

Copyright
by
Hui-Chun Yang
2009

**The Dissertation Committee for Hui-Chun Yang Certifies that this is the approved
version of the following dissertation:**

**Exploring the Complexity of Second Language Writers' Strategy Use
and Performance on an Integrated Writing Test
Through Structural Equation Modeling and Qualitative Approaches**

Committee:

Lia Plakans, Supervisor

Elaine K. Horwitz

Diane L. Schallert

Randy Bomer

Brandon K. Vaughn

**Exploring the Complexity of Second Language Writers' Strategy Use
and Performance on an Integrated Writing Test
Through Structural Equation Modeling and Qualitative Approaches**

by

Hui-Chun Yang, B. A.; M. A.

Dissertation

Presented to the Faculty of the Graduate School of

The University of Texas at Austin

in Partial Fulfillment

of the Requirements

for the Degree of

Doctor of Philosophy

The University of Texas at Austin

May, 2009

Dedication

To my loving parents,

Chun-Cheng Yang and Mei-Li Chen,

who passed on to me an intellectual curiosity of the world around me
and an eternal passion for life

Acknowledgements

Completing this dissertation would not have been possible without the help and support of numerous individuals and a research grant. First and foremost, I give my heartfelt thanks to Dr. Lia Plakans, my supervising professor. She is the best mentor and advisor that any graduate student could hope for. She introduced me to the world of language assessment, inspired me to reach beyond my comfort zone, and challenged me to become a better thinker and a researcher. Her insightful comments on the dissertation and uncompromising high standards in research make this dissertation a research study of which I can be proud.

I am extremely thankful to the members of my committee. A special word of gratitude goes to Dr. Elaine K. Horwitz for her full support and unfailing enthusiasm that guided me throughout this process and for her thought-provoking questions concerning planning and implementation of this study. Dr. Diane L. Schallert, who demystified psychological processing during language comprehension and production in her lectures at UT, gave me the confidence to pursue this study. I also thank both Drs. Horwitz and Schallert for their patience and good cheer as we worked together for months toward the development of questionnaire item bank.

My deep gratitude also goes to Dr. Randy Bomer, whose class, discussions about literacy and multimodality issues, and comments on different drafts of the proposal and dissertation helped me shape the outcome of the study. I owe thanks to Dr. Brandon K. Vaughn for his willingness and patience to respond to my many queries regarding my statistical analyses and for his support throughout.

Additional thanks go to all the UT students for participating in this study, to the ESL/EFL instructors and raters that kindly scored and provided information about the participants in the study, and to my colleagues Heng-Tsung Danny Huang and Pei-Yu Wang for their constant assistance with data collection and analyses.

I am most fortunate to have been sounded by loved ones whose support has been unwavering from the beginning. I would like to extend my gratitude to my friends in the Foreign Language Education program at UT. I am indebted to my parents for their never-ending love, encouragements, and sacrificing supports in making my professional goals happen. Without them, it would have been so much harder to accomplish this dissertation.

This dissertation was funded by a TOEFL Small Grant for Doctoral Research in Second or Foreign Language Assessment. I would like to thank Educational Testing Service (ETS) for its generous financial support. I am also grateful to the University of Texas for the University Fellowship and to the FLE program for its support, without which my course of study and dissertation writing would not have been possible.

**Exploring the Complexity of Second Language Writers' Strategy Use
and Performance on an Integrated Writing Test
Through Structural Equation Modeling and Qualitative Approaches**

Publication No. _____

Hui-Chun Yang, Ph. D.

The University of Texas at Austin, 2009

Supervisor: Lia Plakans

Integrated writing tasks that combine reading, listening, and writing have become increasingly popular in assessing academic writing. These tasks are seen to offer more authenticity, improve fairness, and provide positive washback effects of the test on learning and teaching of English around the globe. However, the integrated nature of these tasks can pose some issues, such as construct-related validity and verbatim source use. Given that the inferences made from test scores depend upon the construct of the measure, it is important to have a working knowledge of how strategies are used on integrated writing tests as part of the process of construct validation.

This study investigates the relationship between second language writers' strategy use and performance on an integrated reading-listening-writing test using structural equation modeling and qualitative approaches. Data were collected from 161 non-native English-speaking students. The students first took an integrated reading-listening-writing test and followed by a strategy inventory on how they thought while completing the test.

Twenty students, ten in the high-performance group and ten in the low-performance group, participated in a retrospective interview.

Exploratory factor analysis (EFA) was used to identify the clusters of items based on three hypothetical factors: Rhetorical, Self-Regulatory, and Test-Wiseness Strategy Use. Confirmatory factor analysis (CFA) was then utilized to test the hypothetical relations between observed and latent variables. Subsequently, structural equation modeling (SEM) was used to model the relationship between students' self-reported strategy use and their test performance. The data collected from retrospective interviews, an open-ended questionnaire, and planning sheets were analyzed to triangulate quantitative results and provide supplementary information in interpreting the quantitative data. The study illuminates the nature of integrated writing strategy use, the nature of integrated writing performance, and the relationship between strategy use and performance on an integrated reading-listening-writing test. The results of the study have implications for second language writing assessment and instruction as well as theory in second language academic writing.

Table of Contents

List of Tables	xiii
List of Figures	xv
Chapter 1 INTRODUCTION.....	1
Background	1
Advantages and Disadvantages of Integrated Writing Tasks	3
Construct Validation in Integrated Writing Tests	5
Strategies in Integrated Writing Tests.....	6
Purpose of the Study	7
Significance of the Study	7
Chapter 2 REVIEW OF THE LITERATURE.....	9
Assessing Academic Writing.....	9
Studies on Processes and Strategies of Integrated Writing.....	11
Distinguishing processes and strategies.....	11
Writing processes and strategies.....	13
Integrated writing and test-taking strategies	16
Integrated writing strategies.....	17
Test-taking strategies	23
Studies on Products of Integrated Writing.....	26
Scoring systems for academic integrated writing tests	26
Textual features and writing test scores.....	30
Chapter 3 RESEARCH METHODS.....	34
Participants.....	34
Recruitment of test takers	35
Demographic data	36
Measurement Instrument and Materials.....	39
The reading-listening-writing integrated writing task	40
The integrated writing scoring rubrics.....	40

The strategy inventory for integrated writing	42
The background questionnaire	45
Study Variables	46
Rhetorical strategy use	47
Self-regulatory strategy use	48
Test-taking strategy use	49
Integrated writing test performance	50
Data Collection	51
Procedures	51
Scoring	52
Data Preparation	53
Data Analysis	54
Computer equipment and software	56
Overview of statistical procedures	56
Descriptive statistics	57
Reliability analyses	57
Exploratory factor analyses	57
Structural equation modeling	58
Retrospective interviews	62
Summary	64
Chapter 4 RESULTS	65
Item-level Exploratory Factor Analyses	65
The strategy inventory for integrated writing	66
Distributions and reliability	67
Item-level EFAs for the strategy inventory	69
The integrated reading-listening-writing test	74
Distributions and reliability	75
Item-level EFAs for the integrated writing test	76
Summary	77
Modeling Integrated Writing Strategy Use and Test Performance	78
The strategy inventory for integrated writing	79

Distributions and reliabilities	79
Confirming the dimensionality of the strategy inventory for integrated writing	81
Discussion of quantitative results: Model 1.3	89
Discussion of qualitative results	91
Summary	100
The integrated reading-listening-writing test	102
Distributions and reliabilities	102
Confirming the dimensionality of the strategy inventory for integrated writing	103
Discussion of quantitative results: Model 2.1	107
Test takers' discussion of source use	108
Summary	110
Testing latent variable models	110
The hypothesized model: Model 3.1	111
The estimation results for Model 3.1	112
The hypothesized model: Model 3.2	113
The estimation results for Model 3.2	114
The hypothesized model: Model 3.3	115
The estimation results for Model 3.3	116
The hypothesized model: Model 3.4	117
The estimation results for Model 3.4	118
The hypothesized model: Model 3.5	119
The estimation results for Model 3.5	120
Comparing competing models	121
Discussion of the Full Latent Model 3.5	122
The integrated writing strategy use model	122
The integrated writing test performance model	125
The relationship between integrated writing strategy use and test performance	126
Summary	131

Chapter 5	DISCUSSION	133
	The Research Questions.....	133
	Research question 1	133
	Research question 2	138
	Research question 3	139
	The structure of integrated writing strategy use	139
	The structure of integrated writing test performance.....	140
	The relationship between strategy use and test performance....	141
	Implications.....	143
	Theoretical implications.....	143
	Methodological implications	145
	Pedagogical implications	146
	Limitations	148
	Future Research	149
	Conclusions.....	151
Appendix A	Recruitment flyer.....	153
Appendix B	Integrated writing scoring rubrics	154
Appendix C	Strategy inventory for integrated writing	156
Appendix D	Background questionnaire.....	159
Appendix E	Interview questions.....	161
References.....		162
Vita		171

List of Tables

Table 3.1:	Participants by native language	37
Table 3.2:	Basic information on interviewees in high-performance group	38
Table 3.3:	Basic information on interviewees in high-performance group	39
Table 3.4:	Taxonomy of the strategy inventory for integrated writing.....	50
Table 3.5:	Procedures for data collection.....	52
Table 3.6:	Symbols used in SEM.....	60
Table 4.1:	Distributions for the integrated writing strategy use variables	68
Table 4.2:	Reliability estimates for the integrated writing strategy use variables	69
Table 4.3:	Variables dropped and recoded from the original strategy inventory...	71
Table 4.4:	EFA results for the strategy inventory for integrated writing: Oblimin rotation	73
Table 4.5:	Composites for the integrated writing strategy use variables	74
Table 4.6:	Distributions for the integrated writing test	76
Table 4.7:	Reliability estimates for the integrated writing test	76
Table 4.8:	Results of EFA for integrated writing ability variable: Oblimin rotation	77
Table 4.9:	Distributions for the integrated writing strategy use variables	80
Table 4.10:	Reliability estimates for the integrated writing strategy use variables	81
Table 4.11:	Comparison of fit indexes for Model 1.2 and Model 1.3.....	89
Table 4.12:	Distributions of the integrated writing performance variables	103

Table 4.13: Reliability estimates for the integrated writing performance	103
Table 4.14: Comparison of fit indexes for Model 2.1 and Model 2.2.....	107
Table 4.15: Comparison of fit indexes for Model 3.2–3.5.....	122

List of Figures

Figure 2.1: The cognitive process model of the composing process.....	14
Figure 2.2: Composing process for reading-to-write test tasks	22
Figure 3.1: A flow chart of statistical procedures used in this study	62
Figure 4.1: Initially-hypothesized model of integrated writing strategy use: Model 1.1	83
Figure 4.2: Hypothesized model of integrated writing strategy use: Model 1.2.....	86
Figure 4.3: Hypothesized model of integrated writing strategy use: Model 1.3.....	88
Figure 4.4: Initially-hypothesized model of integrated writing performance: Model 2.1	104
Figure 4.5: Hypothesized model of integrated writing performance: Model 2.2.....	106
Figure 4.6: Initially-hypothesized model of the relationship between strategy use and test performance: Model 3.1	112
Figure 4.7: Hypothesized model of the relationship between strategy use and test performance: Model 3.2.....	114
Figure 4.8: Hypothesized model of the relationship between strategy use and test performance: Model 3.3.....	116
Figure 4.9: Hypothesized model of the relationship between strategy use and test performance: Model 3.4.....	118
Figure 4.10: Hypothesized model of the relationship between strategy use and test performance: Model 3.5.....	120

Figure 4.11: Integrated writing strategy use model.....	125
Figure 4.12: Integrated writing test performance model.....	126
Figure 4.13: Model of the relationship between strategy use and test performance	131
Figure 5.1: Composing strategies for a reading-listening-writing task	135

Chapter 1

INTRODUCTION

BACKGROUND

In the universities and colleges of English-speaking countries, academic writing is a core competency on which intellectual and disciplinary development is built. For non-native English speakers, poor writing skills can lead to lost opportunities for significant construction and exchange of knowledge, and worse, failure in academic studies. Because of the critical role of academic writing, most standardized university admission tests for international students incorporate writing as an essential test component. Considering that a great deal of writing expected in academic contexts requires students to identify, synthesize, connect and manipulate information derived from source materials, a number of commercial language tests such as the Test of English as a Foreign Language (TOEFL) have started to include reading-listening-writing tasks in their assessment batteries. This integrated form of writing assessment is considered to be more authentic than a conventional independent writing task (Lewkowicz, 1997; Wesche, 1987) and can possibly bring about positive washback¹ effect (Cumming, Grant, Mulcahy-Ernt, & Powers, 2004; Weigle, 2004).

