the respective genre. There are two measured variables for the transcription factors, spelling and correct word sequences, and there are three for the ideation factors, lexical diversity, total t-units, and the ideas rubric. The Chi-Square Test of Model Fit indicates this model lacks a perfect fit with the data, $\chi_2(29) = 57.155$, $p = .001$. The RMSEA for the model is .068, the SRMR is .038, and the CFI is .963. These fit indices indicate a fair fit for the model and data. This model was compared with the second base model and was found to have lower values for AIC and BIC (Base 2: AIC =2396.990,

BIC = 2504.400; SVW Model: AIC = 2370.736. BIC = 2491.573), suggesting the SVW

model as a better model for the data. To further examine this model, modification indices

were investigated and revealed that adding an additional covariance between the error

terms for narrative and expository total t-units would result in a 12.585 drop in the chi-

square value. Based on these results, a second Simple View of Writing model was tested

with this added covariance as it is theoretically conceivable that using the same writing

measurement across genres may result in similar unexplained error.

Analysis of the second Simple View of Writing model (SVW Model 2) with the

additional covariance between t-unit error terms showed a lack of a perfect fit based on the

Chi-Square Test of Model Fit, $\chi_2(28) = 44.977$, $p = .02$. The RMSEA is .053, the SRMR

is .034, and the CFI is .978. These statistics suggest that this model may have a good fit

with the data. In order to test whether there are significant differences between this and

SVW Model 1, the Satorra-Bentler Scaled Chi-Square Difference Test was calculated and

showed a statistically significant difference between the two models, $\chi_2(1) = 8.759$, p

$= .003$. An examination of the modification indices in this model did not reveal any

additional changes to be made that would fit theoretically.

Considering the improved fit statistics, the statistically significant difference

between models, and the lack of additional modifications to be made, SVW Model 2 was

chosen as the final model. Differences between the AIC and BIC statistics for second base

model and SVW Model 2 show lower values for both statistics (Base 2: AIC = 2396.990,

BIC = 2504.400; SVW Model: AIC = 2358.552. BIC = 2482.745), thus suggesting that

the Simple View of Writing better explains the data than the base model.

Figure 3.1 shows a path diagram with standard coefficients for SVW Model 2. When examining standardized coefficients in CFA models, Kline (2016) recommends standardized coefficients above $\beta = .5$. This threshold is met for all paths from the second-order to first-order factors. However, there are two observed variables in this model that fall below this threshold: narrative spelling ($\beta = .456$) and the expository ideas rubric ($\beta = .265$). This suggests that the factors narrative transcription and expository ideation do not account for a substantial amount of the variance in narrative spelling ($R_2 = 20.8\%$) and the expository ideas rubric ($R_2 = 7.0\%$), thus raising questions about their inclusion in the model. However, the lack of variation accounted for in these variables does not hold true for their respective genre counterparts, where expository transcription accounts for 38.3% of variance in expository spelling and narrative ideation accounts for 77.9% of the variance in the narrative ideas rubric. With the exception of narrative spelling and expository ideas rubric, all other observed variables have coefficients above the $\beta = .5$ threshold.
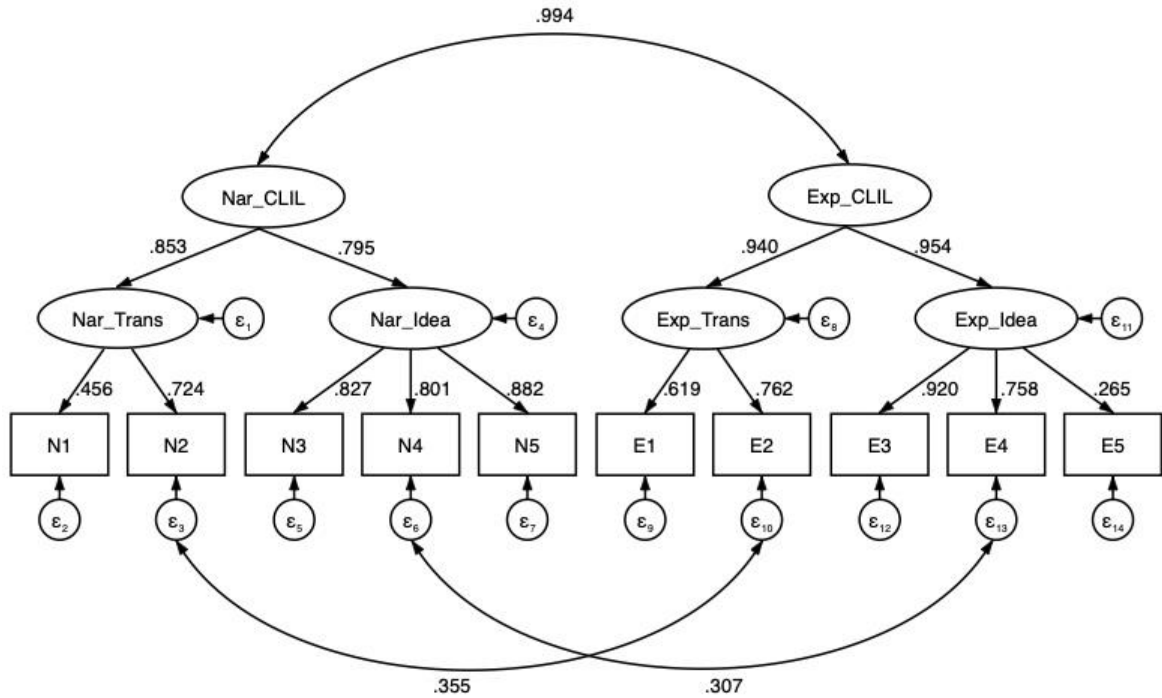
Figure 3.1 Final Simple View of Writing path model with standardized coefficients.

3.5. Discussion

This study used Confirmatory Factor Analysis (CFA) to validate the Simple View of Writing as a theory for evaluating writing as a product in elementary CLIL. The first question asked whether a single-factor model for narrative and expository writing adequately explained CLIL elementary writing data. The analysis suggests that this single-factor model has a poor fit with the data. The second question examined a higher-order factor model driven by the Simple View of Writing. Results show that this model has a good fit with the data. Finally, this study asked whether the Simple View of Writing Model can statistically fit the data better than the single-factor model and found that the Simple

View of Writing Model statistically explains the data more adequately. These models and their implications for further research are discussed below.

As Polio and Park (2016) note, second language writing research that focuses on language development, as much of CLIL writing research does, often is not guided by any specific writing theory, putting the focus on measures. CLIL writing is predominantly explored through individual measures of complexity, accuracy, or through other scoring measures such as rubrics. Using these measurement categories from previous CLIL writing research, the base model examined in this study suggests that considering these measures as individually independent measures of CLIL writing may not be adequate for explaining writing as a product for elementary CLIL students. This may suggest that a more theoretically driven approach may be needed for conceptualizing these measures in CLIL writing development research.

Unlike the base model, the model driven by the Simple View of Writing theory showed a good fit with the data. These results possibly suggest that the Simple View of Writing may theoretically help us better understand and explain CLIL writing as a product for elementary CLIL students when compared with past approaches that were absent of theory. For researchers, the theory provides a framework for measuring writing development in CLIL, putting forth that CLIL development should be measured both in terms of transcription and ideation. This framework can help researchers in ensuring that the measurements chosen for their study adequately address both dimensions and that no study focuses on one of these dimensions exclusively. It also provides a theoretical basis for interpreting results. Rather than examining each metric individually, development could be assessed in terms of transcription growth and ideation growth. By doing so, CLIL

development can be examined more comprehensively. For CLIL practitioners, the Simple View of Writing also provides a guide for CLIL writing instruction. With the principles of CLIL based on an even and integrated balance between language and content, the Simple View of Writing helps drive this balance by conceptualizing writing in terms of transcription (language) and ideation (content).

While the analysis in this study showed a good fit for the model using the Simple View of Writing, a further examination of individual coefficients revealed that not all measures were adequately explained by the transcription and ideation factors and that there were differences between genres. First, the coefficient for narrative spelling was just below Kline's (2016) $\beta = .5$ threshold, measuring at $\beta = .456$. Previous L1 and L2 writing studies have equated transcription exclusively with spelling ability, though sometimes also adding handwriting ability for some younger age groups (Harrison et al., 2016; Kim, Al Otaiba, Wanzek, et al., 2015). With the narrative spelling coefficient below the threshold but the expository spelling coefficient above ($\beta = .619$), spelling as a measure of transcription cannot be completely ruled out. However, its use as a measure for narrative transcription in CLIL elementary narrative writing should be further considered.

There are few possible explanations for why spelling was better explained by the transcription factor in expository writing over narrative writing. Returning to the descriptive statistics, the participants on average spelled 94% of words correctly in their narrative with a slim standard deviation of 1%. In other words, participants generally were very accurate with spelling in their narrative samples with very little variation in spelling performance between participants. Comparing this with expository writing where the average was 90%, which may still be considered high, we see much more variation

between participants with a standard deviation of 3%. This may signal that genre may have an impact on the way the dimensions of the Simple View of Writing should be measured, thus suggesting a need for considering genre theory alongside the Simple View of Writing in future studies. When comparing the type of words needed in narrative and expository writing, it is conceivable that expository would require higher level words, particularly in the case of this study where students were writing about science concepts. This may lead to more spelling errors and subsequently more variation, making for a better indicator of transcription. Narrative writing, on the other hand, may involve words that are more familiar and less technical, resulting in higher accuracy and, subsequently, less variation, weakening its usability as an indicator. In a more homogeneous sample including a single grade level, spelling's strength as an indicator of narrative transcription may change. It is likely that a sample exclusively of lower elementary students may show spelling to be a stronger indicator, whereas one with upper elementary or older may produce an even weaker coefficient. This aligns with Kim et al. (2017) who have suggested that as transcription skills strengthen, its importance in writing diminishes and ideation takes on a stronger role. This could be further extended by saying once word-level transcription ability stabilizes, higher level transcription skills at the clause or discourse level may become more important. However, further empirical research would be needed to substantiate this claim.