While an integrated writing task may be a measure with many strengths, some programs, such as the International English Language Testing Service (Chargé & Taylor, 1997) and institutions (e.g. the University of Texas and the University of Iowa) have concerns about the construct validity of such tasks. Reading-writing or reading-listening-writing tasks have been criticized for confounding the psychometric structure of an

¹ Washback effect refers to the influence of testing on teaching and learning language.

assessment by involving reading and/or listening skills with writing skills. The mutually interdependent nature of the constructs makes test score interpretation and use difficult. Test developers and users need to know more about the construct being measured in order to understand test takers' academic writing proficiency and to use these scores for making admission or placement decisions.

Previous validation studies on integrated writing tasks have analyzed different aspects of construct validity such as task content, discourse characteristics of written responses, rater reliability and scoring criteria (Cumming, Grant, Mulcahy-Ernt, & Powers, 2005; Cumming et al., 2005; Lee & Kantor, 2005; Lewkowicz, 1994; Messer, 1997; Watanabe, 2001). These studies have only examined the written responses derived from the tasks; however, investigating test takers' strategy use in response to these tasks and how their strategy use affects test performance provide evidence for the substantive aspect of construct validity. In other words, it allows examination of whether the reported strategies reflect the types of strategies desired by the test developers. As Cohen (1998) notes, research on test-taking strategies helps testers identify the inherent process whereby test takers arrive at answers, distinguish whether test-takers' interaction with task characteristics affect performance, and thus minimize the construct-irrelevant variance. Chapelle (1999) suggests that the integration of validity evidence from different sources strengthens a validity argument: "A validity argument should present and integrate evidence and rationales from which a validity conclusion can be drawn pertaining to particular score-based inferences and uses of a test" (p. 263).

In an effort to contribute to the validation argument for integrated reading-listening-writing tasks, this study investigates the nature of integrated writing strategy use, the nature of integrated writing test performance, and the relationship between strategy use and test performance. Data were collected and integrated from three sources:

test takers' written responses, strategy inventory data, and qualitative data. Exploratory factor analysis, confirmatory factor analysis, and structural equation modeling were utilized to investigate these issues. The results of this study provide insight into the underlying constructs targeted by these tasks. This information is useful in the development of test tasks and the interpretation and use of test scores.

This chapter presents a rationale for the proposed study by addressing advantages and disadvantages of integrated writing tasks, construct validation in integrated writing tests, strategies in integrated writing tests, the purpose of this study, and research questions.

ADVANTAGES AND DISADVANTAGES OF INTEGRATED WRITING TASKS

“Writing from sources,” or “discourse synthesis,” is an academic writing task found at both the college and university levels. This is a hybrid task that requires students to select, organize, and connect content from multiple source materials as they compose their own text (Bracewell, Frederiksen, & Frederiksen, 1982; Spivey, 1997). In recent years, a number of commercial language tests (e.g. TOEFL iBT, Canadian Academic Language Assessment, Test of English for Educational Purposes, and Certificate of Proficiency in English) have included reading-writing and reading-listening-writing tasks in their assessment batteries. This new form of assessment task is incorporated as part of writing tests because it is reflective of the real use of language that occurs in academic contexts. As noted in the framework for the TOEFL 2000 (Cumming, Kantor, Powers, Santos, & Taylor, 2000), authenticity is the major rationale behind the inclusion of integrated writing tasks in the new TOEFL. The authenticity of these tasks also allows test users and researchers to assume that this type of assessment may improve predictive validity (Lewkowicz, 1997; Wesche, 1987) and bring about a positive washback effect (Cumming, Grant, Mulcahy-Ernt, & Powers, 2004; Weigle, 2004).

Another reason for using integrated writing tasks in assessment is accessibility or fairness. Compared to independent writing tasks where test takers write in response to an essay topic, integrated writing tasks provide test takers with topical knowledge so that they are less likely to be disadvantaged by a lack of background knowledge. Each test taker will have something to write about regardless of their individual, cultural and educational backgrounds (Read, 1990).

Despite the advantages of integrated writing tasks, some critics or researchers questioned their construct validity. Charge and Taylor (1997) reported that the use of reading-writing tasks increased the possibility for confusing the assessment of reading ability with the assessment of writing ability. Enright, Bridgeman and Cline (2002), as cited by Asención (2004), found that the correlations were as high as .80 between scores for the prototype integrated writing tasks for the new TOEFL and reading scores in the paper-based TOEFL. With the increasing use of reading-listening-writing tasks in commercial and in-house language tests, the construct validity of such tasks is worth researching in more depth. However, there is a lack of information about the relationships among reading, listening, and writing in integrated writing tasks.

Verbatim source use is another issue of concern. Researchers on reading-writing assessment tasks have found that students with inadequate reading and summary skills copied source materials wholesale (Lewkowicz, 1994). When evaluating written responses with copied phrases, the reliability of the ratings and the validity in interpreting the test scores can drop. More validation studies are needed to clarify the interdependent construct of integrated writing tasks and to address issues of evaluation given source information use.

CONSTRUCT VALIDATION IN INTEGRATED WRITING TESTS

The proliferation of integrated writing tests increasingly requires test developers to examine various aspects of the integrated writing construct. According to Messick (1989), validity evidence should be collected in terms of six perspectives: content, substantive, structural, generalizability, external, and potential consequences. A number of integrated writing tasks studies have addressed the aspect of content validity by surveying teachers' opinion about the task representativeness of a measure (Asención, 2004; Cumming, Grant, Mulcahy-Ernt, & Powers, 2005). To collect evidence for the substantive aspect of the validation process², several researchers have collected think-aloud protocols to determine whether the process or strategies in which test takers engaged reflect the proposed construct (Asención, 2004; Esmaeili, 2002; Plakans, 2007). New scoring rubrics were developed to examine the structural aspect³ of the construct (Cumming et al., 2005; Watanabe, 2001). The generalizability aspect of the construct has been discussed via examining the consistency of scoring across raters, test takers, and tasks (Asención, 2004). The external validity concerning whether the integrated writing is a construct or several constructs has also been explored (Trites, 2000; Watanabe, 2001). In terms of potential consequences, most studies collect evidence from teachers or students to identify the impact of the inclusion of integrated writing tasks in language tests on teaching and learning in classrooms (Wall & Horák, 2006). These studies also help verify the washback effect of integrated writing tasks (Cumming, Grant, Mulcahy-Ernt, & Powers, 2004; Cumming, Kantor, Powers, Santos, & Taylor, 2000).

² The process of validation includes (1) a validation argument that claims intended and unintended interpretation of test scores, and (2) a validation study that collects evidence to support the claims and reject the counterclaims (Bachman, 2004).

³ Structural aspects of validity relates to validation of the internal structure and scoring criteria of the measure. The interaction among subscales or items should be consistent with the structure of the construct domain.

To support the validity argument of integrated reading-listening-writing tasks, this study investigates the nature of integrated writing strategy use, the nature of integrated writing test performance and the relationship between strategy use and test performance. To gain a better understanding about the construct of test tasks, Bachman (2002, p. 470) suggests (a) the collection of “data on test-takers’ responses to individual assessment tasks,” (b) the collection of “data on processes or strategies that test-takers use in responding to assessment tasks,” and (c) the utilization of “alternatives to current measurement models for modeling quantitative data.” Based on these suggestions, this study collects test takers’ written texts, and integrated writing strategies elicited by a strategy inventory. Finally, structural equation modeling (SEM) is used for modeling the association between strategy use and integrated writing test performance.

STRATEGIES IN INTEGRATED WRITING TESTS

Since the 1990s, there has been a growing interest in the role of strategy in the process of L2 testing, acknowledging strategy use as an important source of variation in language test performance (Anderson, 1991; Anderson, Bachman, Perkins, & Cohen, 1991; Bachman, 1990; Cohen & Upton, 2007; Homburg & Spaan, 1981; Phakiti, 2003; Plakans, 2008; Purpura, 1997; 1998). Strategy use during tests is thought to be different from strategy use during learning in many aspects. Rather than focusing on strategies used to enhance language learning process, strategies used to respond to language tests are used to manage test tasks at hand with an aim to achieve task goals or requirements.

In relation to writing, for instance, some test takers may use test management strategies to help them handle the task in time. Others may use test-wiseness strategies (e.g. practicing writing essays in a particular structure for test preparation) as short-cuts to compensate for a lack of their actual writing ability. In several studies of integrated writing test tasks, test takers were found to copy material directly from source materials

in their response (Braine, 2001; Cumming et al., 2005; Lewkowicz, 1994; Watanabe, 2001), and thus raters had difficulty telling whether the writer actually understood the sources. The variation in test scores owing to the use of test-wiseness strategies is a construct-irrelevant variance (Crocker, 2006; Haladyna & Downing, 2004) that may decrease the precision of scores and contaminate the quality of decisions the test purports to make. Therefore, the data on strategy use during tests are useful in examining the validity of the test tasks (Cohen, 1998). To find out what an integrated reading-listening-writing test is actually measuring, it is crucial for test users and developers to understand the types of strategies test takers use in responding to these tasks.

PURPOSE OF THE STUDY

The purpose of this study is to examine the relationship between strategy use and performance on an integrated reading-listening-writing test, specifically with regard to the influence that test takers' self-reported integrated writing strategies might have on different aspects of their test performance. This study first examines the nature of the rhetorical, self-regulatory, and test-wiseness strategies as measured by a writing strategy inventory, the nature of the integrated writing test performance as measured by an integrated reading-listening-writing task, and finally explores how integrated writing strategy use relates to test performance on an integrated writing test. The goal of the current study is to define the construct validity of integrated reading-listening-writing tasks, and to ultimately contribute to second language academic writing assessment and instruction.

SIGNIFICANCE OF THE STUDY

This study has theoretical, methodological and pedagogical implications for the fields of second language assessment and second language acquisition. From a theoretical perspective, this study identifies the underlying structure of the integrated writing test by

examining whether or not the types of strategies used by test takers reflect the construct discussed in literature on integrated writing. Since the task used in the study was an integrated reading-listening-writing task, this information is useful for the development and validation of similar integrated writing tasks such as the integrated writing section of the new TOEFL. In brief, the current study has implications to the validity of test scores from which inferences are drawn and decisions are made about language learners.

Methodologically, the study utilizes SEM as the primary analytic tool because it accounts for measurement errors in parameter estimations. Several testing researchers (e.g., Bachman, 2002; Kunnan, 1998) have encouraged the use of SEM to investigate validity issues of language assessment. However, most process or strategy studies on integrated writing are small-scale case studies designed to characterize writing behaviors of writers. Few, if any, studies use powerful statistical tools to estimate the underlying construct of integrated writing tasks. The present study fills this gap by using SEM to examine the relationship between test takers' strategy use and reading-listening-writing test performance. Finally, from a pedagogical perspective, this study provides suggestions concerning second language academic writing instruction and writing strategy instruction.

Chapter 2

REVIEW OF THE LITERATURE

This chapter begins by addressing writing assessment for academic purposes, and research on integrated writing tests. Studies on products and processes or strategies of integrated writing will also be discussed. Lastly, test-taking strategies used to respond to language tests and integrated writing tests will be described. The review of literature focuses on the significance of the proposed research questions.

ASSESSING ACADEMIC WRITING

Many large-scale writing assessments have shifted away from discrete-point assessments toward the more authentic performance-based assessments (Grabe & Kaplan, 1996; Hamp-Lyons, 2001). Rather than demonstrate their knowledge about writing, test takers are required to perform actual writing tasks. Performance assessments measure students' ability to respond to real-life tasks, thus they have the potential for predicting students' language use in real-life situations (Jones, 1985). However, several researchers are concerned with the lack of empirical evidence that supports the validity claims of second language performance-based assessments (Bachman, 1990; McNamara, 1996).

Authenticity of academic writing tests is a central issue in designing L2 writing assessments. Conventionally, writing is measured independent of other language skills. Examinees are required to respond to a writing prompt based on their general knowledge and personal experiences. However, studies examining the types of task performed in academic contexts indicate that writing assignments often include genres in which reading and writing are combined (e.g., Whalen & Menard, 1995). A task that assesses students' ability to simply write about what they know is not reflective of real academic

writing tasks. Because of the current calls for more authentic measures of academic skills, more and more language assessments have included integrated writing tasks in their assessment batteries. The integrated writing tasks resemble real-life academic writing tasks that require examinees to integrate multiple language skills. The Internet-based TOEFL examination, for example, asks examinees to comprehend an academic lecture and a reading passage, and to then produce a written text to compare the stimulus listening and reading materials.

There are several reasons for promoting and incorporating integrated writing tasks. First, recreating writing tasks similar to the ones required in real-life academic contexts enhances predictive validity (Wesche, 1987). The test results of such tasks may provide useful information for test score users in understanding how well examinees may perform when carrying out academic writing tasks in the future. Moreover, the relevance and authenticity of integrated writing tasks to the real-life tasks is thought to provide a positive washback effect of testing on second/foreign language instructional programs (Weigle, 2004; Wesche, 1987). The use of integrated writing tests allows administrators, teachers and students to reinforce the learning and teaching of knowledge required in academic contexts and thus prepares students for the real world beyond tests. In addition to the support of stimulus materials, the integrated tasks are developed with a goal to achieve test fairness (Read, 1990; Weir, 1993). When source materials are provided, examinees are less likely to be affected by a lack of information about the topic.

In an effort to build a validity argument for integrated reading-listening-writing tests, a number of studies have garnered evidence (e.g., task representativeness, test use and consequences, linguistic knowledge and strategies) to ensure the inferences drawn from the measure or test score are warranted. In the process of validation, two types of data, writing products and process or strategies, are commonly collected.

STUDIES ON PROCESSES AND STRATEGIES OF INTEGRATED WRITING

The construct of integrated writing appears to be more complex than that of writing without sources because it involves separate language skills such as reading, listening and writing. For the purpose of test development, use and interpretation, it is necessary to examine the processes or strategies in which test takers engage. As suggested by Bachman (2002), researchers should accrue data on test takers' processes or strategies in response to performance-based assessment tasks to support test use and inferences. Cumming et al. (2005) also suggest that process studies provide important construct information about tasks, such as the relationship between writing processes and products and the way writing processes and products distinguish proficient and less proficient examinees. In recent years, many researchers have started examining the processes or strategies test takers use to complete integrated writing assessment tasks. This section of the chapter reviews research on distinguishing processes and strategies, writing processes and strategies, and integrated writing and test-taking strategies

Distinguishing processes and strategies

Process is typically seen as an umbrella term that entails various human mental activities. Strategy generally refers to an individual's conscious and willful management of specific techniques used to facilitate achieving a particular goal. In a language learning context, if a learner apply certain tactics habitually or subconsciously, the behavior is viewed as a common process rather than a strategy (Cohen, 1998). While processes are largely automatic and demand little attentional resources, strategies act as a compensatory system. This system exerts conscious control over the general processes with a goal of maximizing effectiveness in performing tasks. For instance, when a learner's general approaches or techniques are not adequate to resolve an issue or a problem, intentional strategies are deployed to moderate and govern the situation. According to strategy

researchers (Chamot, 1987; Cohen, 1998), the primary distinctions between processes and strategies center on automaticity and purposefulness. The two terms, process and strategy, will be used interchangeably in this chapter, which refers to a possible swap of these operations during writing.

Most discussions on SLA processes focus on the role of learning strategies. In other words, L2 researchers are interested in what a good language learner does to solve a learning and communication problem. A wide range of taxonomies of language learning strategies have been published. Two types of classifications have received considerable attention: Oxford's (1990) strategy system and O'Malley and Chamot's (1990) strategy system. Oxford (1990) distinguishes between direct and indirect strategies: Direct strategies relate to target language use (subcategorized into memory, cognitive and compensation strategies), whereas indirect strategies support language learning (further divided into metacognitive, affective and social strategies). In O'Malley and Chamot's (1990) strategy system, they distinguish three general types of learning strategies: cognitive, metacognitive, and social-affective strategies with various sub-strategies in each type respectively.

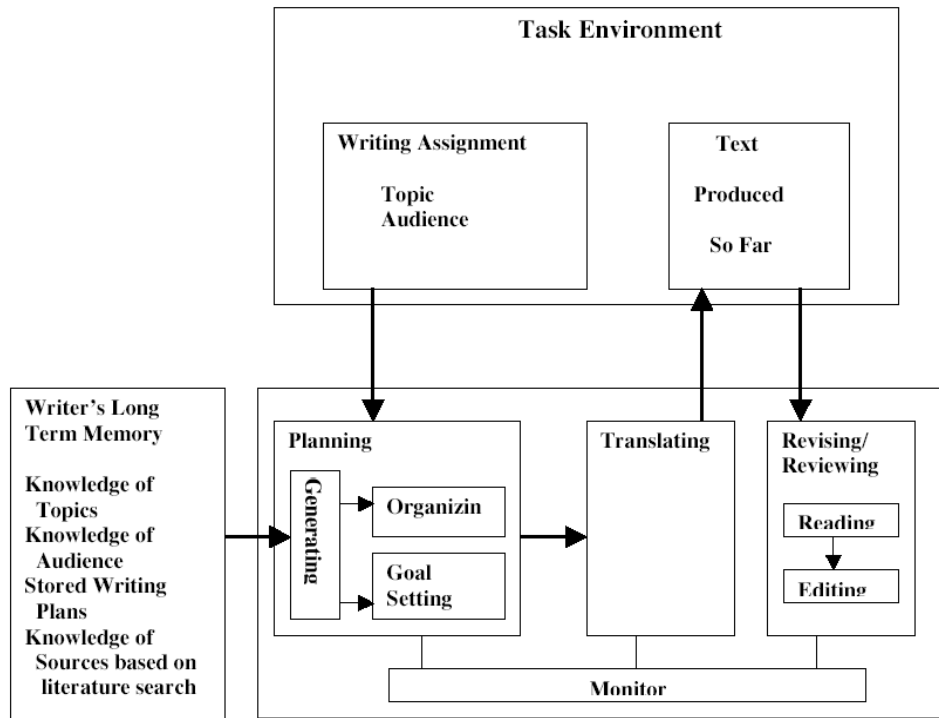
While these taxonomies are useful in examining a variety of language learning phenomena, it is unclear whether they are appropriate to be applied across different L2 situations. As individuals try to solve problems arising from a domain or a particular task, they must utilize domain- or task-specific strategies in order to be effective. Strategy use should be task-bound and context-specific (Glaser, 1984; Pressley, Goodchild, Fleet, Zajchowski, & Evans, 1989). In an L2 strategy study, Chamot, Kupper, Impink-Hernandez (1988) found that successful learners matched different strategies with different tasks, suggesting that task variation plays an important role in the selection of strategy use.

Given the complex nature of strategy use, it needs to be acknowledged that there are a number of possible interactions between strategies and contexts. Therefore, language learning strategies are not identical to language use or test-taking strategies. With these considerations in mind, several researchers (Cohen, 1998; Ellis, 1994; Hsiao & Oxford, 2002) have promoted the development of task- or context-based strategy surveys to elicit a wider variety of strategies that learners use to handle different tasks under different situations.

Writing processes and strategies

To delve into the possible strategies writers use, it is important to have an in depth knowledge of what constitutes the writing process. Most process research is built upon earlier L1 writing models. To date, the cognitive view of the writing processes is the most prevalent framework in L2 research. This strand of research owes a great deal to cognitive psychology. The first cognitive model of writing (see Figure 2.1), proposed by Flower and Hayes (1981), states that writing is a problem-solving, goal-directed and recursive activity. Writers move from writer-based texts to reader-based texts through three interweaving composing elements: the task environment, the cognitive writing process, and the writers' long-term memory. The task environment includes elements exterior to the writer such as writing assignments and texts that writers have produced. The cognitive writing process consists of three cognitive sub-processes: planning, text generation, and revision. The writers' long-term memory consists of the writers' knowledge of topic, knowledge of audience and stored writing plans.

Figure 2.1 The cognitive process model of the composing process (Flower & Hayes, 1981)



Subsequently, Bereiter and Scardamalia (1987) delineated the writing behaviors of novice and expert writers. They propose that novice writers tend to employ the “knowledge-telling model” which involves linear processes of retrieving content knowledge from memory and generating new ideas based on the phrases and sentences written down. Such writers employ few planning, revising or self-regulation techniques during writing. They often neglect the needs of their audience. In contrast, expert writers tend to use the “knowledge-transforming model” where more reflective and recursive composing processes, such as goal-setting and problem-solving, take place.

Hayes (1996) modified the earlier cognitive model (Flower & Hayes, 1981) of writing to present a more encompassing framework—the individual-environmental model. The new model includes two components: task environment and individual. The

task environment describes the social environment and physical environment surrounding the writer. The individual component includes motivation/affect, working memory, long-term memory and cognitive processes. In contrast to the previous model, the new framework integrates the writers' motivation and affect with the cognitive processes due to their important influences on the cognitive processes. The subcategories of cognitive processes—planning, text generation and revision—are replaced by text interpretation, reflection, and text production. The component of revision, within the earlier cognitive writing processes, is replaced by text interpretation. Text interpretation emphasizes the interaction among: writers, source texts, text representations, and task descriptions. Rather than serve as part of working memory, planning, decision-making, and inferencing serve as parts of the reflection process.

Drawing heavily on previous research (Baddeley, 1986; Flower & Hayes, 1980), Kellogg (1996) focuses specifically on the function of working memory in the writing processes. This model consists of three systems—formulation, execution, and monitoring—with two components involved in each system. Formulation entails planning—writers set their goals and outline ideas related to the goal for later translating. Execution involves programming the output of translation using appropriate motor systems such as handwriting or typing. Monitoring refers to reading where a writer comprehends words and sentences in his or her written texts and makes local or global revisions. With multiple systems working simultaneously, individuals have a limited capacity of cognitive resources available. Therefore, they must trade off attention between tasks based on their priority.

While there are several common features shared by these models, each model has a different focus within the writing process. Kellogg's model focuses on the role of working memory. Hayes' individual-environmental model deals with the interaction

among affect, social context, and cognition. Bereiter and Scardamalia's model centers on different approaches skilled and less-skilled writer adopt in a writing task. These models are useful for identifying and interpreting processes or strategies involved in L2 writing because of some similarities in L1 and L2 writing processes. However, to understand the interactions between writers and writing contexts in L2 settings, models concerning L2 writers' characteristics, task characteristics and related factors should be examined. For the purpose of this study, research on integrated writing and test-taking strategies will be reviewed in the following section.

Integrated writing and test-taking strategies

As part of the process of construct validation, it is important to gain a better understanding of how integrated writing and test-taking strategies are used on tests. Bachman and Palmer (1996) have noted, "unless we can demonstrate that the inferences we make on the basis of language tests are valid, we have no justification for using test scores for making decisions about individuals" (p. 95). Cohen (2006) added that test-taking strategy research can provide valuable information about test construct by distinguishing language learner strategies from test-taking strategies. According to Cohen (2006), language learner strategies relate to the ways examinees use their language skills to approach test tasks. In the current study, language learner strategies refer to test takers' repertoire of synthesizing, evaluating and summarizing source texts in responding to the integrated writing test. Test-taking strategies include test-management and test-wiseness strategies. Test-management strategies refer to a set of techniques examinees use to achieve task goals and purposes. In this study, test management strategies refer to approaches test takers use to identify task requirements and respond meaningfully to the task. Test-wiseness strategies are seen as tricks examinees use to arrive at answers to the question rather than cognitive operations expected by test designers. In this study, test-

wiseness strategies refer to techniques for using knowledge of test formats (e.g., source texts) and other extraneous information (e.g., writing templates) to respond to the task without going through the cognitive and linguistic processes desired by test developers.

As indicated at the outset, the current study investigates examinees' use of strategies to validate the integrated format of emergent writing tests. Consequently, a review of research on integrated writing strategies and test-taking strategies will provide insight into the nature of integrated writing.

Integrated writing strategies

Integrated writing tasks require coordination of writing with other skills. Examining writing frameworks alone cannot provide a whole picture of the interdependent processes and strategies involved in reading-writing or reading-listening-writing tasks. Integrated frameworks exploring the interactions of multiple skills are needed to disentangle the nature of integrated writing. According to Grabe (2001), reading-writing tasks engage reader/writers in both reading comprehension (e.g., word recognition; monitoring; making connections between background knowledge and texts) and writing processes and strategies (e.g., planning, revising, monitoring). Although each process/strategy is no different from regular reading or writing processes or strategies, a sum of reading and writing processes/strategies is not sufficient to perform reading-writing tasks effectively. They are merely threshold abilities required to complete integrated writing tasks. An ability to recognize task goals and contexts, manipulate source materials, and integrate source information in writing is essential to successful writing.