The coefficient for the expository ideas rubric also was below the $\beta = .5$ threshold with a measure of only $\beta = .265$. This quite a bit lower than the violation observed for the narrative spelling coefficient, raising more concerns. Like with spelling, the rubric produced different results between genres, again highlighting the possible need of genre

theory in coordination with the Simple View of Writing. The narrative ideas rubric has a coefficient well above $\beta = .5$, measuring at $\beta = .882$. There are two possible reasons for such a large discrepancy—one developmental and one psychometric. Similar to the explanation above, participant writing performance may be the reason for the disparity in the rubric scores as an indicator between genres. It was noted in the results section that the gap between averages for the narrative and expository ideas rubric appeared to be much wider than in other measures. The expository ideas rubric score indicated that participants, on average, lacked the ability to go beyond presenting simple ideas in expository writing, compared with the much higher average for the narrative samples with clear ideas and some description. Given the large difference in discourse-level ability as a result of genre, the elementary CLIL participants as a group may not have strong enough abilities in discourse-level ideation for this to be a good indicator of writing at this level. The stronger coefficients for word- and clause-level ideation may also support the case for this explanation. This may also suggest the need for CLIL teachers to dedicate time toward teaching academic writing genres in their CLIL classrooms, even at the elementary level, a sentiment expressed by other scholars (Whittaker et al., 2011).

Another explanation for the discrepancy in coefficients between genres for the ideas rubric is one of psychometric validity. Though Education Northwest (2013) purports the rubric is designed for use across genres and the rubric has been used in previous studies, this is the first study that has used this rubric with writing in an elementary CLIL setting. While the model's results suggest that its use may be valid for evaluating narrative writing for CLIL elementary students, the validity of its use for expository is in question. Given the developmental explanation presented above, this study does not necessarily rule

out the rubric's validity, but it does suggest the need for a validation study for the rubric's use in CLIL expository and other types of writing.

3.6. Conclusion

This study has shown that the Simple View of Writing, comprised of the dimensions of transcription and ideation, may be useful as a theoretical basis for investigating and understanding CLIL writing development. However, there are a few limitations to this study that should be considered. First, though the field of second language writing has a history of drawing theories from L1 writing research (Cumming, 2016), differences between L1 and L2 writing should be considered and appropriate adaptations to the theory should be made. As noted by Silva (1993), while there are certain similarities between L1 and L2 writing, there are also differences that should be acknowledged. Second, caution should be taken when applying the findings of these results to other populations. This study uses a sample of Taiwanese elementary CLIL students in an intensive CLIL program. Validation of this theory for other populations, particularly older CLIL populations, is needed before applying it broadly to CLIL research. Finally, further consideration of the types of indicators for each dimension, especially when considering different genres, is necessary. With some indicators not meeting the appropriate threshold for coefficients in this study, more research is needed to understand the possible underlying cause of this and the theoretical considerations for choosing indicators for each of the dimensions of transcription and ideation. Genre theory and other language development theories in combination with the Simple View of Writing may help lead the way toward more clarity in this matter.

Despite these limitations, the Simple View of Writing seems to provide an adequate basis for uniting language development studies in CLIL and ensuring comprehensive evaluation of writing development. When combined with genre theory, the Simple View of Writing provides a framework with enough flexibility to accommodate different writing genres by allowing for various indicators of transcription and ideation to be used while at the same time providing enough guidance for researchers and practitioners in addressing the complex dimensions of writing that can produce results comparable across studies and populations. It is recommended that CLIL writing studies focusing on writing development consider the use of the Simple View of Writing as a theoretical basis in future investigations.

3.7. References

Ackerl, C. (2007). Lexico-grammar in the essays of CLIL and non-CLIL students: Error analysis of written production. *Vienna English Working Papers, 16*(3), 6-11. Retrieved from https://anglistik.univie.ac.at/research/views/archive/

Agustín Llach, M. P., & Jiménez Catalán, R. M. (2007). Lexical reiteration in EFL young learners' essays: Does it relate to the type of instruction? *International Journal of English Studies, 7*(2), 85-104. Retrieved from https://revistas.um.es/ijes/article/view/49001

Artieda, G., Roquet, H., & Nicolás-Conesa, F. (2017). The impact of age and exposure on EFL achievement in two learning contexts: formal instruction and formal instruction + content and language integrated learning (CLIL). *International Journal of Bilingual Education and Bilingualism*, 1-19. https://doi.org/10.1080/13670050.2017.1373059

Berninger, V. W., & Swanson, H. L. (1994). Modifying Hayes and Flower's model of

skilled writing to explain beginning and developing writing. *Advances in Cognition*

and Educational Practice, 2, 57-81.

Camacho, A., & Alves, R. A. (2017). Fostering parental involvement in writing:

Development and testing of the program Cultivating Writing. *Reading and Writing:*

An Interdisciplinary Journal, 30(2), 253-277. https://doi.org/10.1007/s11145-016-

9672-6

Carroll, J. B. (1964). *Language and thought.* Engelwood Cliffs, NJ: Prentice Hall.

Celestén Pérez, M., & Basse, R. (2015). Analysing Errors of CLIL and non-CLIL Primary

School Students in their Written and Oral Productions: A Comparative Study.

Paper presented at the 32nd International Conference of the Spanish Association of

Applied Linguistics (AESLA): Language Industries and Social Change, Madrid,

Spain. https://doi.org/10.1016/j.sbspro.2015.02.023

Chansri, C., & Wasanasomsithi, P. (2016). Implementing CLIL in higher education in

Thailand: The extent to which CLIL improves agricultural students' writing ability,

agricultural content, and cultural knowledge. *PASAA: Journal of Language*

Teaching and Learning in Thailand, 51, 15-38. Retrieved from

https://eric.ed.gov/?id=EJ1112239

Cumming, A. (2016). Theoretical orientations to L2 writing. In R. M. Machón & P. K.

Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 65-88).

Boston: De Gruyter. https://doi.org/10.1515/9781614511335-006

Dafouz, E., Camacho, M., & Urquia, E. (2014). 'Surely they can't do as well': A

comparison of business students' academic performance in English-medium and

Spanish-as-first-language-medium programmes. *Language and Education, 18*(3),

223-236. https://doi.org/10.1080/09500782.2013.808661

Dearden, J. (2015). English as a medium of instruction —a growing global phenomenon.

Retrieved from British Council: https://ora.ox.ac.uk/objects/uuid:4f72cdf8-b2eb-

4d41-a785-4a283bf6caaa

Education Northwest. (2013). Grades K–2 Traits Rubric for Ideas. Retrieved from

https://educationnorthwest.org/sites/default/files/gradesK-2-6pt-rubric.pdf

Eurydice. (2006). Content and language integrated learning (CLIL) at school in Europe.

Retrieved from Brussels:

http://www.indire.it/lucabas/lkmw_file/eurydice/CLIL_EN.pdf

Gené-Gil, M., Juan-Garau, M., & Salazar-Noguera, J. (2015a). Development of EFL

writing over three years in secondary education: CLIL and non-CLIL settings.

*Language Learning Journal: Journal of the Association for Language Learning,
43*(3), 286-303. https://doi.org/10.1080/09571736.2015.1053278

Gené-Gil, M., Juan-Garau, M., & Salazar-Noguera, J. (2015b). Writing development under

CLIL provision. In M. Juan-Garau & J. Salazar-Noguera (Eds.), *Content-based
language learning in multilingual educational environments* (pp. 139-161). Cham:

Springer. https://doi.org/10.1007/978-3-319-11496-5_9

Graham, K. M., Choi, Y., Davoodi, A., Razmeh, S., & Dixon, L. Q. (2018). Language and

content outcomes of CLIL and EMI: A systematic review. *Latin American Journal
of Content & Language Integrated Learning, 11*(1), 19-37.

https://doi.org/10.5294/laclil.2018.11.1.2

Harrison, G. L., Goegan, L. D., Jalbert, R., McManus, K., Sinclair, K., & Spurling, J. (2016). Predictors of spelling and writing skills in first- and second-language learners. *Reading and Writing, 29*, 69-89. https://doi.org/10.1007/s11145-015-9580-1

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

Hyland, K. (2008). Writing theories and writing pedagogies. *Indonesian JELT, 4*(2), 1-20. http://dx.doi.org/10.25170%2Fijelt.v4i2.145

Järvinen, H. M. (2010). Language as a meaning making resource in learning and teaching content: Analysing historical writing in content and language integrated learning. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 145-168). Amsterdam: John Benjamins. https://doi.org/10.1075/aals.7.08jar

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78*(4), 243-255. https://doi.org/10.1037/0022-0663.78.4.243

Kim, Y.-S., Al Otaiba, S., & Wanzek, J. (2015). Kindergarten predictors of third grade writing. *Learning and Individual Differences, 37*, 27-37. https://doi.org/10.1016/j.lindif.2014.11.009

Kim, Y.-S., Al Otaiba, S., Wanzek, J., & Gatlin, B. (2015). Toward an understanding of
dimensions, predictors, and the gender gap in written composition. *Journal of
Educational Psychology, 107*(1), 79-95. https://doi.org/10.1037/a0037210

Kim, Y.-S., Gatlin, B., Al Otaiba, S., & Wanzek, J. (2017). Theorization and an empirical
investigation of the component-based and developmental text writing fluency
construct. *Journal of Learning Disabilities, 51*(4), 320-335.
https://doi.org/10.1177/0022219417712016

Kim, Y.-S., & Schatschneider, C. (2017). Expanding the developmental models of writing:
A direct and indirect effects model of developmental writing (DIEW). *Journal of
Educational Psychology, 109*(1), 35-50. https://doi.org/10.1037/edu0000129

Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). New
York: The Guilford Press.

Lahuerta, A. (2017). Analysis of accuracy in the writing of EFL students enrolled on CLIL
and non-CLIL programmes: The impact of grade and gender. *The Language
Learning Journal*, 1-12. https://doi.org/10.1080/09571736.2017.1303745

Llinares, A., & Whittaker, R. (2007). Talking and writing in a foreign language in CLIL
contexts: A linguistic analysis of secondary school learners of geography and
history. *Volumen Monográfico*, 83-91.

Llinares, A., & Whittaker, R. (2010). Writing and speaking in the history class: A
comparative analysis of CLIL and first language contexts. In C. Dalton-Puffer, T.
Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL
classrooms* (pp. 125-143). Amsterdam: John Benjamins.
https://doi.org/10.1075/aals.7.14dal

Macaro, E. (2018). *English medium instruction*. Oxford, U.K.: Oxford University Press.

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.).

Mahwah, NJ: Lawrence Erlbaum Associates.

https://doi.org/10.1177/014272370002006006

MacWhinney, B. (2019). *Tools for analyzing talk part 2: The CLAN program*.

https://doi.org/10.21415/T5G10R

McMaster, K. L., Shin, J., Espin, C. A., Jung, P.-G., Wayman, M. M., & Deno, S. L.

(2017). Monitoring elementary students' writing progress using curriculum-based

measures: Grade and gender differences. *Reading and Writing: An*

Interdisciplinary Journal, 30(9), 2069-2091. https://doi.org/10.1007/s11145-017-

9766-9

Miret Pallarols, M. (2009). *CLIL effects on writing performance and overall proficiency of*

secondary school students. (Unpublished master's thesis), University of Barcelona,

Barcelona, Spain.