The interaction between sources and writing can be illustrated from a constructivist perspective. Constructivism portrays readers as active participants in the creation of meaning and knowledge (Kintsch & van Dijk, 1978). Three key operations

are involved in constructing meaning from sources: organizing, selecting and connecting (Spivey, 1990; Spivey & King, 1989). Constructivist approaches identify instances of organizing, selecting, and connecting by examining textual characteristics of essays from reading-writing tasks (Spivey, 1997). Although these operations have not been researched as strategies, they will be described as strategies in this study in an effort to explain the interactions of literacy skills involved in integrated writing tasks.

The first type of the strategies is organizing. When writers compose from source texts, they use their prior knowledge about general text organization (e.g., discourse cues and patterns) to make meaning of the texts, understand the links between and across meaning units and different sources, and to finally formulate mental representations and structures for their writing (Spivey, 1990, 1997). Although the structures identified in the reading passage may be replicated in the texts created, they are often reconfigured to meet task requirements and goals. The organizing strategies follow the concept that the goals and expectations writers have in mind shape the way they comprehend the source texts and they frame their writing (Schallert, 1987). Similarly, Kintsch's (1998) situation model, drawn from the construction-integration model, elaborates that the information in the text created is largely molded by writers' knowledge-base interpretations.

Selecting is the second operation central to the process of constructing meaning from sources (Spivey, 1990, 1997). In the process of performing integrated tasks, writers use their criteria of relevance to identify key points in sources. They apply varying degrees of attention toward source text given the task goals, demands, purposes or specifications. In a number of summary studies (e.g., Brown & Day, 1983), researchers found when writers summarize sources texts, they manipulate information available by extracting major points and ignoring redundant or trivial information. Kennedy (1985) compared processes and strategies employed by fluent and not-so-fluent readers. She

found that fluent readers actively selected important information by taking notes, and were able to use ideas in the notes for their writing. On the other hand, not-so-fluent readers tended to passively extract information by rereading the sources during writing.

According to Spivey (1990, 1997), connecting strategies are employed by writers who digest source information, link this external information with their previous knowledge, and finally elaborate a combination of these concepts from an integral perspective. The process of connecting can be related to Bereiter and Scardamalia's (1987) "knowledge-telling model" and "knowledge-transforming model." Writers applying knowledge-telling strategies report information from source texts without further illustration and explanation; writers applying knowledge-transforming strategies expand and elaborate known information from source texts through connections to their background knowledge. In reading-to-write tasks, elaborating (i.e., connecting) is considered a major element in the construction of one's own texts (Stein, 1990). However, not all tasks demand background knowledge for content elaboration. For example, summary tasks requiring brief or highly distilled reports of main ideas from source texts do not promote inclusion of previous knowledge for content extension. In the present study, the integrated reading-listening-writing test task requires test takers to write solely based on two provided sources (i.e., one reading passage and one lecture); test takers are not encouraged to incorporate background knowledge or experience in their writing. In order to explain test takers' mental operations for retrieving information from two source texts of different modalities, this study describes connecting strategies as linking ideas across sources.

Spivey's constructivist approach provides insight about the mental operations involved in reading-writing literacy events. However, it fails to address the self-regulatory mechanisms that govern and orchestrate these operations. For the organizing,

selecting, and connecting operations or strategies to occur, writers' metacognition needs to be activated. Metacognitive control allows writers to be aware of: task goals, plans to go about the task, resources available, and strategies they could use to improve task effectiveness. Previous reading-to-write research has studied three major metacognitive operations: monitoring, planning, and evaluating.

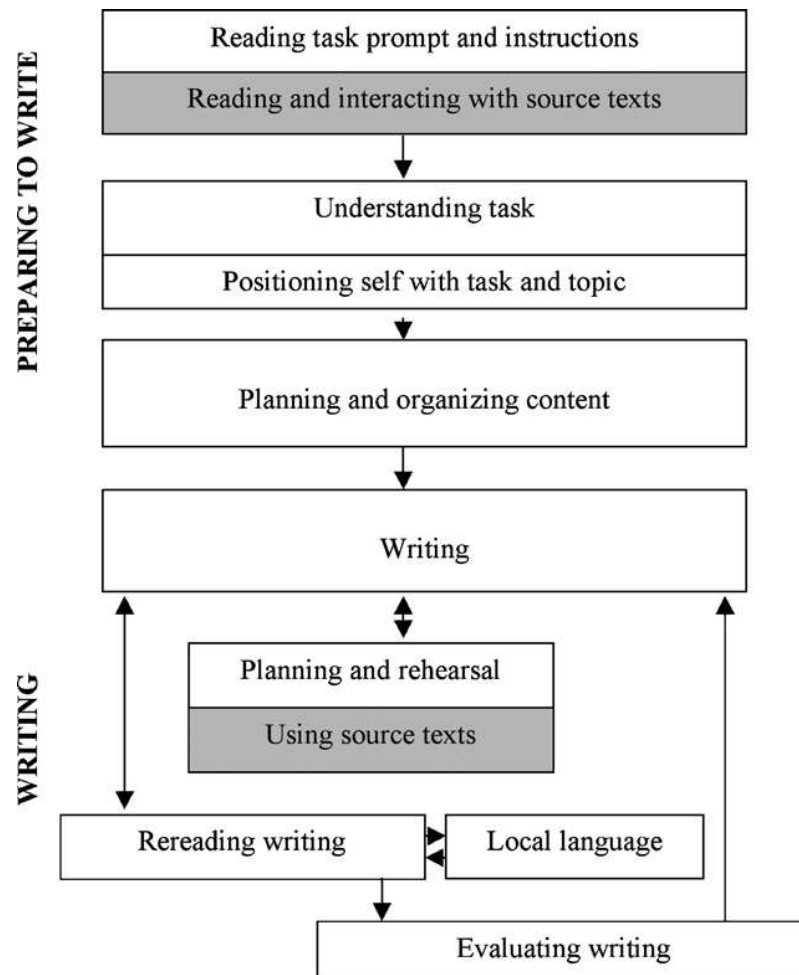
In a validation study, Asención (2004) used think-aloud protocols to analyze native and nonnative English speakers' composing processes in response to two reading-writing tasks. Of all these processes, monitoring appeared to occur most frequently (over 50%) across groups. Planning occurred less frequently (over 20%). Selecting, organizing, and connecting were the least frequently reported processes (below 10%) across groups. Stein (1990) also investigated the roles of planning and monitoring strategies in reading-to-write tasks. Planning is found to be important in constructing texts from sources because it allows students to recognize text features, extract main ideas from source texts and information from background knowledge, and develop a sense of rhetorical purpose. Planning orchestrates the organizing, selecting and connecting strategies described in Spivey's work. On the other hand, monitoring is also considered a central element in reading-to-write tasks. It allows writers to conceptualize a problem-solving situation in which they make decisions about how information from sources should be organized, selected and connected. It also enables writers to handle unexpected issues such as comprehension difficulties or conflicting contextual goals.

Evaluating strategies have also been reported in reading-writing research. Esmaeili (2002) investigated ESL adults' strategy use when the reading and writing on the test were thematically related to each other. Based on interview protocols and retrospective questionnaires, he found that many students reported reconsidering:

previous goals, planned thoughts, written texts and local and global changes undertaken to written texts.

Some studies have evaluated how reading and writing strategies interact with each other in reading-writing tasks. Plakans' (2008) process study identifies three major categories of reading strategies: goal-setting, comprehending the source texts, and mining information for use in writing. Each category constitutes a number of strategies. The occurrence frequency and occurrence time point (prewriting, writing, and revising) of each strategy was reported in detail. To explain how writers moved from one stage to another, Plakans designed a model representing two stages of writing: preparing to write and writing. In the first stage, writers prepare themselves for reading-to-write tasks in a linear process. In the writing stage, writers follow a series of nonlinear processes: planning, rehearsing phrases, rereading source materials, and examining mechanics and language use.

Figure 2.2 Composing process for reading-to-write test tasks (Plakans, 2008)



In a follow-up study, Plakans (2009) investigated the relationship between reading strategies and writing performance. The study reveals that participants in the high-performance group used a wider range of strategies. They also used considerably more global reading strategies, such as mining strategies (e.g., scanning and skimming), to draw information from source texts for their writing. Participants in the lowest performance group, however, used more word-level reading strategies and rarely used

mining strategies. Both of Plakans' studies reveal intricate connections among writers, strategy use, source texts, and task characteristics.

While these studies provide insight into the processes or strategies examinees use for reading-writing tasks, examinees' mental operations in response to reading-listening-writing tasks are relatively unexplored in the area of second language writing and language testing. This study set out to explore test takers' response strategies when interacting with a reading text, a listening text, and their own texts.

Test-taking strategies

A major strand of test-taking strategy research centers on test-management strategies and test-wiseness strategies (Allan, 1992; Cohen, 1998; Cohen & Upton, 2007) as well as the relationship between strategy use and test performance (Phakiti, 2003; Purpura, 1997; Tsai, 2004).

Most studies investigating test-taking strategies focus on strategies examinees use in reading comprehension tests. To examine the impact of test-wiseness on construct validity in reading comprehension tests, Allan (1992) designed a test-wiseness scale that measured EFL/ESL students' use of test-wiseness strategies. The 33-item multiple-choice instrument included stem option cues, grammatical cues, similar options, and item giveaways. In a review of test-taking strategies, Cohen (2006) examined an in-house TOEFL validation study that investigated how test takers at different English proficiency levels interact with integrated reading-writing tests. He also evaluated issues raters had when rating these types of essays. Researchers found that respondents had difficulty providing information about how they transformed reading materials into their own words. Raters were found to have difficulty identifying students' own words versus their verbatim source use. Cohen reported that the results of the study led to the removal of the integrated reading-writing component from the test. The research does not, however,

account for the processes and strategies test takers use to respond to reading-writing tasks, and how strategy use affects their writing performance.

Only a relatively small number of studies have addressed test-management and test-wiseness strategies examinees use in responding to integrated writing test tasks. In a study that focuses on reading strategies examinees use in reading-to-write tasks, Plakans (2009) found that many examinees utilized “goal-setting” strategies—going back to reread and interpret writing instructions to clarify the purpose of writing. Examinees checked source integration and citation in their essays as they reread the writing instructions. These strategies helped them determine how to use the source texts in their writing. They also confirmed that they were completing the reading-writing task appropriately. Studies examining textual features of response essays (Cumming et al., 2005; Lee & Kantor, 2005) reported that more instances of verbatim source use were found in essays responding to reading-writing tasks than listening-writing tasks. Instances of inappropriate source use may reflect some strategies test takers use to improve scores such as copying (Campbell, 1990; Currie, 1998; Pennycook, 1996) and patchwriting strategies (Howard, 1995; Hyland, 2005). These goal setting and source use strategies have not been researched as test-management strategies and test-wiseness strategies, but for the purpose of this study, they will be described as such in order to distinguish these two types of operations from test takers’ repertoire of reading, listening and writing strategies.

Most research studies focus on the relationship between test takers’ strategy use and L2 reading test performance (Phakiti, 2003; Purpura, 1997; 1998). However, the relationship between strategy use and writing test performance is rarely addressed in language testing research. Tsai (2004) explores the cognitive and metacognitive strategies 214 ESL learners used in response to UCLA’s placement writing test. Students were

asked to respond to one of two argumentative essays within 50 minutes. The multi-faceted Rasch measurement, exploratory factor analyses, and confirmatory factor analyses were used to examine the relationships among test takers, raters, writing prompts, and rating scales. The results show that although the metacognitive operations have no direct effects on L2 writing ability, they have positive direct effects on the comprehension and text generating processes. The results also indicate that the more test takers engaged in comprehending processes⁴, the better they were able to perform the task. However, the more test takers used text processing strategies⁵, the worse they performed on timed essay tests. Tsai further emphasizes that only a good control over both cognitive and metacognitive strategies leads to successful essays. The results of the study provide some insight into how different types of strategies affect L2 writing test performance. However, this study focuses on strategy use in writing-only tests. Therefore, no information was provided concerning the interactions between test takers' strategy use and source materials.

The studies reviewed in this section provide valuable information about how test takers use strategies to approach writing-only and reading-writing test tasks. However, most of these studies examine only processes or strategies examinees use in response to test tasks. As Chapelle (1999) noted, a validation argument can be strengthened when different sources of evidence are combined. An investigation that accounts for both writing strategies and products will provide a clearer picture about the construct of integrated writing.