Muthén, L. K., & Muthén, B. O. (2017). Mplus 8.4 [Computer software]. Los Angeles,

CA: Muthén and Muthén.

National Development Council. (2018). *Blueprint for developing Taiwan into a bilingual*

nation by 2030. Retrieved from

https://www.ndc.gov.tw/en/Content_List.aspx?n=D933E5569A87A91C&upn=963

3B537E92778BB.

Navés, T. (2011). How promising are the results of integrating content and language for

EFL writing and overall EFL proficiency? In Y. Ruiz de Zarobe, J. M. Sierra, & F.

Gallardo del Puerto (Eds.), *Content and foreign language integrated learning:*

Contributions to multilingualism in European contexts (pp. 155-186). Bern: Peter Lang. https://doi.org/10.3726/978-3-0351-0171-3

Olsson, E., & Sylvén, L. K. (2017). Validity in high-and low-stakes tests: A comparison of academic vocabulary and some lexical features in CLIL and non-CLIL students' written texts. *Oslo Studies in Language, 9*(3), 127–146. https://doi.org/10.17011/apples/urn.201512234129

Pessoa, S., Miller, R. T., & Kaufer, D. (2014). Students' challenges and development in the transition to academic writing at an English-medium university in Qatar. *International Review of Applied Linguistics in Language Teaching, 52*(2), 127-156. https://doi.org/10.1515/iral-2014-0006

Polio, C., & Park, J.-H. (2016). Language development in second language writing. In R. M. Machón & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 287-306). Boston: De Gruyter. https://doi.org/10.1515/9781614511335-016

Ruiz de Zarobe, Y. (2010). Written production and CLIL: An empiricial study. In C. Dalton-Puffer, T. Nikula, & U. Smit (Eds.), *Language use and language learning in CLIL classrooms* (pp. 191-209). Amsterdam: John Benjamins. https://doi.org/10.1075/aals.7.10rui

Silva, T. (1993). Toward an understanding of the distinct nature of L2 writing: The ESL research and its implications. *TESOL Quarterly, 27*(4), 657-677. https://doi.org/10.2307/3587400

StataCorp. (2019). Stata statistical software: Release 16. College Station, TX: StataCorp LLC. https://doi.org/10.7717/peerj.4598/supp-10

Tai, H.-Y. (2015). Writing development in syntactic complexity, accuracy and fluency in a content and language integrated learning class. *International Journal of Language and Linguistics, 2*(3), 149-156. Retrieved from http://ijllnet.com/journal/index/2190

Vidal, K., & Jarvis, S. (2018). Effects of English-medium instruction on Spanish students' proficiency and lexical diversity in English. *Language Teaching Research*, 1-20. https://doi.org/10.1177/1362168818817945

Whittaker, R., & Llinares, A. (2009). CLIL in social science classrooms: Analysis of spoken and written productions. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán (Eds.), *Content and language integrated learning: Evidence from research in Europe* (pp. 215-234). Bristol, U.K.: Multilingual Matters. https://doi.org/10.21832/9781847691675-015

Whittaker, R., Llinares, A., & McCabe, A. (2011). Written discourse development in CLIL at secondary school. *Language Teaching Research, 15*(3), 343-362. https://doi.org/10.1177/1362168811401154

Xanthou, D.-C. (2017). *The effect of CLIL on the writing and vocabulary skills of 3rd graders in the 3rd Experimental Primary school of Thessaloniki*. (Unpublished master's thesis), Aristotle University of Thessaloniki, Thessaloniki, Greece.

4. GENDER, TRANSCRIPTION, AND IDEATION AS PREDICTORS OF CLIL WRITING PERFORMANCE

Writing is an indispensable skill in today's world, used for a variety of tasks such as sending a digital message through social media or email, writing a report for educational or professional purposes, among others (Camacho & Alves, 2017; Graham, 2008; Graham, Gillespie, & McKeown, 2013). For many English-users around the world for whom English is not their first language, English writing often provides a gateway to educational or professional opportunities at home and abroad. More and more higher education institutions are being conducted through English medium instruction (Dearden, 2015), and admission is often predicated on English proficiency tests with English writing components, often with prompts for a variety of genres, or writing types. Professionally, English writing is often used for connecting with other professionals and disseminating knowledge internationally.

Acknowledging writing's importance, one critical mission of schools around the world is to help develop student writing, often both in the first language (L1) and in a second or additional language (L2), which in many programs is usually English. However, two issues often inhibit this goal: (1) writing is a difficult skill to develop, especially for young learners (Camacho & Alves, 2017), and (2) the current knowledge base on writing development is sparse and incomplete (Graham et al., 2013), especially for school-aged L2 learners (Lee, 2016). Previous research on writing has shown that early success in writing can often predict future success in school and beyond, but early struggles can often lead to

low motivation and self-efficacy (Berninger & Graham, 1998; Camacho & Alves, 2017; Graham, 2008).

One type of language program proposed to prepare students for the multiple contexts in which English is used is Content and Language Integrated Learning (CLIL). Rather than only addressing general English, as is typically done in traditional English language programs, CLIL provides opportunities for students to learn both language and content simultaneously, thus providing students the skills to use English in a variety of disciplinary contexts. Especially in terms of writing, CLIL often offers opportunities to develop skills in a variety of writing genres that are discipline specific. However, current evidence on CLIL success, particularly in terms of English writing outcomes, is mixed (Graham, Choi, Davoodi, Razmeh, & Dixon, 2018).

In light of the need for developing L2 writing skills for a variety of contexts and genres through CLIL, this study examines writing development of two genres, narrative and expository writing, from a cognitive component skills perspective, grounded in the Simple View of Writing. Previous research suggests that stronger abilities in various component skills associated with transcription and ideation contribute to more successful writing (Harrison et al., 2016). Based on this literature, this study examines the contribution of component skills along with gender to writing for English L2 learners in a primary CLIL setting.

4.1. Theoretical Framework

In his chapter on theoretical orientations of L2 writing, Cumming (2016) explains that much of L2 writing theory has been influenced by theories from L1 writing. One such case is studies conducted from a cognitive theory perspective, which have been influenced

greatly by the seminal L1 process model of Hayes and Flowers (1980). This model describes the process a writer goes through as they compose a piece of writing, including planning, translating, and reviewing. As influential as this model has been in both L1 and L2 writing research, Berninger and Swanson (1994) note that "Hayes and Flower's model is about skilled writing and does not capture the uniqueness of beginning and developing writing" (p. 58). It is from this gap that the Simple View of Writing emerged as a theory of L1 writing (Berninger & Swanson, 1994; Juel, Griffith, & Gough, 1986).

The Simple View of Writing theory posits that writing is comprised of two components—ideation and transcription. Ideation is how ideas are communicated through language whereas transcription is the orthographic component, or the translation of these ideas into written form. From a cognitive theory perspective, both of these processes and their associated component skills are seen as equally important in writing. If ideas are absent, there is nothing to write about, and if skills in transcription are lacking, ideas cannot be communicated through writing.

Conceptualizing the component skills that represent ideation and transcription has been challenging for many researchers. Transcription has been relatively straightforward, being most often measured through spelling ability. However, conceptualizing ideation has provided a bit more difficulty. Generally, researchers have used oral language as a representation of ideation ability (Kim, Al Otaiba, Wanzek, & Gatlin, 2015; Kim, Gatlin, Al Otaiba, & Wanzek, 2017; Kim, Park, & Park, 2015). Vocabulary has also been used as a measure of ideation as well (Coker, 2006; Harrison et al., 2016; Kim, Al Otaiba, & Wanzek, 2015; Kim, Al Otaiba, Wanzek, et al., 2015). In regard to using oral language as a conduit for ideation, Kim, Al Otaiba, Wanzek, et al. (2015) suggest:

Generated ideas cannot be produced without being translated into oral language because the child has to express ideas using appropriate words, encode them using appropriate syntactic structure, and organize and present them in a logical sequence. Therefore, oral language proficiency would determine how the generated ideas are adequately expressed (p. 81).

In support of vocabulary, it can be seen in the above quote that for the encoding, organizing, and presenting of ideas through oral language to occur, "appropriate words" are needed as a prerequisite. In other words, vocabulary at the word level supports oral language at the discourse level. Thus, in this study transcription will be represented by spelling ability and ideation by vocabulary and discourse-level oral language (speaking) ability.

Interestingly, while the Hayes and Flowers' (1980) study has received much attention and has been adapted for L2 writing, the Simple View of Writing has not. This is despite the criticism of the process model as not representative of developing writers (Berninger & Swanson, 1994). With a growing population of young English learners around the world and the relative lack of research on young L2 writers (Lee, 2016), the Simple View of Writing may be useful in driving inquiries into young learners' L2 writing processes. When examined together with genre theories of writing, the component processes associated with young learner writing and how they differ across the types, or genres, of writing can be investigated. Particularly in regard to ideation, it is possible that the processes of young learners are not static and may differ as students write in various genres.

4.2. Literature Review

4.2.1. Spelling Ability and Writing

Spelling is a fundamental skill needed for writing. Nation (2009) opines, "If learners have poor spelling skills, they will typically avoid writing tasks, and when writing will avoid words that they find difficult to spell" (p. 18). Various studies have reported correlations between spelling and writing for English L2 learners. Babayiğit (2014) reported a correlation of $r = .5$ for English L2 primary school learners. However, two studies showed correlations at the lower and higher ends. Abu-Rabia (2003) reported a small correlation ($r = .29$) between scores on a spelling sentence dictation task and a writing task scored for thematic maturity for L2 English learners in Israel. On the other end, Schoonen, van Gelderen, Stoel, Hulstijn, and de Glopper (2011) reported a large correlation ($r = .78$) for L2 English high school learners in the Netherlands, where spelling was assessed by identifying the correctly spelled words. It is possible that these discrepancies may be associated with the spelling assessment methods used, though further investigation would be required to substantiate this claim.

In today's world of technology, some may question the need for spelling skills with features such as text-to-speech and spell checkers embedded within today's computers and smartphones. Though interventions with technology have been shown to improve English language learners' writing (Lin, Liu, & Paas, 2017), these tools enhance skills rather than replace them. These technologies do not guarantee accurate transcription (Baker, 2017), and it is still often upon the writer to choose the appropriate spelling or even decide to make a correction (Heift & Rimrott, 2008). For students with low skill levels, there is no assurance that the technology can present the correct feedback nor that the writer will

105

recognize the appropriate change to be made. While today's technology may assist learners with spelling in their writing, a certain degree of spelling ability is still required (Baker, 2017).

4.2.2. Oral Language Ability and Writing

Second language research has shown that a reciprocal relationship exists between oral language and writing ability (Rubin & Kang, 2008; Williams, 2008). Writing skills support oral language development, and oral language skills support writing development. The examination of oral language ability as a cognitive component process in writing has been examined in L1 contexts, but it has received limited attention in L2 research. When considering the connection between L2 oral language and writing, most research has been driven by sociocultural theories (Hirvela & Belcher, 2016; Williams, 2008). As such, oral language has been mainly viewed as a tool for collaboration, feedback, and presentation, or as a way to facilitate and enhance writing.

The few studies that have looked at the influence of L2 oral language ability on L2 writing have shown positive results. Harrison et al. (2016) found that oral language and lexical retrieval predicted over 14% of variance in L2 written content and structure for English as a second language primary students in Canada. With Korean adult learners, Pae and O'Brien (2018) found that oral language independently predicted 43% of variance in writing. When controlling for listening and reading skills, the authors note that oral language explains 4% above and beyond those variables. Waluyo (2018) found a similar impact of oral language skills on writing for Thai university students, reporting oral language and writing correlations of $r = .59$ and $r = .47$ for beginner and intermediate

students, respectively, suggesting oral language skills may independently account for 34.8% and 22.1% of variation in writing for these students.

4.2.3. Vocabulary Ability and Writing

Of the three skills examined in this study, the importance of vocabulary in regard to L2 writing has by far received the most attention, and with good reason. Nation and Newton (2009) note, "The most important 2000 to 3000 word families make up such a large portion of both spoken and written use that it is difficult to use the language effectively without a good knowledge of them" (p. 33). In other words, a lack of vocabulary can greatly inhibit the ability to write.

A handful of recent studies have documented the relationship of vocabulary and writing for primary students. Harrison et al. (2016) reported a small correlation of .04 for primary school English L2 learners in Canada. In contrast, Zhang, McBride-Chang, Wagner, and Chan (2014) showed much stronger correlations between vocabulary and L2 writing for students in Hong Kong at various ages. (age 5 = .38; age 6 = .45; age 7 = .50; age 8 = .45; age 9 = .59). Recent studies have also examined the impact of vocabulary on the L2 writing of secondary students. Wu, Dixon, Sun, and Zhang (2019) reported large correlations between vocabulary and writing for eighth grade (receptive = .52; productive = .55) and ninth grade students (receptive = .74; productive = .74) in China. Similar findings were shown in Schoonen et al. (2011) with Dutch students in grade 8 ($r = .53$) and grade 9 ($r = .57$).

In order to better understand the wide variation reported in current research, Graham and Eslami (2019) conducted a meta-analysis of studies reporting correlations between L2 vocabulary and writing. Results of the analysis estimates a correlation of $r = .5$

for the two constructs. However, there was a significant amount of variation ($I_2 = 82.49$)

left unaccounted for in the model. Further moderator analysis revealed that some of this

variation may be related to study setting, meaning countries where English is spoken as the

first language of the majority versus those where the population generally has a first

language other than English. Level of schooling (or age groups) was not found to explain

the unaccounted variation. Though studies report a varying degree of influence of

vocabulary on writing and reasons for this variation remain unknown, what is fairly certain

is that vocabulary does have an impact on L2 writing.

4.2.4. Gender and Writing

   While age may be a factor in the effect of transcription and ideation, gender has

also been shown to be a factor in writing achievement and language learning. In English

L1 studies of elementary writing, both Kim, Al Otaiba, Wanzek, et al. (2015) and

McMaster et al. (2017) show females outperforming males in writing performance and

growth. In L2 English contexts, a study by Roquet, Llopis, and Pérez-Vidal (2016) showed

similar findings for general language learning where females outperformed males, both in

traditional English as a foreign language (EFL) courses and Content and Language

Integrated Learning (CLIL) contexts. Lahuerta (2017) found contrasting results with the

studies above when specifically studying writing in CLIL, showing no significant

differences between female and male CLIL students; however, in this same study females

did outperform males in the non-CLIL group. The author suggests that "CLIL helps

balance gender differences" (Lahuerta, 2017, p. 11). One possible explanation for this

could be related to motivation. Fontecha and Alonso (2014) found that while females were

slightly more motivated than males in EFL courses, the reverse was found to be true in

CLIL. However, not all studies have found linguistic and/or motivation differences between genders in CLIL (Fernandez-Fontecha, 2014). Thus, further research is needed to better understand the differences between genders in CLIL writing.

4.3. Current Study

Though some is known about the role spelling, vocabulary, and oral language independently play as component skills in the process of L2 writing, almost no studies exist in which these skills are examined as dimensions of the Simple View of Writing— transcription and ideation—in a single model. This is especially true in CLIL contexts. Additionally, there are questions in regard to the role of gender in the process of writing and its effects in CLIL instruction. This study looks to address these gaps in knowledge about gender and the component skill processes involved in writing for CLIL students. Specifically, this research will address the following questions:

1. How does writing performance for both narrative and expository writing in CLIL differ by gender?

2. Above and beyond gender, to what degree does spelling, a measure of transcription, explain the variations in writing performance for both narrative and expository writing in CLIL?

3. Above and beyond gender and transcription, to what degree does vocabulary, a measure of word level ideation, and discourse level oral language ability independently explain the variation in writing performance for both narrative and expository writing in CLIL?

4.4. Methods

4.4.1. Participants

Participants in this study include 56 Taiwanese sixth grade elementary students enrolled in a private school in northern Taiwan. The school's program follows a Mandarin Chinese curriculum provided by Taiwan's Ministry of Education, and their English curriculum is based on the principles of Content and Language Integrated Learning (CLIL). Beginning in first grade, students attend seventeen 40-minute periods of English CLIL per week. Classes in English/Language Arts (ELA) (eight periods per week) use a textbook series from the USA. The writing class (two periods per week) generally uses activities from ELA or other courses to inform the writing curriculum. The math curriculum (two periods per week) mirrors the curriculum used in the Mandarin Chinese class, with students learning the math concepts first in their L1 and then again in English. The course uses materials from a math textbook published in Singapore along with supplemental material. Similar to ELA, the Social Studies/Science class (three periods per week) follows a United States curriculum and uses various textbooks published in the U.S. as materials for the course. Finally, Physical Education (PE) and the Enrichment Course (EC) have an open curriculum (each one period per week). The teacher is given freedom to decide which athletic activities to do with the class in PE and hands-on activities (e.g., arts and crafts, project-based activities) to engage in during EC. By the end of sixth grade, most students' English proficiency levels on the Common European Framework of Reference for Languages (CEFR) are between A2 (upper beginner) and B1 (lower intermediate) with a few students reaching B2 (upper intermediate), based on scores on the Cambridge English Assessment Exams (Cambridge Assessment English, 2019a).

4.4.2. Measures

4.4.2.1. Spelling Ability

English spelling ability was measured using the spelling subtest of the Woodcock-Johnson Tests of Cognitive Abilities III (McGrew, Schrank, & Woodcock, 2007). The test administrator read a word, an example sentence with the word, and repeated the word a second time. Students were asked to dictate the word they heard on their test paper. According to the testing manual, the testing concludes once the test ceiling is reached (i.e., missing six words in a row). Scores were calculated for each student by giving one point for each correct word until the testing ceiling. A calculation of Cronbach's alpha for internal consistency showed excellent internal reliability, $\alpha = .94$.

4.4.2.2. Vocabulary Ability

English vocabulary ability was measured using an adapted version of the first four levels of the Updated Vocabulary Levels Test (Webb, Sasao, & Ballance, 2017). The test is designed to measure English language learners' vocabulary knowledge on the most common words used in English and has been used in previous CLIL vocabulary development research (Reynaert, 2019). The test is broken into sections testing sets of the most common 1,000 words. Due to age and time considerations, a selection of half of the questions from each of the sections were administered from the first four sections of the test (1,000-4,000 words), totaling 60 questions. Students were shown a grid with three definitions in the first column and six words in subsequent rows. Students were directed to put a check mark in the cell corresponding to the correct definition and word. Students received one point for each correct answer. A calculation of Cronbach's alpha for internal consistency showed excellent internal reliability, $\alpha = .95$.

111

4.4.2.3. Oral Language Ability

Oral language discourse ability was measured using scores from Cambridge Assessment English Exams. Cambridge Assessment offers standardized English proficiency exams for young learners through adults at all levels of English proficiency. All of these exams are aligned to the Cambridge English Scale, allowing for comparisons between tests (Cambridge Assessment English, 2019a). As a result of the range of English proficiencies in sixth grade, students took either the Young Learner English (YLE): Movers, YLE: Flyers, Key English Test (KET) for Schools, Preliminary English Test (PET) for Schools, or First Certificate in English (FCE) exam. The exams are administered annually on the school's campus and all students are encouraged, though not required, to sit the exam most appropriate for their current level, as determined by the school's teachers and administration. Final speaking scores for the students were provided to the researcher by Cambridge via the school, thus reliability for the sample could not be calculated. Cambridge Assessment English (2019b) reports the following inter-rater correlation for the speaking sections of each test: KET $r = .86$, PET $r = .85$, and FCE $r = .84$. Reliability measures for the speaking sections of the YLE Movers and Flyers are not reported on Cambridge's website.

4.4.2.4. Writing Ability

Students were administered two writing assessments, one for narrative writing and one for expository writing. For the narrative writing assessment, students were asked to write a story about a child who broke the teacher's glasses (Camacho & Alves, 2017). For the expository writing assessment, students were asked to write about something they learned about in science class during the previous year. The expository prompt was

intentionally left open in order to allow students to choose a topic in science they were comfortable with, rather than restricting them to a specific topic where they may lack the appropriate knowledge to complete the task. This decision was made considering recommendations by Whittaker and Llinares (2009), who suggest that writing prompts should involve "tasks for which the students were prepared, on topics they enjoyed, to make sure that there would be sufficient production" (p. 219).

Prior to the assessment, both prompts were checked by the school's teachers and administrators, who deemed the topics appropriate and accessible for the students. For both assessments, students were given three minutes to plan, where they were instructed they could draw or write, and five minutes to write (McMaster et al., 2017). Students were notified when one minute was left for writing so they may complete their text.

Given the complexity of writing as a construct, factor scores were calculated using Mplus 8.4 (Muthén & Muthén, 2017) with the following measures: spelling accuracy, correct word sequences, lexical diversity, total t-units, and writing quality (ideas). Details on the factor loadings for each variable can be found in Section 3 of this dissertation. The factor scores were transformed into T scores with a mean of 50 and standard deviation of 10 in order to assist with interpretation of the results.