⁴ Tsai (2004) specified that comprehending processes include strategies of inferencing, clarifying, and invoking prior knowledge.

⁵ Text processing strategies include organizing information (e.g., take notes), composing and revising in Tsai's (2004) framework.

STUDIES ON PRODUCTS OF INTEGRATED WRITING

The inclusion of integrated writing tasks in language tests provides information about examinees' abilities to comprehend source materials presented in different modalities (e.g., reading passage and lecture) and abilities to evaluate sources, analyze, compile and synthesize ideas into written works. Because integrated writing tests assess multiple and complex literacy skills, it is important to differentiate the aspects or features of integrated writing products so that test scores can be properly interpreted and used. Previous research has shown that response essays are commonly measured in terms of content, organization and language use. Examinees' use of source materials are either not included in the scoring systems or measured inconsistently across different tests. This section of the chapter reviews the existing scoring systems for academic integrated writing. It further discusses the relationship between textual features of writing and test scores.

Scoring systems for academic integrated writing tests

Integrated writing tasks have been increasing in popularity in commercial, in-house tests as well as research tests. Some commercial language tests (e.g., Certificate of Proficiency in English – COPE) include reading-writing tasks to measure English learners' ability to understand a reading passage and express ideas in writing. Other tests (e.g., Ontario Test of English as a Second Language – OTESL, Test of English as a Foreign Language Internet-based test – TOEFL iBT, and Canadian Academic Language Assessment – CAELA) incorporate reading-listening-writing tasks to measure English learners' comprehension ability for two forms of source materials and their ability to compose from sources in a logical and accurate fashion. Regardless of source text modalities, source texts are usually thematically-related so that examinees can write based on an integration of ideas drawn from sources.

Although all integrated writing tasks expect some responsible use of source information, the degree to which test takers must manipulate source information differs from test to test. Tests such as TOEFL iBT⁶ and OTESL⁷ require examinees to synthesize information from both a reading passage and a lecture. Similarly, the CAELA⁸ provides an essay topic at the beginning of the test so that examinees can keep it in mind while reading and listening to source texts. Examinees are then asked to use information from the reading and listening sections⁹ of the test to support their positions about the topic. Varied from other tests reviewed in this section, the COPE¹⁰ writing test requires examinees to write an essay based on the ideas in the summary examinees have completed as part of the reading section of the test.

With regard to the scoring rubrics used to assess source-based writing, these tests launch similar criteria to assess the quality of essays. Holistic or analytic, the rubrics are based on three aspects of writing: relevance and adequacy of content, essay organization, and accuracy of language use. The scoring rubrics for the TOEFL iBT and CAELA tests evaluate the extent to which the reading and lecture are integrated in examinees' essays, though not for all score levels. In the TOEFL iBT scoring rubrics¹¹, an essay at the score level of "5" is described as a response with the following features:

A response at this level successfully selects the important information from the lecture and coherently and accurately presents this information in relation to the relevant information presented in the reading. The response is well organized, and occasional language errors that are present do not result in inaccurate or imprecise presentation of content or connections.

⁶ See TOEFL practice test in <http://www.ets.org> – Retrieved April 6, 2009

⁷ See more information from Wesche (1987).

⁸ See writing section in CAELA <http://www.cael.ca/taker/writingresponse.shtml> – Retrieved April 6, 2009

⁹ Examinees are asked to complete reading and listening comprehension sections of the test before they respond to the writing section of the test.

¹⁰ See COPE writing test in <http://www.copetest.com/prepform.htm> – Retrieved April 6, 2009

¹¹ See TOEFL integrated writing rubrics in http://www.ets.org/Media/Tests/TOEFL/pdf/Integrated_Writing_Rubrics_2008.pdf – Retrieved April 6, 2009

Examinees who “omit one major key point made in the lecture” receive a score of “3.” These criteria indicate that a threshold listening comprehension ability is required to complete the task successfully; a lack of listening comprehension is guaranteed to result in a lower score.

In the CAELA test, a writer at the band “70” is considered an “adept writer” who “responds readily to the demands of the topic and presents information clearly and logically; uses the reading passages and the lecture effectively to support the thesis; demonstrates understanding of the requirements of academic writing.”¹² The CAELA test prioritizes both reading and listening comprehension abilities for task completion.

In other tests such as the OTESL and the COPE test, the interactions among sources and writing are not emphasized in the analytic scoring rubrics. Essays at the score level of “5” in terms of content show “systematic development of topic with well-structured main and subordinate themes and relevant supporting statements.”¹³ In the COPE test, essays that score “3” show

A superior ability to communicate effectively; the ideas presented can be followed easily; the main ideas and supporting details can be easily identified and are logically and coherently developed in an appropriate organizational framework; the writing shows a range of sentence types and vocabulary; occasional minor errors in grammar or word choice do not obscure meaning.¹⁴

These scoring rubrics focus how source information should be integrated in response essays instead of addressing issues of plagiarism.

A number of in-house English tests include integrated tasks in which writing is based on reading passages (e.g., exit test, Braine, 2001; UAWA, Feak & Dobson, 1996), or reading passages and lectures (e.g., UIUC-EPT, Cho, 2003). In these tests, examinees are required to write from the sources. The scoring rubrics of these tests focus on

¹² See CAELA scoring criteria for writing in <http://www.cael.ca/pdf/C4.pdf> – Retrieved April 6, 2009

¹³ See scoring criteria in Wesche’s (1987, p. 43) *Combined Writing Scale*

¹⁴ See COPE scoring criteria for writing in <http://www.copetest.com/prepform.htm> – Retrieved April 6, 2009

common features of the writing such as content, organization, and language use. In terms of content, for example, Braine (2001) focused on relevancy of content with test takers' own ideas or positions and Cho (2003) focused on development of ideas and relevance of content with the given topic. Feak and Dobson (1996) examined the integration and synthesis of source materials. The criteria used to assess organization and language use are similar to other commercial language tests reviewed before.

Previous review of scoring criteria for commercial and in-house integrated writing tests found the major features assessed in these tests to be similar to those assessed in conventional writing-only tests: content, organization, and language use. One feature of writing assessed in addition to these three aspects is the use of source information. However, the essence of the source use measure in these scoring rubrics is varied from test to test. In some tests, the only criterion explicitly related to the use of source information is content. For example, in TOEFL iBT, the criterion for good use of sources relates to accurate and coherent presentation of source information in test takers' writing. In CAELA test, the criterion regarding the evaluation of source use is vaguely addressed in the rubrics: test takers are required to "effectively" use source information in their writing. The UAWA test described in Feak and Dobson's (1996) study required test takers to synthesize source materials and not copy the exact phrases from the sources. The scoring rubrics provided by the OTESL, COPE test and Braine's (2001) Exit Test do not specifically include a criterion for assessing the use of source materials. In contrast to the other tests, the EPT described in Cho's (2003) study uses an analytic scoring rubric to assess different writing features; source use is one of the features under assessment. The criterion evaluates to the accuracy of source information presented in test takers' writing.

The interdependent construct of these tasks complicates the development of criteria used to score integrated writing. Some scoring rubrics are kept the same as the

ones used to assess conventional writing-only types of tasks, and thus do not truthfully reflect the construct being measured. For those tests that incorporate source use as a criterion in their scoring standards, however, there is a discrepancy as to what constitutes good source use. The criterion used to measure source information in writing varies from the presentation of accurate source materials to full synthesis of source information, with and without mentioning issues of direct copying from source materials.

Scoring should reflect the given construct being measured, the task, the test purpose, and the test takers. Although most rubrics provide some evidence of test takers' academic writing ability, they do not satisfactorily tie to the nature of integrated writing test performance in which multiple language skills (e.g., reading, listening and writing) interact. This study represents an additional effort to elucidate the nature of reading-listening-writing test tasks.

Textual features and writing test scores

Some product studies have focused on comparisons of scores and textual features between two types of writing tasks: conventional writing-only and integrated writing tasks (Cumming et al., 2005; Lewkowicz, 1994). Lewkowicz (1994) compared holistic scores assigned to essays in response to a reading-to-write task and a writing-only task. No statistically significant differences were found in the scores. In terms of textual features, the length of reading-to-write essays was similar to the length of writing-only essays, while there was a significant difference in the number of ideas used in the essays written by these two groups. The study concluded that although reading material provided students with ideas, it did not help improve the quality of the written essays. Students without texts relied on the ideas from the reading and did not develop their main points fully. In contrast, students without the reading text were able to elaborate their ideas more extensively even though they did not provide as many points. These results

are limited for interpretation because the study lacks information about other textual aspects of the essays or details about the rating criteria used. The major features of essays with high or low ratings were not discussed.

In an analysis of TOEFL essay discourse features, Cumming et al. (2005) compared grammatical accuracy, lexical/syntactic complexity, rhetoric, and pragmatics between responses for independent (writing-only) and integrated tasks (reading-writing and listening-writing) across score levels 3, 4 and 5. The results indicate that the independent writing task response essays and the integrated writing task response essays are significantly different for subcategories such as length and quality of claims across proficiency groups. In terms of the overall textual features of reading-writing essays, effective examinees tended to a) summarize reading materials rather than paraphrase them, b) state the main position in the beginning and in the end, and c) avoid using verbatim strings from source readings. Similarly, effective examinees responding to listening-writing tasks summarized the lecture rather than using verbatim phrases or paraphrasing from the lecture. The least effective examinees, for both types of integrated tasks, used verbatim phrases and failed to cite the sources. They were also likely to write from personal experience without addressing the prompt question.

The textual features of examinees' essays varied extensively across score levels. The results indicate that examinees may approach these tasks differently. In addition, a single holistic score may assume examinees have equivalent ability across different aspects of writing, and may not reflect the examinees' ability on different aspects of the essay. To provide a fuller understanding of the nature of the tasks and examinees' responses toward the tasks, it is necessary to explore the underlying processes or strategies test takers use to carry out the tasks, and how these characteristics of strategies interact with different aspects of writing.

To analyze the relationship between textual features and resulting essay scores, Watanabe (2001) used one writing-only and two reading-to-write essays students completed for an English program placement exam. All reading-to-write essays were analyzed in terms of text length, the use of ideas from sources or experience or knowledge, the use of quotation marks and citation, and overall organization. He concluded that an essay from an integrated task receiving a high rating usually has a clear introduction, several body paragraphs discussing points in the two source passages and personal experience, a conclusion, and is of sufficient length.

According to the studies reviewed in this section, it appears that the textual features of an essay affect students' final scores. Effective examinees tend to incorporate source information to support their positions. They also avoid plagiarism by citing the sources in their essays. However, examining writing products alone may not elicit a full picture of the integrated writing construct. To interpret scores or refine the scoring scheme for integrated writing, Cumming et al. (2005) suggest future researchers examine the writing processes or strategies examinees use at different score levels. Similarly, Bachman (2004) supports process studies and sees them as means to evaluate the usefulness of the task: "There are two aspects of test performance that we need to investigate in our evaluation of test usefulness: the processes or strategies test takers use in responding to specific test tasks and the product of those processes or strategies—individuals' responses to the test tasks and the scores that they obtain" (p. 5). This review raises the following research questions:

1. What is the nature of strategy use in an integrated reading-listening-writing test?
2. What is the nature of test performance as measured by an integrated reading-listening-writing test?

3. What is the relationship between strategy use and test performance on an integrated reading-listening-writing test?

Chapter 3

RESEARCH METHODS

To investigate the research questions outlined in Chapter Two, the study employed a correlational research design, using a series of exploratory factor analyses (EFAs) and structural equation modeling (SEM) as the primary analytic approach as well as interviews, planning sheets, and an open-ended questionnaire as supplementary analytic techniques. EFA was first used to identify composite variables from the strategy inventory and then to generate baseline models for the SEM. Subsequently, SEM was used to assess the quality of these models. The qualitative data were used to cross-validate quantitative findings and provide further information about the strategy use of test takers in the high- and low-performance groups. This section presents the description of the participants, the measurement instruments and materials, the data collection, and analysis.

PARTICIPANTS

The sampling of participants was purposive based upon the given context—the admission test for North American academic institutions, the findings of previous research studies, and pilot testing. The goal of the study was to provide validity evidence for integrated reading-listening-writing tasks by collecting process (i.e. integrated writing strategy use) and products of writing (i.e. written texts). This type of task was introduced in the TOEFL iBT test in 2005. Therefore, participants who had or would take the TOEFL for university admissions were recruited. The raters were selected from a pool of experienced ESL teachers. The criteria and rationale for selecting test takers are discussed in detail below.