With the exception of writing quality, scores for each metric were calculated using CLAN (MacWhinney, 2000). The writing samples were first transcribed and then transferred into CLAN for coding of spelling and clausal errors. The EVAL command was then used in CLAN to retrieve the metrics used in the factor score. Spelling accuracy is the percentage of correctly spelled words. Correct word sequences is the percentage of word sequences, or adjacent words, that are morpho-syntactically correct. Spelling errors were

not considered in this score as they were accounted for in the previously described spelling accuracy score. Lexical diversity was calculated using a formula proposed by Carroll (1964) where types (unique words) is divided by the square root of tokens (total words) multiplied by two. Total t-units is the number of t-units, or units of text that can stand alone as a sentence within the text.

The final score, writing quality, was determined using an adapted version of the ideas rubric from Education Northwest (2013). The rubric has a six-point range. One point was awarded for texts where ideas were not presented through letters or texts. Two points were rewarded for texts with comprehensible letters but either incomprehensible words or no complete sentences (i.e., list of words). Three points were rewarded for texts with one t-unit. Four points were rewarded for texts with multiple t-units that lacked focus on a central idea, and five points were rewarded for texts where a clear focus on a central idea was present. Finally, six points were awarded for any texts that were well focused with rich descriptions of the topic.

Every writing sample was scored by the lead researcher and a trained graduate student. The lead researcher and graduate student calibrated on the rubric and independently scored a stratified sample of 20 percent of the writing samples, stratified by genre and English class. Using quadratic weighted Kappa as a metric for inter-rater reliability, kappa scores of $\kappa = .82$ for narrative writing and $\kappa = .81$ for expository writing were reached. Discrepancies were discussed and resolved. Then, the lead researcher scored the remaining samples independently.

4.4.3. Data Collection

Data were collected during the second week of school in August, with the exception of the Cambridge Exams, which were administered in March of the previous semester. Each of the assessments were given during regular English classes and proctored by their classroom teacher. Prior to administering the assessments, each proctor was trained by the lead researcher and provided a written script to use when giving the assessment. Each assessment was given in separate class periods over a period of one week.

4.4.4. Data Analysis

Two hierarchical regression models (Cohen, Cohen, West, & Aiken, 2002), one for the narrative and one for the expository genre, were run using STATA 16 (StataCorp, 2019). In determining the order of variables to be entered into the hierarchical regression models, the principals of causal priority were used. In the first model, gender was entered. This was entered first both to address the research question regarding gender differences as well as to act as a covariate in all subsequent models. In the second model, spelling, representing a measure of transcription, was entered. Entering this variable second was based on the assertion of Kim et al. (2017) that transcription skills are foundational for subsequent ideation in writing. In models three and four, variables of ideation were entered. In model three, vocabulary as a word-level measure of ideation was entered given that words are the foundation of discourse. Finally, discourse-level ideation through oral language was entered into model four.

Prior to the regression analyses, assumption checking for regression was conducted. First, linearity was checked using a residual versus predictor plot with lowest

fit. Second, normality of residuals was checked using a QQ plot, statistics for skewness and kurtosis, and a Shapiro-Wilk test. Finally homoscedasticity was checked using a residual versus fitted plot and a Breusch-Pagan/Cook-Weisberg test for heteroskedasticity. The results of the various tests suggest that all of the assumptions of regression were met.

Given that the oral language assessment administered by the school was non-compulsory, the data set contained missing values for 18 participants (32.14%). Little's MCAR test was conducted and was non-significant, $\chi_2 (4) = 5.89$, $p = .21$, suggesting that the data may be missing completely at random. To address the missing data, multiple imputation using multivariate normal regression (MVN) was used to impute missing oral language values.

Finally, the data were examined for outliers using the following diagnostic tests: leverage, studentized residuals, Cook's D and DFBETAS. An analysis of the results suggests that there may be three cases that could be outliers. However, due to the limited sample size of this study, the decision was made to leave these cases in for analysis. To compensate for the possible influence of the outliers, the regression analyses were conducted using the robust variance estimation method, which should minimize any effect the outliers may have on the results.

4.5. Results

Descriptive statistics for each of the measures are reported in Table 4.1. The three independent variables follow a normal distribution with skewness close to 0 and kurtosis around 3. However, both the narrative and expository writing measures exhibited a high skewness and kurtosis and both were negatively skewed.

Table 4.1 Descriptive Statistics for Language Ability Measures ($N = 56$)

Measure	Mean	Median	SD	Sk	Ku	Min	Max
Spelling	14.63	14.5	7.55	0.37	2.28	4	32
Vocabulary	36.02	36.5	12.49	-0.21	2.18	7	57
Oral Language	141.11	135	17.20	0.38	2.63	105	178
Narrative	50	53.14	10	-2.15	8.91	10.60	64.69
Expository	50	53.12	10	-2.16	9.01	10.41	64.85

Note. Due to missing data, oral language N = 38; Narrative and Expository scores are standardized T scores.

Table 4.2 displays a correlation matrix for the two writing genres, spelling, vocabulary, and oral language. All correlations in the table are statistically significant at $p < .001$. The correlation between the two writing genres is near perfect, $r = .99$, suggesting that there may be little difference in the subsequent regression models for the two genres. The correlations between many of the dependent variables in the model are also quite high, $r > .7$, suggesting the model may contain multicollinearity. This concern may be particularly warranted for spelling and vocabulary, which has a correlation of $r = .90$.

Table 4.2 Correlation Matrix for Language Ability Measures

	Spelling	Vocabulary	Oral Lang.	Narrative	Expository
Spelling	1.00				
Vocabulary	.90	1.00			
Oral Language	.73	.78	1.00		
Narrative	.64	.76	.61	1.00	
Expository	.64	.76	.61	.99	1.00

Note. All correlations are statistically significant, p < .001.

To test for the possible presence of multicollinearity, variance inflation factor (VIF) values were calculated for the model with gender, spelling, vocabulary, and oral language (Model A). There is little agreement between researchers on the exact VIF cutoff value for identifying multicollinearity. Though many agree that any VIF value above 10 is certainly a concern, Vatcheva, Lee, McCormick, and Rahbar (2016) show that VIF below 5 can also present problems, and Gordon (2015) notes that any VIF values above 4 may need attention. As can be seen in Table 4.3, the VIF values for vocabulary and spelling are above the threshold presented by Gordon (2015), thus indicating multicollinearity could be a factor. When multicollinearity occurs, one possible remedy is to remove one of the overlapping variables (Kock & Lynn, 2012); therefore, a second model with vocabulary removed was assessed (Model B). The choice to remove vocabulary rather than spelling was based on the fact that this study is testing the Simple View of Writing and the removal of the spelling variable would in turn eliminate transcription from the model, thus disallowing a test of the theory. The VIF values for Model B were found to all be below 4, suggesting that multicollinearity may not be an issue for Model B (Table 4.3). In light of the possible multicollinearity issue in Model A, regression analyses for both Model A and Model B will be reported.

Table 4.3 Variance Inflation Factor (VIF) Values for Language Ability Measures

Variable	Model A VIF	Model B VIF
Vocabulary	5.83	-
Spelling	4.86	2.17
Oral Language	2.72	2.23
Gender	1.06	1.05

4.5.1. Hierarchical Regression Models

In the following paragraphs, results of a hierarchical regression analysis using robust estimation for narrative and expository writing will be presented. First, results for Model A which has gender, spelling, vocabulary, and oral language as predictor variables entered one by one will be presented. Then, Model B which excludes vocabulary based on a potential violation of multicollinearity will be examined.

Model A for narrative writing accounted for 64.9% of the variation in narrative writing and was statistically significant, $F(4, 43.5) = 8.19$, $p < .001$ (Table 4.4). In the first step, gender was entered and found to account for 3.2% of variation in narrative writing. This was not statistically significant, $p = .176$. Adding spelling to the model increased the variation accounted for by 40.4% for a total of 43.6% of variation predicted by the two variables. A one-point increase in spelling achievement is predicted to raise the narrative writing score of females by 0.84, a statistically significant effect, $p < .001$. Gender remained a non-significant predictor when controlling for spelling, $p = .110$, suggesting both genders would experience similar gains in writing with increased spelling achievement scores.

Table 4.4 Hierarchical Regression Estimates for Narrative Writing (Model A)

Variable	B	SE	p	95% CI	R^2	ΔR^2
Step 1					.032	
Gender	-3.57	2.60	.176	-8.78, 1.65		
Step 2					.436***	.404***
Gender	-3.20	1.97	.110	-7.14, 0.75		
Spelling	0.84	0.17	< .001	0.51, 1.17		
Step 3					.624***	.188***
Gender	-3.58	1.68	.038	-6.94, -0.21		
Spelling	-0.34	0.25	.172	-0.84, 0.15		
Vocabulary	0.80	0.21	< .001	0.38, 1.22		
Step 4					.649***	.025
Gender	-3.91	1.80	.035	-7.53, -0.29		
Spelling	-0.37	0.24	.136	-0.86, 0.12		
Vocabulary	0.69	0.23	.006	0.22, 1.17		
Oral Lang.	0.09	0.13	.503	-0.19, 0.36		

With the addition of vocabulary to the model, there was an 18.8% increase in variation accounted for, raising the model to 62.4% of variance accounted for. Accounting for both gender and spelling, vocabulary is predicted to raise narrative writing scores by .80 for every one-point increase in vocabulary achievement for females. This is statistically significant, $p < .001$. With the addition of vocabulary into the model, spelling is no longer a statistically significant predictor in the model. However, when controlling for spelling and vocabulary achievement, gender became a statistically significant predictor, $p = .038$, suggesting that females with average vocabulary and spelling achievement outperform their male counterparts in narrative writing by an estimated 3.57 points.

The final model added oral language to the model, which was found to not be statistically significant, $p = .503$, and only added an additional 2.5% of variance accounted for in the model. Vocabulary, $p = .006$, and gender, $p = .035$, remained statistically significant in the final model while spelling remained non-significant, $p = .136$. As expected based on the high correlation between the two genres, the expository writing model was nearly identical to the narrative model. Given the similarity between the two models, the statistics for the model are reported in Table 4.5 but will not be detailed here in order to avoid redundancy.

Model B for narrative writing is statistically significant, $F(3, 41.9) = 6.74, p < .001$, and accounts for 52.5% of the variation in narrative writing (Table 4.6). Steps 1 and 2, where gender and spelling are the same as in Model A, are detailed in the preceding paragraph. With the removal of the vocabulary variable due to potential multicollinearity, the final step 3 adds the variable oral language to the model with gender and spelling. This model accounted for an additional 8.9% of variation in narrative writing, a non-significant change.

Table 4.5 Hierarchical Regression Estimates for Expository Writing (Model A)

Variable	B	SE	p	95% CI	R2	ΔR2
Step 1					.031	
Gender	-3.48	2.60	.187	-8.70, 1.74		
Step 2					.432***	.401***
Gender	-3.11	1.97	.121	-7.07, 0.85		
Spelling	0.84	0.17	< .001	0.50, 1.17		
Step 3					.622***	.190***
Gender	-3.49	1.68	.043	-6.86, -0.11		
Spelling	-0.35	0.25	0.165	-0.85, 0.15		
Vocabulary	0.80	0.21	< .001	0.38, 1.22		
Step 4					.646***	.024
Gender	-3.82	1.81	.040	-7.45, -0.18		
Spelling	-0.38	0.25	.130	-0.87, 0.12		
Vocabulary	0.70	0.23	.006	0.22, 1.18		
Oral Lang.	0.09	0.13	.508	-0.19, 0.36		

Table 4.6 Hierarchical Regression Estimates for Narrative Writing (Model B)

Variable	B	SE	p	95% CI	R2	ΔR2
Step 1					.032	
Gender	-3.57	2.60	.176	-8.78, 1.65		
Step 2					.436***	.404***
Gender	-3.20	1.97	.110	-7.14, 0.75		
Spelling	0.84	0.17	< .001	0.51, 1.17		
Step 3					.525***	.089
Gender	-4.20	2.11	.053	-8.46, 0.05		
Spelling	0.43	0.28	.142	-0.16, 1.03		
Oral Lang.	0.21	0.13	.134	-0.07, 0.49		
Step 4					.461***	
Gender	-4.99	2.24	.031	-9.49, -0.49		
Oral Lang.	0.34	0.08	< .001	0.17, 0.51		

While this final model is statistically significant, gender, $p = .053$, spelling, $p = .142$, and oral language, $p = .134$, were all found to be non-significant when controlling for the other variables in the model. Given these results, a final step 4 analysis was conducted to better understand the contribution of oral language to narrative writing without controlling for spelling. Controlling for gender only, oral language was found to be statistically significant, $p < .001$. An increase of one point in oral language achievement is predicted to have a .34-point increase in narrative writing for female students. In this model, gender is a significant predictor when controlling for oral language, $p = .034$. Females with an average oral language ability are predicted to outperform their male counterparts in narrative writing by 4.11 points. As was the case with Model A, the results for expository writing are almost identical to narrative writing. Statistics for Model B expository writing can be found in Table 4.7.

Table 4.7 Hierarchical Regression Estimates for Expository Writing (Model B)

Variable	B	SE	p	95% CI	R^2	ΔR^2
Step 1					.031	
Gender	-3.48	2.60	.187	-8.70, 1.74		
Step 2					.432***	.401***
Gender	-3.11	1.97	.121	-7.07, 0.85		
Spelling	0.84	0.17	< .001	0.50, 1.17		
Step 3					.522***	0.09
Gender	-4.11	2.12	.059	-8.39, 0.16		
Spelling	0.43	0.28	0.145	-0.17, 1.03		
Oral Lang.	0.21	0.13	0.136	-0.07, 0.49		
Step 4						
Gender	-4.90	2.24	.034	-9.41, -0.38	.458***	
Oral Lang.	0.34	0.08	< .001	0.17, 0.51		

4.6. Discussion

This study examined writing as a cognitive component skill process by investigating the effect of gender, transcription, and ideation on narrative and expository writing for elementary CLIL students in Taiwan. Results of a hierarchical analysis suggest that these variables may have an effect on CLIL narrative and expository writing for grade 6 primary students, but future research is needed to confirm these results due to study limitations described below. As the effects of these variables were found to be almost identical for narrative and expository writing and the genres were highly correlated ($r = .99$), the following paragraphs will discuss the results generally in terms of CLIL writing rather than discuss the two genres separately.

4.6.1. The Effect of Gender on CLIL Writing

Statistically significant differences in gender alone were not found for CLIL writing in this study. The non-significant results of gender as a predictor seem to align with Lahuerta (2017), however a post-hoc power test for the model suggests caution should be taken when interpreting these results. Considering the low R-squared value for this model (Narrative $R_2 = .032$; Expository $R_2 = .031$), a sample size of 240 participants would be needed to have adequate power of .8. As this study is far below that number, its results may be subject to error.

When controlling for transcription ability through spelling, gender remained a non-significant predictor. This model showed adequate power ($> .99$), accounting for 43.6% of variation in narrative writing and 43.2% in expository writing. This model may provide a stronger argument for supporting lack of differences between genders in CLIL (Lahuerta, 2017). However, when controlling for ideation measures of vocabulary and/or oral

124

language skills, gender emerged as a statistically significant predictor of writing, with CLIL females outperforming CLIL males. This aligns with previous research showing gender as a predictor of language achievement and writing with females showing stronger performance than males (Kim, Al Otaiba, Wanzek, et al., 2015; McMaster et al., 2017; Olinghouse, 2008; Roquet et al., 2016). Viewed through the Simple View of Writing, these contrasting results suggest ideation as a factor in the differences between females and males in writing.

Considering the contrasting results above, further research is needed on the effect of gender on CLIL writing and CLIL performance in general. If it is confirmed that females outperform males, further intervention studies will be needed to help close the gender gap. On the other hand, if CLIL is found to balance language outcomes between genders as suggested by Lahuerta (2017), this may provide positive evidence for implementing CLIL in schools.

4.6.2. The Simple View of Writing as a Theory for L2 Cognitive Writing Processes

In addition to gender, the current study looked to test whether the Simple View of Writing as a theory of cognitive component skills for primary CLIL could explain variation in writing performance. Based on previous literature, it was hypothesized that spelling ability represented the transcription component and that vocabulary and oral language ability, as respective word-level and discourse-level variables, correspond with ideation. Scholars have previously suggested that transcription skills should initially be a strong predictor of writing, but as young learners develop and transcription skills become automatic, ideation becomes a stronger predictor (Kim et al., 2017). While this study may provide evidence for using the Simple View of Writing as a theory for L2 writing

125

processes, there are conflicting results on the interaction of transcription and ideation at this level. Future research examining different grade levels and/or language proficiency may help us better understand how the interaction of transcription and ideation changes at different levels of study and proficiency.

In terms of the balance of transcription and ideation skills in sixth grade CLIL writing, results suggest that transcription skills may still play an important role in writing development. When controlling for gender, the transcription measure of spelling was a statistically significant predictor of CLIL writing. However, when ideation variables were added to the model, spelling was no longer a significant predictor. In the case of adding oral language alone as a measure of ideation (Model B), neither variable was significant when controlling for gender, despite the model accounting for 52.5% of variation in writing. This may indicate transcription and ideation are competing for significance. This is further supported when oral language was examined alone with gender and was also found to be statistically significant. With both spelling and oral language significantly predicting variation in writing independently, it could be suggested that both processes deserve attention in L2 writing instruction. This may also show the final year of elementary school as a possible threshold for the Simple View of Writing, where importance is transitioning from transcription to ideation. To substantiate this, future research should examine this model with both older and younger students in order to understand how these processes interact at different levels of schooling.

In contrast, Model A may present a competing theory. Vocabulary was found to be the only significant component skill predictor of writing in Model A when controlling for gender, spelling, and oral language ability. This could indicate that the sixth grade students

have already passed the threshold where ideation skills are now a stronger predictor of writing. This result also could suggest that vocabulary is a stronger predictor of ideation at the elementary level than discourse-level oral language ability. However, these interpretations may be questionable in light of the multicollinearity issues in this model and caution is urged.

The potential multicollinearity issue with vocabulary and spelling in Model A raise questions about the Simple View of Writing and vocabulary. Theoretically, vocabulary was posited as a measure of ideation, yet the multicollinearity occurred with the transcription variable of spelling. There are a few conclusions that could be drawn from this. One possibility is that variables were correctly specified in Model A, but the sample size was inadequate to overcome multicollinearity. Vatcheva et al. (2016) have suggested that a small sample size could result in multicollinearity and that a larger sample size could reduce the potential for this error. Studies with a larger sample size could help confirm this explanation.

An alternative explanation is that variables were not correctly specified and vocabulary, given its multicollinearity, is a measure of transcription ability, or perhaps more adequately termed as a measure of language ability. There is a history of using L1 theories in L2 writing research (Cumming, 2016), but these theories often require adaptation for L2 contexts. Perhaps an L2 Simple View of Writing would be comprised of what might be termed language skills, rather than transcription exclusively, and ideation ability. In this conceptualization, it may make sense for overlap to occur between vocabulary and spelling, resulting in issues of multicollinearity. With a larger sample size and more indicators that represent transcription/language and ideation, this could be

empirically tested using Confirmatory Factor Analysis (CFA) to validate the constructs followed by a structural equation model to test the Simple View of Writing. Though the previous section (Section 3) found a good model fit for vocabulary as a measure of ideation for writing as a product, it is possible that the Simple View of Writing may fit differently for L2 writing as a process.

One final interpretation is that vocabulary acts as a global variable of language ability for young English language learners, encompassing both transcription and ideation, thus possibly suggesting the lack of need for the Simple View of Writing as a theory of L2 writing as a process. Though the Simple View of Writing has been used as a theory of L1 writing, the differences between L1 and L2 learners may make the theory incompatible for L2 settings. Under this line of thought, vocabulary ability alone may be adequate for understanding elementary L2 writing development, therefore eliminating the need for a theory such as the Simple View of Writing with dimensions of transcription and ideation. Model B may suggest that this is not the case and that the Simple View of Writing may still have value as a theory of L2 writing processes. However, the misapplication of this theory to L2 writing cannot be ruled out until further empirical investigations have been conducted.

4.7. Conclusion

This study examined writing as a process through the dimensions of the Simple View of Writing and gender. Results point to an effect for gender, spelling, vocabulary, and oral language on variation in CLIL writing, though further investigation is needed to better understand the relationship of theses process variables to each other when predicting writing outcomes. It is recommended that future research build on the limitations of this

study in order to further our understanding of how the Simple View of Writing and gender affect CLIL writing. In regard to gender, language instructors should keep in mind the possibility for a gender gap in L2 language and writing development and experiment with ways to differentiate instruction so all students can reach high levels of achievement. In addition, this study shows some promise for the Simple View of Writing as a theory for L2 writing processes. For practitioners, this theory may be able to provide a basis for planning skill instruction in the L2 writing classroom. Attention to transcription and ideation skill building in the L1 classroom has shown promise, and similar attention in L2 settings may lead toward a similar increase in writing achievement.