Recruitment of test takers

ESL students enrolled in the University of Texas at Austin during fall 2008 were chosen to represent a range of proficiency levels, fields of studies, and various language backgrounds. To provide reasonable estimates for relevant statistics, this study involved 161 participants.

Level of English proficiency of the participants is an important element of the sampling plan. Previous studies have noted that L2 proficiency appears to account for the difference in strategy use (de Larios, Manchon, Murphy, & Marin, 2008; Sasaki, 2000) and products of writing (Pennington & So, 1993; Raimes, 1985; Sasaki & Hirose, 1996). For this study, the sample population was identified and recruited from ESL students who enrolled in ESL programs, undergraduate programs, and graduate programs.

Cultural variation in writing practice such as the use of source materials is another issue commonly addressed in previous studies. Some researchers have reported ESL students with Asian backgrounds appear to use source material verbatim (Currie, 1998; Deckert, 1993; Moore, 1997), other researchers (Pennycook, 1994; Scollon, 1994) believed variations do exist across cultural and language backgrounds. Since the test takers of the TOEFL are likely to have different backgrounds, language, and culture, backgrounds of the participants were not controlled for the present study.

The academic status of participants is another issue of concern in this study. The focus of the TOEFL test is on writing in academic settings. Both ESL, undergraduate and graduate students are prospective test takers of the TOEFL. Thus, students of these statuses were recruited for this study.

The recruitment e-mail message was sent to potential participants through different student association listservs. I also visited university ESL academic writing classes and handed out a flyer (see Appendix A) to recruit students. To ensure a high

participation rate, participants were offered a gift certificate worth 10 dollars as a monetary incentive.

Demographic data

A purposive sample of 161 students at the University of Texas at Austin took part in this study. Participants were recruited based on their adult ESL learner status, including international and immigrant students from a wide variety of language backgrounds. Table 3.1 shows the participant frequencies sorted by native language.

In terms of the participants' gender, approximately 66% of the participants were female (N = 106) and 34% of the participants were male (N = 55). With regard to participants' educational backgrounds, about 54.7% of the participants were graduate students (N = 88), about 29.2% of the participants were undergraduate students (N = 47), and the remaining 16.1% of the participants (N = 26) were students enrolled in the ESL programs at UT.

Regarding the English proficiency of the participants, all graduate and undergraduate student participants enrolled in UT have a minimum TOEFL score of 550 (paper test), 213 (computer-based test), or 79 (Internet-based test). For students enrolled in the UT ESL program, five of them have not taken the TOEFL. The rest of these students have a wide range of TOEFL scores ranging from 123 to 237 (computer-based test)¹⁵. The average score was 181.27 with a standard deviation of 38.

¹⁵ The TOEFL scores of these participants were converted to computer-based test scores to allow comparison.

Table 3.1 Participants by native language

Native language	Number	Percent
Chinese (including Mandarin, Taiwanese and Cantonese)	80	49.7
Korean	45	28.0
Thai	7	4.3
Japanese	5	3.1
Arabic	4	2.5
Vietnamese	4	2.5
French	3	1.9
Spanish	2	1.2
Turkish	2	1.2
Portuguese	2	1.2
Uzbek	2	1.2
Bengali	1	.6
Urdu	1	.6
Swedish	1	.6
Farsi	1	.6
Hindi	1	.6
Total	161	100.0

Within a week after they completed the task, thirty-eight students volunteered to participate in a retrospective interview about their strategy use during the test and their language learning experiences. After the interview data were collected, they were classified into two groups based on their performance¹⁶ in the integrated writing test: high-performance group and low-performance group. For the purpose of this study, the interview data from the ten top scorers (13, 14 and 15 points) and from the ten bottom

¹⁶ Performance here refers to test takers' total scores on content, organization and language use features of the essay. Source use score were not used as a criterion for grouping so that the issues of inappropriate source use and general writing features can be examined separately.

scorers (4, 6, 7, and 9 points) were used. The data gathered from individuals who received a score ranging from 10 to 12 were not utilized. Interviewees' basic information is represented in Table 3.2 and Table 3.3. The interviewee data were used to triangulate with quantitative findings regarding the relationship between strategy use and test performance.

Table 3.2 Basic information on interviewees in high-performance group

Writing Scores	Verbatim Source use	Name ^a	Academic Status	Major	TOEFL	First language
13/15	10	Yawen	ESL	Chemical Engineering	230	Chinese
13/15	8	Haley	ESL	Finance	230**	Chinese
14/15	7	Feng	Under-graduate	Civil Engineering	270*	Chinese
15/15	12	Tonnie	Graduate	Foreign Language Education	260	Chinese
13/15	0	Dana	Graduate	Instructional Technology	none	Chinese
13/15	17	Richard	Graduate	Journalism	253*	Korean
15/15	3	Fulcher	Graduate	Foreign Language Education	277	Chinese
14/15	0	Lossie	Graduate	Educational Psychology	247*	Chinese
13/15	4	Liu	Graduate	Geophysics	233*	Chinese
14/15	9	Yvonne	Graduate	Foreign Language Education	250*	Chinese

^a All names are pseudonyms.

* These scores are converted from paper-based to computer-based TOEFL to allow comparison

** These scores are converted from Internet-based to computer-based TOEFL to allow comparison.

Table 3.3 Basic information on interviewees in low-performance group

Writing Scores	Verbatim Source use	Name ^a	Academic Status	Major	TOEFL	First language
9/15	0	Kim	Graduate	Physical Education	253	Korean
9/15	11	LJ	ESL	Speech-language pathology	193**	Chinese
7/15	38	Kate	ESL	Computer Science and Statistics	170**	Korean
7/15	20	Sara	Graduate	Special Education	260	Korean
6/15	16	Wokie	ESL	Business	207*	Korean
9/15	27	Jimmy	Under-graduate	Industrial Engineering	173*	Korean
4/15	15	Chih	ESL	Mechanical Engineering	none	Chinese
9/15	0	Elle	ESL	Education	none	Chinese
4/15	0	MD	ESL	Pharmacy	197**	Chinese
7/15	24	Tori	ESL	Japanese Literature	177*	Japanese

^a All names are pseudonyms.

* These scores are converted from paper-based to computer-based TOEFL to allow comparison

** These scores are converted from Internet-based to computer-based TOEFL to allow comparison.

MEASUREMENT INSTRUMENTS AND MATERIALS

Four types of instruments were used in this study: (a) the TOEFL integrated writing test task¹⁷, (b) the integrated writing scoring rubrics (see Appendix B), (c) the strategy inventory for integrated writing (see Appendix C) and (d) the background questionnaire (see Appendix D). The integrated writing scoring rubrics were developed for analytically scoring the data elicited from the reading-listening-writing test. The strategy inventory for integrated writing was designed to measure test takers' self-reports

¹⁷ The task used in the study was provided by ETS and is used in the TOEFL. It cannot be published due to security restrictions and thus is not included here.

of their strategy use during the test. The background questionnaire was designed to elicit participants' basic information. The details of these measures are described below.

The reading-listening-writing integrated writing task

The purpose of the study was to provide validity evidence for an integrated writing test, thus a large-scale current integrated writing task selected from the *TOEFL® iBT Data Set 3: Writing* (2008)¹⁸ was used as the test material in this study. Since the TOEFL iBT is intended to measure the language proficiency in academic settings, the writing section of this test was developed by ETS to simulate the types of writing in which participants engage in university-level academic settings.

To complete the reading-listening-writing task, all test takers were first asked to read a short reading passage approximately 255 words long for about 2 minutes and listen to part of a lecture related to the reading passage for about 2 minutes. Then, they were asked to write a response¹⁹ based on the relationship between the content of the reading passage and the lecture. The reading passage was available to the test takers while they were writing the response essays. They were also allowed to take notes during the lecture. In total, they had 20 minutes to plan and write their response. An expected length of the essay was about 150 to 225 words.

The integrated writing scoring rubrics

Analytic scoring has been considered useful in capturing strengths and weaknesses of L2 language test takers who are likely to have uneven developmental proficiencies across different aspects of writing (Hamp-Lyons, 1991; Raimes, 1990; Sasaki & Hirose, 1999). In order to examine how strategy use may account for different

¹⁸ The TOEFL iBT public use materials were developed to advance knowledge in the field of English language assessment by Educational Testing Service.

¹⁹ After they read the reading passage and listened to the lecture, they were asked to respond to the question presented right after the lecture: "Summarize the points made in the lecture, being sure to explain how they challenge the specific points made in the reading passage."

aspects of the essays, an analytic Integrated Writing Scoring Rubrics (Appendix B) was developed by adapting the TOEFL iBT integrated writing rubrics and the *LanguEdge Courseware* scoring rubric used by Cumming et al. (2005). These rubrics were intended for the evaluation of integrated writing. Although they were holistic scoring rubrics, several features of writing could be identified from the statements for each band. Based on these major features, these scoring rubrics were modified for analytic scoring particularly for this study to address four dimensions of integrated writing: content, organization, language use, and verbatim source use. There are 6 bands (0-5) under each category.

In this study, the band descriptors for *content* are accurate presentations of principal ideas from two sources and connections between two sources. According to the TOEFL task requirements, test takers are required to show their understanding of the major ideas and the relationships among ideas discussed in the reading passage and the lecture. Therefore, the TOEFL iBT integrated writing rubrics and *LanguEdge Courseware* scoring guidelines include these features in their holistic rubrics to determine effectiveness of task fulfillment. The band descriptors for *organization* are cohesion at the paragraph and essay levels. In terms of *language use*, grammar, mechanics and overall understanding for the written language are criteria for evaluation. While the TOEFL iBT integrated writing rubrics do not include *source use* in their statements for each band, the source use in essays should be evaluated because an appropriate use of source materials is expected in all academic writing contexts. Following the previous research on TOEFL integrated tasks (Cumming et al., 2005), the integrated writing scoring rubrics in this study evaluated the aspect of verbatim source use in the essays. Verbatim source use was defined as a string of three words or more directly drawn from the sources into student essays. The number of instances of a string of three words or

more copied verbatim from source texts was tallied. A written response that displays no verbatim source use receives a score of 5. A response that shows less than two strings of three words drawn from source texts receives a score of 4. A response that contains less than three strings of three words but more than two strings of three words copied verbatim from sources receives a score of 3. The rest may be deduced by analogy. Verbatim phrases in essays were checked to make sure they were not a collocation of words, such as technical terms or idiom phrases.

The strategy inventory for integrated writing

Language learning strategy use is typically assessed by self-reported questionnaires because strategies are often defined as techniques and behaviors learners consciously apply to efficiently carry out their learning tasks (Cohen & Upton, 2007; Ellis, 1994; Schmidt, 1994). For this study, the strategy inventory for integrated writing was developed to measure test takers' strategy use during the integrated writing task.

As noted by Flower et al. (1990) and Flower (1994), composing processes are context-sensitive and context-dependent. The construct of writing is situational and multidimensional. Learners' strategy use may be different in a non-testing writing condition and in a testing situation. For example, some test takers may employ test-wiseness tricks as short-cuts to achieve answers (e.g. not responding to the task as instructed but simply reciting from writers' previously-memorized writing templates). Such test-wiseness strategies have been reported by reading researchers (Cohen, 1998; Phakiti, 2003) while little has been studied in the context of writing tests. The strategy inventory for integrated writing was developed in an attempt to capture different types of strategies, both positive and negative, that test takers used to respond the reading-listening-writing test task.

The preliminary taxonomy of the inventory items was based on theoretical frameworks of the integrated operations (Spivey, 1997; Spivey & King, 1989), empirical studies on processes engaged in responding to integrated writing tasks (Asención, 2004; Plakans, 2008, 2009; Spivey, 1990, 1997; Spivey & King, 1989; Stein, 1990), and test-taking strategies (Cohen, 1998; Cohen & Upton, 2007). Inspired by these studies, organizing, selecting, connecting, planning, monitoring, evaluating, test management and test-wiseness were major types of strategies identified for this investigation. The categories were introduced for two purposes: (a) to ensure the inventory was comprehensive and accounted for as many strategies specific to integrated writing as possible and (b) to provide clearer descriptions and grouping of instances of strategy use. The development of the inventory also drew from the L1 and L2 literature in writing, integrated writing, learning, and test-taking strategies (Cohen, 1998; Esmaili, 2002; Flower & Hayes, 1981; Raimes, 1985; Sasaki, 2000; Torrance, Thomas, & Robinson, 1994; Trites, 2000; Watanabe, 2001).