4.8. References

Baker, E. A. (2017). Apps, iPads, and literacy: Examining the feasibility of speech recognition in a first-grade classroom. *Reading Research Quarterly, 52*(3), 291-310. https://doi.org/10.1002/rrq.170

Berninger, V. W., & Graham, S. (1998). Language by hand: A synthesis of a decade of research on handwriting. *Handwriting Review, 12*, 11-25.

Berninger, V. W., & Swanson, H. L. (1994). Modifying Hayes and Flower's model of skilled writing to explain beginning and developing writing. *Advances in Cognition and Educational Practice, 2*, 57-81.

Camacho, A., & Alves, R. A. (2017). Fostering parental involvement in writing: Development and testing of the program Cultivating Writing. *Reading and Writing: An Interdisciplinary Journal, 30*(2), 253-277. https://doi.org/10.1007/s11145-016-9672-6

Cambridge Assessment English. (2019a). *International language standards*. Retrieved from https://www.cambridgeenglish.org/exams-and-tests/cefr/

Cambridge Assessment English. (2019b). *Quality and accountability*. Retrieved from https://www.cambridgeenglish.org/research-and-validation/quality-and-accountability/

Carroll, J. B. (1964). *Language and thought*. Engelwood Cliffs, NJ: Prentice Hall.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2002). *Applied multiple regression/correlation analysis for the behavioral sciences* (3rd ed.). Mahwah: Lawrence Erlbaum Associates, Inc.

Coker, D. (2006). Impact of first-grade factors on the growth and outcomes of urban school children's primary-grade writing. *Journal of Educational Psychology, 98*(3), 471-488. https://doi.org/10.1037/0022-0663.98.3.471

Cumming, A. (2016). Theoretical orientations to L2 writing. In R. M. Machón & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 65-88). Boston: De Gruyter. https://doi.org/10.1515/9781614511335-006

Dearden, J. (2015). *English as a medium of instruction —A growing global phenomenon*. Retrieved from British Council: https://ora.ox.ac.uk/objects/uuid:4f72cdf8-b2eb-4d41-a785-4a283bf6caaa

Education Northwest. (2013). *Grades K–2 traits rubric for ideas*. Retrieved from https://educationnorthwest.org/sites/default/files/gradesK-2-6pt-rubric.pdf

Fernandez-Fontecha, A. (2014). Motivation and gender effect in receptive vocabulary learning: an exploratory analysis in CLIL primary education. *Latin American*

Journal of Content and Language Integrated Learning, 7(2), 27-49.

https://doi.org/10.5294/laclil.2014.7.2.2

Fontecha, A. F., & Alonso, A. C. (2014). A preliminary study on motivation and gender in

CLIL and Non-CLIL types of instruction. *International Journal of English Studies,*

14(1), 21-36. https://doi.org/10.6018/ijes/14/1/156681

Gordon, R. A. (2015). *Regression analysis for the social sciences*. New York: Routledge.

Graham, K. M., Choi, Y., Davoodi, A., Razmeh, S., & Dixon, L. Q. (2018). Language and

content outcomes of CLIL and EMI: A systematic review. *Latin American Journal*

of Content & Language Integrated Learning, 11(1), 19-37.

https://doi.org/10.5294/laclil.2018.11.1.2

Graham, K. M., & Eslami, Z. R. (2019). Does the simple view of writing explain L2

writing development?—A meta-analysis. Manuscript submitted for publication.

Graham, S. (2008). *Effective writing instruction for all students*. Retrieved from

Renaissance Learning: http://doc.renlearn.com/KMNet/R004250923GJCF33.pdf

Graham, S., Gillespie, A., & McKeown, D. (2013). Writing: Importance, development, and

instruction. *Reading and Writing: An Interdisciplinary Journal, 26*(1), 1-15.

https://doi.org/10.1007/s11145-012-9395-2

Harrison, G. L., Goegan, L. D., Jalbert, R., Mcmanus, K., Sinclair, K., & Spurling, J.

(2016). Predictors of spelling and writing skills in first- and second-language

learners. *Reading and Writing, 29*(1), 69-89. https://doi.org/10.1007/s11145-015-

9580-1

Hayes, J., & Flowers, L. (1980). Identifying the organization of writing processes. In L. W. Gregg & E. R. Steinberg (Eds.), *Cognitive processes in writing* (pp. 3-30). Hillsdale, NJ: Erlbaum. https://doi.org/10.1017/s0142716400006585

Heift, T., & Rimrott, A. (2008). Learner responses to corrective feedback for spelling errors in CALL. *System, 36*(2), 196-213. https://doi.org/10.1016/j.system.2007.09.007

Hirvela, A., & Belcher, D. (2016). Reading/writing and speaking/writing connections: The advantages of multimodal pedagogy. In R. M. Machón & P. K. Matsuda (Eds.), *Handbook of second and foreign language writing* (pp. 587-612). Boston: De Gruyter. https://doi.org/10.1515/9781614511335-030

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78*(4), 243-255. https://doi.org/10.1037/0022-0663.78.4.243

Kim, Y.-S., Al Otaiba, S., & Wanzek, J. (2015). Kindergarten predictors of third grade writing. *Learning and Individual Differences, 37*, 27-37. https://doi.org/10.1016/j.lindif.2014.11.009

Kim, Y.-S., Al Otaiba, S., Wanzek, J., & Gatlin, B. (2015). Toward an understanding of dimensions, predictors, and the gender gap in written composition. *Journal of Educational Psychology, 107*(1), 79-95. https://doi.org/10.1037/a0037210

Kim, Y.-S., Gatlin, B., Al Otaiba, S., & Wanzek, J. (2017). Theorization and an empirical investigation of the component-based and developmental text writing fluency construct. *Journal of Learning Disabilities, 51*(4), 320-335. https://doi.org/10.1177/0022219417712016

Kim, Y.-S., Park, C., & Park, Y. (2015). Dimensions of discourse level oral language skills

and their relation to reading comprehension and written composition: an

exploratory study. *Reading and Writing, 28*(5), 633-654.

https://doi.org/10.1007/s11145-015-9542-7

Kock, N., & Lynn, G. S. (2012). Lateral collinearity and misleading results in variance-

based SEM: An illustration and recommendations. *Journal of the Association for

Information Systems, 13*(7), 546-580. https://doi.org/10.17705/1jais.00302

Lahuerta, A. (2017). Analysis of accuracy in the writing of EFL students enrolled on CLIL

and non-CLIL programmes: The impact of grade and gender. *The Language

Learning Journal*, 1-12. https://doi.org/10.1080/09571736.2017.1303745

Lee, I. (2016). EFL writing in schools. In R. M. Machón & P. K. Matsuda (Eds.),

Handbook of second and foreign language writing (pp. 113-140). Boston: De

Gruyter. https://doi.org/10.1515/9781614511335-008

Lin, P.-H., Liu, T.-C., & Paas, F. (2017). Effects of spell checkers on English as a second

language students' incidental spelling learning: A cognitive load perspective.

Reading and Writing: An Interdisciplinary Journal, 30(7), 1501-1525.

https://doi.org/10.1007/s11145-017-9734-4

MacWhinney, B. (2000). *The CHILDES project: Tools for analyzing talk* (3rd ed.).

Mahwah, NJ: Lawrence Erlbaum Associates.

https://doi.org/10.1177/014272370002006006

McGrew, K. S., Schrank, F. A., & Woodcock, R. W. (2007). *Woodcock-Johnson III

Normative Update Technical Manual*. Rolling Meadows, IL: Riverside Publishing.

McMaster, K. L., Shin, J., Espin, C. A., Jung, P., Wayman, M. M., & Deno, S. L. (2017).

    Monitoring elementary students' writing progress using curriculum-based

    measures: Grade and gender differences. *Reading and Writing, 30*(9), 2069-2091.

    https://doi.org/10.1007/s11145-017-9766-9

Muthén, L. K., & Muthén, B. O. (2017). Mplus 8.4 [Computer software]. Los Angeles,

    CA: Muthén and Muthén.

Nation, I. S. P. (2009). *Teaching ESL/EFL reading and writing*. New York: Routledge.

Nation, I. S. P., & Newton, J. (2009). *Teaching ESL/EFL listening and speaking*. New

    York: Routledge.

Olinghouse, N. G. (2008). Student- and instruction-level predictors of narrative writing in

    third-grade students. *Reading and Writing, 21*(1), 3-26.

    https://doi.org/10.1007/s11145-007-9062-1

Pae, H. K., & O'Brien, B. (2018). Overlap and uniqueness: Linguistic componential traits

    contributing to expressive skills in English as a foreign language. *Reading

    Psychology, 39*(4), 384-412. https://doi.org/10.1080/02702711.2018.1443298

Reynaert, B. (2019). CLIL and productive vocabulary acquisition in the Czech context

    Advances in Language and Literary Studies, 10(4), 153-160.

Roquet, H., Llopis, J., & Pérez-Vidal, C. (2016). Does gender have an impact on the

    potential benefits learners may achieve in two contexts compared: Formal

    instruction and formal instruction + content and language integrated learning?

    International Journal of Bilingual Education and Bilingualism, 19(4), 370-386.

    https://doi.org/10.1080/13670050.2014.992389

Rubin, D. L., & Kang, O. (2008). Writing to speak: What goes on across the two-way street. In D. Belcher & A. Hirvela (Eds.), *The oral-literate connection: Perspectives on L2 speaking, writing, and other media interactions* (pp. 210-225). Ann Arbor: University of Michigan Press. https://doi.org/10.3998/mpub.231182

StataCorp. (2019). Stata statistical software: Release 16. College Station, TX: StataCorp LLC

Vatcheva, K. P., Lee, M., McCormick, J. B., & Rahbar, M. H. (2016). Multicollinearity in regression analyses conducted in epidemiologic studies. *Epidemiology, 6*(2), 1-20. https://doi.org/10.4172/2161-1165.1000227

Waluyo, B. (2018). Vocabulary acquisition through self-regulated learning on speaking and writing development. *International Journal of Language Teaching and Education, 2*(3). https://doi.org/10.22437/ijolte.v2i3.5747

Webb, S., Sasao, Y., & Ballance, O. (2017). The updated Vocabulary Levels Test: Developing and validating two new forms of the VLT. *International Journal of Applied Linguistics, 168*(1), 34-70. https://doi.org/10.1075/itl.168.1.02web

Whittaker, R., & Llinares, A. (2009). CLIL in social science classrooms: Analysis of spoken and written productions. In Y. Ruiz de Zarobe & R. M. Jiménez Catalán (Eds.), *Content and language integrated learning: Evidence from research in Europe* (pp. 215-234). Bristol, U.K.: Multilingual Matters. https://doi.org/10.21832/9781847691675-015

Williams, J. (2008). The speaking-writing connection in second language and academic literacy development. In D. Belcher & A. Hirvela (Eds.), *The oral-literate connection: Perspectives on L2 speaking, writing, and other media interactions*

(pp. 10-25). Ann Arbor: University of Michigan Press. https://doi.org/10.3998/mpub.231182

Wu, S., Dixon, L. Q., Sun, H., & Zhang, P. (2019). Breadth or depth: the role of vocabulary in Chinese English-Language beginning writers' development. *International Journal of Bilingual Education and Bilingualism*, 1-17. https://doi.org/10.1080/13670050.2019.1572066

Zhang, J., McBride-Chang, C., Wagner, R. K., & Chan, S. (2014). Uniqueness and overlap: Characteristics and longitudinal correlates of native Chinese children's writing in English as a foreign language. *Bilingualism: Language and Cognition, 17*(2), 347-363. https://doi.org/10.1017/s1366728913000163

5. CONCLUSIONS

5.1. Introduction

This dissertation examines Content and Language Integrated Learning (CLIL) writing through the lens of the Simple View of Writing (Berninger & Swanson, 1994; Juel, Griffith, & Gough, 1986). Through three studies, one systematic review and two empirical investigations, it has been shown that the Simple View of Writing may provide a useful framework for understanding writing outcomes and processes of CLIL. In this final section, a summary of the three studies will be provided along with suggestions for CLIL researchers and practitioners.