Following Dörnyei (2003), items were listed in logical sequences (strategies used before, during, and after writing). Some items were interwoven with others because of the dynamic and interactive nature of writing processes. According to Kline (2005), items that have more than one major thought, called *double-barreled* items, should be avoided. With this goal in mind, each item in the strategy inventory for integrated writing is intended to describe a single thought as closely as possible. Kline (2005) also suggests that items should be brief and worded in simple language. All items in the strategy inventory for integrated writing, thus, were kept short and written in first-person narrative form. Furthermore, because one process may be closely related to the other, the categorization does not imply disassociation. A number of inventory items were adapted from Esmaili's (2002) *Checklist of Writing Strategies*. In developing the inventory,

negation and reversed items were not included because explicit negation demands additional cognitive loadings from respondents and thus increases the occurrence of misresponse (Canale & Swain, 1980) and decreases reliability (Herche & Engelland, 1996). To make the inventory meaningful and understandable for respondents, strategy items were expressed as specifically as possible and their level of specificity was refined to be consistent with other items. As suggested by Kline (2005), Cohen (1998) and Hsiao and Oxford (2002), items should be expressed as explicitly and context-specific as possible. In addition, the extent of specificity of one item should be closely aligned with another to ensure systematic and consistent responses from participants.

Because the inventory (see Appendix C) was given to students after they had completed the test, the statements that asserted the use of writing strategies were expressed in past tense. To indicate the degree to which students used a strategy during the test, a 5-point Likert scale was used: 1 (Very rarely), 2 (Rarely), 3 (Occasionally), 4 (Often) and 5 (Very often). The inventory was written in English, as this is the language commonly shared by all participants.

Based on the standard criteria for the development of valid and reliable inventory (Brown & Day, 1983; Dörnyei, 2003; Gilham, 2000), the preliminary inventory was piloted and analyzed for validity before actual use in the present study. Of various methods for establishing validity recommended by measurement researchers, three types of validity were addressed in this study: content, construct, and response validity. To establish content validity, the questionnaire draft was presented to four experts who are researchers and teachers of academic writing in English. Following Dörnyei (2003), redundancy, clarity, and readability of items were scrutinized on the basis of experts' perspective. To balance linguistic backgrounds, cultural backgrounds and gender on the panel, two male and three female native and nonnative English experts were selected to

examine the items. Definitions of measurement constructs were provided to experts as guidelines for examining the items. Experts' opinions were used to revise instructions, interpretations or wordings of the questionnaire. To obtain students' feedback, the inventory was piloted with seven potential participants. They were asked to fill out the strategy inventory for integrated writing after they completed the reading-listening-writing test. Emergent strategies for carrying out the task were added to the item pool. Adjustments for wording or conceptual problems were made based on students' feedback.

The construct validity of the inventory was established based on previous theoretical frameworks. As discussed in the earlier part of this section, the development of the inventory was informed by theories of L1 and L2 writing processes and language learning strategies. A number of composition researchers have called attention for investigators to be flexible in terms of applying models or constructs to their studies because there is still a lack of a single adequate model in explaining writing in a first or second language (Cumming, 1998; Flower, et al., 1990; Grabe, 2001). Current conceptualizations view writing as constructive and contextualized processes. They emphasize the impact of interactions among the contextual factors on the cognitive processes of writers (Flower et al., 1990; Grabe & Kaplan, 1996). Hence, based on the complexity of the present assessment context, various theories and models that focus on the interactions of cognitive processes, source information and written texts were used. Further validation data for the inventory are discussed in the data analysis and result section.

The background questionnaire

The background questionnaire (see Appendix D) given along with the strategy inventory for integrated writing was used to elicit information on test takers' background

information including their native language, academic status, writing courses and experiences, opinions about writing in English, and their English proficiency test reports (e.g., TOEFL scores or IELTS scores).

Though little writing research has been conducted to explore writing strategies used within and across cultural and linguistic groups, strategy research has indicated that L2 learners tend to use socially- or culturally-encouraged strategies as they learn a second language (Bedell & Oxford, 1996; Oxford, 1996). According to previous studies, writing expertise also proves to be an indicator of variation in students' writing behaviors such as process or strategy use (Sasaki, 2000; Yang & Shi, 2003). Students' academic status, graduate or undergraduate, has been found to influence their choice of source use in their written products (Watanabe, 2001). With respect to L2 proficiency, a number of writing studies (Connor, 1987; Connor & Kramer, 1995; de Larios, Murphy, & Manchon, 1999; Manchon & de Larios, 2007; Sasaki, 2000) found that L2 proficiency may explain the difference in strategy use, and the frequency and quality of strategy use lead to variance in written products. The background questionnaire was used to collect demographic information of test takers.

STUDY VARIABLES

This section describes the procedure for operationalizing the variables used in this study. This study contains four sets of variables: the rhetorical strategy use variables (measured by 12 items), the self-regulatory strategy use variables (measured by 14 items), the test-taking strategy use variables (measured by 8 items), and the integrated writing test performance variables (measured by 4 items). The first three types of variables represent the strategy use variables in this study. The fourth variable represents the various aspects of integrated writing ability.

Rhetorical strategy use

Based on the review of Spivey's work (1990, 1997) and several related studies (Asención, 2004; Esmaili, 2002; Plakans, 2009), three strategy-type variables were selected for this study to examine the mental operations used by the examinees during the integrated writing test: the organizing, selecting, and connecting strategy types. These strategies represent a series of interactions among writers, source texts and written texts. Depending on the nature of these strategies and writing context (i.e., an integrated writing task), this set of strategy types is operationalized as Rhetorical Strategy Use representing examinees' various approaches to recognizing the rhetorical context and using rhetorical resources to meet task goals. The rhetorical strategy use scale consists of 12 strategy items. Each strategy-type variable is represented by four inventory items. An averaged composite of all items was computed to generate different strategy types. For example, the organizing strategy type is an averaged composite of items 9, 10, 11, and 13. The strategy-type variables and corresponding items are displayed in Table 3.4. A description of each strategy-type variable in the study is as follows:

Organizing (ORG): This variable is designed to measure the extent to which the test takers seek to recognize organizational patterns and textual cues of source materials to frame their own texts based on their discourse goals and purposes.

Selecting (SEL): This variable is designed to measure the extent to which test takers seek to identify important information from sources based on task requirements such as textual relevance or other criteria. The selected information is deemed important for inclusion in test takers' own texts.

Connecting (CON): This variable is designed to measure the extent to which test takers seek to explore links among ideas from diverse sources and integrated information from these sources based on task goals and purposes.

Self-regulatory strategy use

Based on the review of literature (Asención, 2004; Esmaeili, 2002; Plakans, 2008, 2009; Stein, 1990) discussed in Chapter Two, three strategy-type variables were selected for this study to examine metacognitive operations of examinees during the integrated writing test: the planning, monitoring, and evaluating strategy type. These strategies represent a set of self-regulation techniques examinees used to orchestrate their cognitive operations and to help them complete their task at hand effectively. Based on the nature of these three strategy types and the task context, this set of strategy types is operationalized as Self-Regulatory Strategy Use representing examinees' various techniques to control over the process of thinking in response to the integrated writing test. The self-regulatory strategy use scale consists of 14 strategy items. Each strategy-type variable is represented by at least three inventory items. An averaged composite of all items was computed to generate different strategy types. The taxonomy of strategy-type variables and their corresponding items are displayed in Table 3.4. A description of each strategy-type variable in the study is as follows:

Planning (PLAN): This variable is designed to measure the extent to which test takers seek to generate plans to meet the external task demands such as time, physical resources, and internal task demands.

Monitoring (MON): This variable is designed to measure the extent to which test takers seek to check the progress and the effectiveness of their own performance when completing the task. Monitoring may also involve identifying potential problems of a task at hand.

Evaluating (EVAL): This variable is designed to measure the extent to which test takers seek to reexamine task effectiveness and fulfillment by evaluating task

requirements, purposes for the writing, planned thoughts, written texts, and changes made to the text.

Test-taking strategy use

Based on the review of literature (Allan, 1992; Cohen & Upton, 2007) discussed in Chapter Two, two strategy-type variables, the test management and test-wiseness strategy type, were selected for this study to examine test-specific techniques examinees used to approach the integrated writing test. This set of strategy types is operationalized as Test-Taking Strategy Use that represents examinees' various techniques to handle an integrated reading-listening-writing test task. The test-taking strategy use scale consists of eight strategy items. Each strategy-type variable is represented by at least three inventory items. An averaged composite of all items was computed to generate different strategy types. Table 3.4 shows the taxonomy of strategy-type variables and their corresponding items. A description of each strategy-type variable in the study is as follows:

Test-Management (TM): This variable is designed to measure the extent to which test takers seek to understand task goals by using clues available from the task instructions.

Test-Wiseness (TW): This variable is designed to measure the extent to which test takers utilize the formats or characteristics of a test or a test condition to compensate the ability being measured.

Table 3.4 Taxonomy of the strategy inventory for integrated writing (34 items)

Variables	No.	Item used
1. Rhetorical Strategy Use		
• Selecting (SEL)	4 items	3, 6, 12, 15
• Organizing (ORG)	4 items	9, 10, 11, 13
• Connecting (CON)	4 items	4, 5, 8, 17
Subtotal	12 items	
2. Self-regulatory Strategy Use		
• Planning (PLAN)	3 items	2, 7, 27
• Monitoring (MON)	6 items	21, 24, 25, 26, 30, 32
• Evaluating (EVAL)	5 items	23, 29, 31, 33, 34
Subtotal	14 items	
3. Test-Taking Strategy Use		
• Test-Management (TM)	3 items	1, 16, 19
• Test-Wisness (TW)	5 items	14, 18, 20, 22, 28
Subtotal	8 items	
Total	34 items	

Integrated writing test performance

The review of scoring criteria in Chapter Two indicates that content, organization, and language use are the most common measures of integrated writing. A measure of verbatim source use was added to examine test takers' retention of verbatim phrases from sources in their own writing. Integrated Writing Test Performance constitutes four observed variables of the writing test, each representing one feature of writing in the scoring rubrics used to score essays from the integrated writing task.

Content: This variable is designed to measure the test taker's ability to select principal information from the lecture to support ideas and connect key points in the lecture and the reading passage. Considering the nature of the reading-listening-writing task and several scoring criteria used by commercial and in-house tests discussed previously, appropriate content in response to the task needs to be relevant to ideas in sources, accurate in incorporating source information, and well-developed based on main

points or arguments drawn from sources. These criteria are essential in assessing test takers' ability to integrate source information fully into their essays.

Organization: This variable is designed to measure test takers' ability to write an organized, logical, and coherent response at the paragraph and essay level. Effective organizational framework in response to the task needs to be well-developed based on a clear introduction, several supporting body paragraphs discussing ideas presented in two sources, a concluding remark, and a sufficient length.

Language Use: This variable is designed to measure test takers' ability to use accurate and appropriate sentence formation and word forms. An effective writing shows a range of sentences types and vocabulary with occasional minor grammatical or word choice errors that do not obscure meaning.

Verbatim Source use: This variable is designed to measure the extent to which test taker use verbatim phrases from source texts expressed as the number of verbatim strings of three words or more from the source materials. These verbatim strings of words are checked to make sure they were not a collocation of words.

DATA COLLECTION

The purpose of this study was to investigate the relationship between writing and test-taking strategies test takers used and their test performance. Various types of data were collected in the current study: test takers' essays, planning sheets, the strategy inventory for integrated writing, the background questionnaire, and interview protocols.

Procedures

Participants came to the computer lab in several small groups. As with the TOEFL test, students first read the instructions for the writing task, and then they wrote on the computer. Students were encouraged to ask questions regarding the instructions or vocabulary in the instructions that they did not understand. Explanations or translations

were provided to answer the questions. Then, students were asked to complete the integrated writing test, the strategy inventory, and the background questionnaire.

To complete the task, each student was directed to a website designed for this study. This website integrated all elements—the reading passage, the lecture, the strategy inventory, and the background questionnaire—necessary for this study. Test takers were required to read a passage on an academic topic and listen to part of a lecture on the related topic. Then they were asked to write a response to a question that asked them to summarize ideas from the lecture and explain the relationship between ideas from the lecture and the reading passage. They could take notes as they read or listened. They were allowed two minutes to read the passage then listen to the lecture. They had twenty minutes to plan and write their responses. Immediately after the writing test, all students were asked to complete the strategy inventory for integrated writing and the background questionnaire. All participants had up to one hour to complete both the test and the strategy inventory. Test takers were recruited to participate in follow-up retrospective interviews. After the interview data were collected, these data were then classified into high- and low-performance groups with ten test takers in each identified for further analysis.