5.2. Summary of Studies

Section 2 details the findings of a systematic review on measures and outcomes reported in previous research on CLIL writing. Measurement in writing was found to be conducted through four broad categories: vocabulary, rubrics, complexity/accuracy/fluency (CAF), and content analysis. While some studies in the review combine multiple types of writing measures, these measures are rarely examined together and often do not account for the dimensions of transcription and ideations of which writing is comprised, as put forth by the Simple View of Writing. Given the complexity of writing, this lack of comprehensive measurement within studies may lead to incomplete or misleading findings on CLIL writing.

In regard to CLIL writing outcomes, two types of studies were present in the literature: those that examine CLIL versus a non-CLIL group and those that measure CLIL growth. Findings of studies investigating CLIL versus non-CLIL writing outcomes are

mixed. The majority of studies show CLIL is equal or exceeds non-CLIL in most measures of outcomes, with measures of accuracy as an exception where non-CLIL groups were found to exceed CLIL groups. As for CLIL growth, studies have generally shown that CLIL instruction leads toward growth in some, but not all, metrics. However, studies have also noted that student writing products have generally shown a lack of content knowledge mastery and lack of characteristic genre features.

Considered together, Section 2 highlights the potential benefits of applying the Simple View of Writing to examinations of CLIL. For one, the theory may help in designing studies that account for and examine the dimensions of transcription and ideation together for a more holistic examination of writing. Second, given that CLIL students may underperform non-CLIL students in accuracy, a focus on transcription, of which accuracy is a possible component, within CLIL writing research and instruction is needed. Alongside transcription, the reviewed studies also point toward a need for attention toward ideation, with CLIL students inadequately displaying content knowledge and genre feature awareness within writing. In sum, while the current research on CLIL writing has begun to answer questions about the effectiveness of CLIL writing outcomes, the Simple View of Writing may provide a framework that will assist in a more comprehensive understanding of the complex construct of writing.

Section 3 examines a model based on the Simple View of Writing for evaluating CLIL writing as a product. A higher-order factor model of CLIL writing was evaluated using Confirmatory Factor Analysis (CFA). The model was comprised of covarying second-order factors for CLIL Narrative Writing and CLIL Expository Writing which both had two first-order factors for transcription and ideation, the dimensions of the Simple

View of Writing. Transcription was made up of two observed variables: spelling accuracy and correct word sequences. Ideation was comprised of three observed variables: lexical diversity, number of t-units, and a holistic rubric score for ideas.

The results of the CFA analysis suggest that the Simple View of Writing Model provides a better fit for the CLIL writing data than a single-factor model that does not account for the dimensions of transcription and ideation. However, an analysis of individual coefficients in the model reveal some coefficients below $\beta = .5$, indicating a weak relation to the latent factors. Given that these instances of low coefficients were genre specific, the section provides possible developmental and psychometric explanations that are in need of examination in future research. Overall, the Simple View of Writing seems to provide a good fitting model that may be used in future evaluations of CLIL writing as a product.

Section 4 uses the Simple View of Writing to examine writing as a cognitive process driven by component skills through two hierarchical regression analyses for narrative and expository writing. The component skills included in the study are spelling ability, which represents the transcription dimension of writing, and vocabulary and oral language ability, which represent ideation. These component skills are considered along with gender, which has been found in previous language research to be a factor in second language writing achievement.

The results of the analysis show that the component skills of the Simple View of Writing may predict variability of writing for primary school grade six CLIL students. However, issues of possible multicollinearity between spelling and vocabulary within the model lead toward multiple explanations of the possible impact of transcription and

139

ideation skills on CLIL writing. In the model including vocabulary, ideation as represented

through vocabulary, is found to be a significant predictor of writing variation, with

transcription as a non-significant predictor. In contrast, with vocabulary removed, the

model shows that transcription and ideation may be equally competing factors in

predicting writing. In regard to the effects of gender, it seems gender differences may exist

when accounting for ideation skills.

Overall, the Simple View of Writing shows promise as a model for examining

CLIL writing as a process of component skills. However, given the multiple interpretations

that emerge from the results of this study, more research is needed to understand how these

component skills interact to predict CLIL writing variation.

5.3. Implications for Future Research

The research presented in this dissertation lays the path for future research using

the Simple View of Writing as a theory for CLIL writing. When considering writing as a

product, future research in CLIL writing should look to account for both transcription and

ideation in evaluations of CLIL writing products. This research has provided a validated

model for general use across various types of writing and groups, but researchers are

encouraged to explore combinations of other observed variables for transcription and

ideation that may be appropriate for different age/language proficiency groups and writing

genres. Further, particularly as it pertains to the dimension of ideation, additional genre-

specific word, clause, and discourse measures should be explored.

Additionally, the validation of the Simple View of Writing as a model of CLIL

writing as a product opens the door to more statistically advanced modeling of CLIL

writing through Structural Equation Modeling (SEM). As research into writing outcomes

through CLIL continues, it is encouraged that future studies consider writing as a complex construct and utilize multiple measures to represent it within statistical models. One criticism presented in Section 2 of this dissertation is the focus on singular measures of writing in past CLIL writing research to determine the effectiveness of CLIL. The Simple View of Writing provides a framework that may lead to a more comprehensive understanding of writing and, subsequently, a more complete understanding of the effectiveness of CLIL.

In regard to using the Simple View of Writing as a theory for examining CLIL writing as a cognitive process, there is much more research needed to understand if and how component skills act together as a process for producing writing in CLIL. First, the role of vocabulary as a measure of ideation, and its relation to spelling and oral language, needs further attention. This could be first done through a study similar to the one in this dissertation, but with an increased sample size, in order to reduce the possible effect of multicollinearity. Another approach would be to use a CFA model, similar to the one in Section 3 but with measures of component skills rather than measures of a written product, to validate these observed skills as measures of transcription and ideation. Should such a CFA model be validated, an SEM model with the latent variables of transcription and ideation as predictors of writing could be evaluated.

Assuming that these component skills are validated as predictors of writing through one or both of the suggested studies above, a subsequent line of research should address how the effect of transcription and ideation skills change over time. Based on previous first language (L1) writing research (Kim, Gatlin, Al Otaiba, & Wanzek, 2017), it is predicted that at lower ages and/or language levels, transcription will take on a strong predictive role,

and as the learner's transcription skills become more automatic, ideation may exert a stronger influence on writing. It is also likely that the orthography of the learner's first language and their strength in first language literacy may determine the degree of influence of transcription and ideation on writing. All of these hypotheses are avenues for further empirical research.

5.4. Implications for CLIL Writing Curriculum and Instruction

This dissertation also has implications for CLIL practitioners. Based on the findings of the studies presented in the previous sections, it is recommended that the Simple View of Writing be used as a theory for driving curriculum and instruction decisions in the CLIL writing classroom. CLIL is intended to be a form of instruction that balances both language and content outcomes. The Simple View of Writing accounts for both with transcription as a construct of language and ideation as a representation of content. When considering the teaching of writing in the CLIL classroom, CLIL practitioners should include instruction and assessment of the written product both in terms of transcription elements, such as spelling and grammar, and ideation, through the incorporation of elements such as content-specific vocabulary, sentence structure, and organization. Researchers have documented the struggle of practitioners to balance language and content in content-based classrooms such as CLIL (Cammarata, 2010). It is hoped that the Simple View of Writing as a framework for informing curriculum design, instruction, and assessment may help practitioners achieve this balance when teaching writing in a CLIL setting.

5.5. Conclusion

Taiwan's *Blueprint for Developing Taiwan into a Bilingual Nation by 2030* (National Development Council, 2018) puts forth ambitious goals for English development of the citizens of Taiwan, particularly in regard to English writing. While there is much more research to be done, the studies in this dissertation have begun to set the foundation for CLIL writing research and instruction in Taiwan, and elsewhere, so that effective writing development through CLIL education may be achieved. It is hoped that the research presented here can inspire future work by researchers and practitioners of CLIL to contribute toward furthering our knowledge base and instructional practices, leading toward increased CLIL writing outcomes for students both in Taiwan and around the world.

5.6. References

Berninger, V. W., & Swanson, H. L. (1994). Modifying Hayes and Flower's model of skilled writing to explain beginning and developing writing. *Advances in Cognition and Educational Practice, 2*, 57-81.

Cammarata, L. (2010). Foreign language teachers' struggle to learn content-based instruction. *L2 Journal, 2*(1), 89-118. https://doi.org/10.5070/l2219063

Juel, C., Griffith, P. L., & Gough, P. B. (1986). Acquisition of literacy: A longitudinal study of children in first and second grade. *Journal of Educational Psychology, 78*(4), 243-255. https://doi.org/10.1037/0022-0663.78.4.243

Kim, Y.-S., Gatlin, B., Al Otaiba, S., & Wanzek, J. (2017). Theorization and an empirical investigation of the component-based and developmental text writing fluency

construct. *Journal of Learning Disabilities, 51*(4), 320-335.

https://doi.org/10.1177/0022219417712016

National Development Council. (2018). Blueprint for developing Taiwan into a bilingual

nation by 2030. Retrieved from

https://www.ndc.gov.tw/en/Content_List.aspx?n=D933E5569A87A91C&upn=963

3B537E92778BB