Table 3.5 Procedures for data collection

Phase	Procedures
Phase 1	Integrated reading-listening-writing task
Phase 2	Strategy inventory for integrated writing
Phase 3	Background questionnaire
Phase 4 (optional)	Follow-up interviews

Scoring

Prior to the scoring sessions, raters, two experienced ESL instructors and one EFL instructor, were sent a booklet to guide them in scoring the essays. The booklet contained

the test directions and example responses that had been scored analytically based on 6-point (0-5) scoring rubrics. Raters were instructed to become acquainted with the rubrics so they were familiar with the scoring criteria before the actual training and scoring.

In the training and practice scoring session, a two-stage process was employed in the study. First, two raters read the scoring rubrics together and discussed their ideas about each category and each band. Then they independently read a test taker's essay and assigned a single score on categories of content, organization, language use, and verbatim source use of the essays. If an agreement in ratings between two raters was achieved, two scores were averaged to form the final judgment of the performance. In the cases when discrepancies in ratings occurred—scores were considered discrepant when they were at least three points apart (Cooper et al., 1984)—a third rater conducted a blind review of the essay, and the score assigned by the third rater was summed and averaged with the closest original rating (Myford & Wolfe, 2002; Weigle, 1999).

DATA PREPARATION

Data preparation involves entering data into the computer, checking the data for accuracy, checking for missing data, and transforming the data.

Two different sources of data, strategy inventory data and test performance data, were input in separate Excel files on the computer. The accuracy of data entry was checked. Based on the results of data screening, five test takers' data were dropped because their essays showed that they had not made a reasonable attempt to respond to the task. The rest of the data were examined for missing values. No missing values were detected in these data sets. As a result, five cases were dropped from the study, leaving 161 cases to be used in the current study.

Once the data sets for strategy inventory and test performance were formatted, they were transformed for further analyses. To form strategy use composite variables, test

takers' item scores of the same scale were averaged, producing a single score ranging from 1 to 5 for each strategy-type variable. For example, scores from each strategy item measuring Organizing were averaged to form an Organizing composite variable.

In the case of the integrated reading-listening-writing task, the test takers' raw scores from two raters were averaged to form three test performance variables, producing a single score ranging from 0 to 5 for each writing test performance variable. For instance, scores assigned by two raters measuring the content feature of the essays were averaged to form a composite variable, Content.

DATA ANALYSIS

Primary ex post-facto correlational techniques used in this study were exploratory factor analysis (EFA), confirmatory factor analysis (CFA), and structural equation modeling (SEM). Previous L2 writing and assessment research provides heuristic inquiry into the nature of the integrated writing tasks and strategies students use to carry out the tasks. In this study, EFA was used to uncover the underlying factor structure of the measures, and CFA was used to generate baseline models for the subsequent SEM, whose analyses served as invaluable tools to examine the quality of the overall model. To probe more deeply into participants' strategy use in response to an integrated reading-listening-writing test, follow-up interviews were transcribed and analyzed.

Previous research investigated the relationship of strategy use and L2 performance by means of frequency counts, correlational analyses, analyses of variance, or t-tests. However, these standard statistical methods such as correlational and regression analyses are limited in that they do not directly account for measurement errors, and thus affect the generalization of the research. Bachman (1991; 2002) proposed that SEM is an abundant approach to the construct validation process. The SEM approach also facilitates the examination of the relationship between L2 test performance and test

takers' strategy use (Bachman, 1991; Cohen, 2006). To analyze the qualitative data, this study followed inductive approaches described by Charmaz (2004, 2006). The data were examined for patterns, and these patterns were evolved into codes of the test takers' strategy use. Both emerging codes and theories on integrated writings (Asención Delaney, 2008; Plakans, 2008; Spivey & King, 1989; Watanabe, 2001) were considered for analysis. These data were used to provide further information about test takers' strategy use and to cross-validate the strategy inventory items. Following the previous recommendations, three statistical analyses—EFA, CFA and SEM—were used to address research questions presented in Chapter Two.

1. What is the nature of test takers' use of integrated writing strategies in an integrated reading-listening-writing test?

This question concerns the trait structure of rhetorical, self-regulatory, and test-taking strategies²⁰ used by test takers in response to an integrated writing test. To assess item grouping and examine how these items are related to the hypothesized clustering, a series of item-level EFAs were carried out on the strategy use inventory data. Once the clusterings of the observed variables were determined, a series of CFAs were performed to examine the nature of strategy use. The quantitative data were cross-validated with the retrospective interview data.

2. What is the nature of the integrated writing test performance as measured by an integrated reading-listening-writing test?

This question addresses the trait structure of an integrated reading-listening-writing test. A series of item-level EFAs were first conducted to determine the grouping of observed variables represented in an integrated reading-listening-writing test. This step

²⁰ Based on EFA results, the test-management strategy type was dropped. Thus, the test-taking strategy category was also dropped, leaving the test-wiseness strategy type for further SEM analyses.

was followed by a series of CFAs in order to determine the underlying construct of the writing test.

3. What is the relationship between the test takers' use of integrated writing strategies and their performance on an integrated reading-listening-writing test?

Based on substantial literature on L2 writing and language assessment, I proposed and tested a number of competing models concerning the relationships among self-regulatory strategy use, rhetorical strategy use, test-wiseness strategy use and integrated writing test performance. A series of separate CFAs were conducted to examine the links between the measured and latent variables in each measurement model. On the basis of each measurement model, structural models were constructed to explore the relationships among the underlying constructs by testing them for best statistical fit of the data using SEM.

Computer equipment and software

Microsoft Excel 2003 was used to input the scores and subscores of the integrated writing test and to format the background information and strategy inventory data. These data then were exported to another statistical program, the SPSS 15.0 for PC. The program was used to compute descriptive statistics, reliability estimates, and to perform exploratory factor analyses on the strategy inventory and the integrated writing test. The AMOS 16.0 was used to perform confirmatory factor analyses and structural equation modeling.

Overview of statistical procedures

Several statistical procedures were utilized in this study. The procedures were described as follows. A flow chart of these procedures is presented in Figure 3.1.

Descriptive statistics

Descriptive statistics were computed first. To check the normality assumptions of the variables, the means, standard deviations, skewness and kurtosis of each variable were examined. The univariate normality of the scores was inspected because the dominant method—maximum likelihood estimation (ML)—in SEM for estimating path coefficients requires normally distributed endogenous variables. The violation of normality assumption affects the results of the chi-square test, resulting in inflated or deflated chi-square values.

Reliability analyses

Internal consistency estimates (i.e. Cronbach's Alpha coefficient) were computed for strategy variables and for the four averaged scores of the integrated writing test. The reliability of a measure was judged by estimating how well different items for the same construct within each measure yielded consistent results. If respondents answer two items designed to measure the same thing in a similar fashion, the measure is said to have internal consistency. A widely-accepted social science cut-off alpha value .70 was selected for the reliability analysis.

Exploratory factor analyses

Exploratory factor analysis (EFA) is a set of statistical procedures used as a means of identifying underlying clusters of a set of variables based on common hypothetical factors or variables. A number of studies aimed at investigating learner strategies have utilized EFA for reducing variables and defining the dimension of the underlying constructs (Phakiti, 2003; Purpura, 1997).

In this study, a set of EFAs was performed to investigate the factor structures of inventory items and test items so that the items best measuring the same and separate underlying constructs could be identified. According to Kim and Mueller (1978), EFA is

performed in three steps: (a) preparation of Pearson product-moment correlation matrix, (b) factor extraction and (c) rotation and interpretation. Correlations among various strategy use items were analyzed and the matrix of product-moment correlations was generated for factor analysis. Bartlett's test of sphericity, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy, and determinant of correlation matrices were performed to examine the factor structure of both variables in the strategy use data and scores on the writing test.

Once the number of factors were identified and retained, another factor analysis was run to obtain the loadings for each of the factors. The principal axis factoring (PAF), also known as common factor analysis, was used to extract the initial factors. Common factor analysis, instead of principal component analysis, was selected for the analyses of this study because common factor analysis is a better technique than principal component analysis in that it takes errors into consideration. The extraction procedure produced a factor loading for every item on each extracted factor.

After determining the number of underlying factors to be extracted, the loadings were rotated. Rotation maximized high loadings and minimized low loadings in order to generate simple structures, with most inventory items that had larger loading on one underlying factor but smaller loadings on the other factors. The results of these EFAs served as the baseline models for the subsequent SEM analyses.

Structural equation modeling

Structural equation modeling is a sophisticated statistical procedure used to determine if a model is valid for the data given hypotheses derived from prior research. In this study, SEM was utilized as the primary procedure for statistical analysis because it provides a number of advantages over other standard multivariate procedures (e.g., ANOVA, canonical correlational analysis, regression analysis, path analysis or EFA).

First, SEM takes a more confirmatory approach, which facilitates hypothesis testing. Also, unlike classical statistical procedures, SEM evaluates and corrects measurement errors, thereby providing explicit estimates of parameters (Jöreskog & Sörbom, 1989; Stevens, 1992). Finally, SEM allows observed and latent variables to be tested simultaneously (Byrne, 1994).

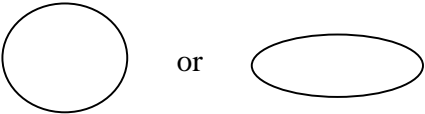
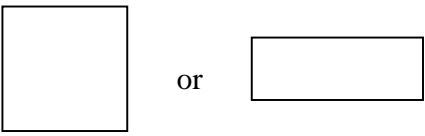
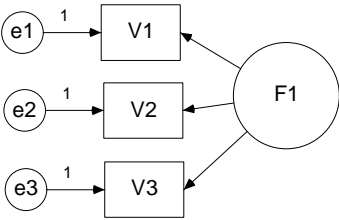
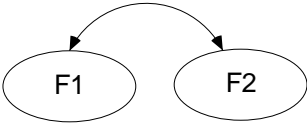

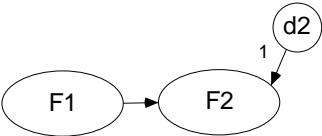

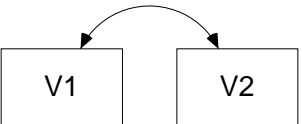

SEM is an umbrella term of three processes (Keith, 2006; Schumacker & Lomax, 1996): (a) path analysis, (b) confirmatory factor analysis, and (c) structural models. Path analysis refers to analysis of structural models of observed variables. Confirmatory factor analyses (CFA) are analyses of *priori* measurement models where the number of latent factors and their link to the indicator variables are specified. Once the measurement models are specified, structural models are used to determine the relationship between latent variables. The analysis of both measurement and structural models refers to the full latent model. Measurement models, structural models and the full latent variable models were utilized in this study.

SEM involves measurement models and structural models. Measurement models describe how latent variables are measured by observed variables. For example, in this study, integrated writing test performance is measured in terms the four features of the integrated writing test, whereas rhetorical strategy use is measured in terms of three rhetorical strategy types. Subsequently, structural models describe how one hypothesized latent variable relate to another. In the present study, this procedure addresses the relationship between integrated writing strategy use and integrated writing test performance.

Table 3.6 summarizes the symbols used in SEM. Latent variables or unobserved variables are depicted using circles or ovals. Observed variables or indicator variables are depicted using squares or rectangles. Single-headed arrows represent a presumed causal

relationship between two observed variables or unobserved variables. Double-headed arrows represent a noncausal relationship between two observed variables or unobserved variables.

Table 3.6 Symbols used in SEM

Symbols	Interpretation
	Unobserved or latent variable (circles or ellipses) that can be inferred from measured variables (e.g., construct of a measure)
	Observed or indicator variable (rectangles or squares) that are measured through tests, observations, questionnaires or other methods
	<p>→: Presumed causal relationship</p> <p>e1, e2, e3: Residual or measurement errors associated with the observed variables</p>
	 : Presumed non-causal relationship between two latent variables
	 : Disturbance in prediction of a factor
	 : Presumed non-causal relationship between two observed variables

Suggested by SEM experts, there are six steps necessary in model testing in addition to data collection: model specification, identification, estimation, evaluation, and

modification (Keith, 2006; Kline, 2005; Weston & Gore, 2006). The statistical modeling in this study was conducted as follows based on those steps.

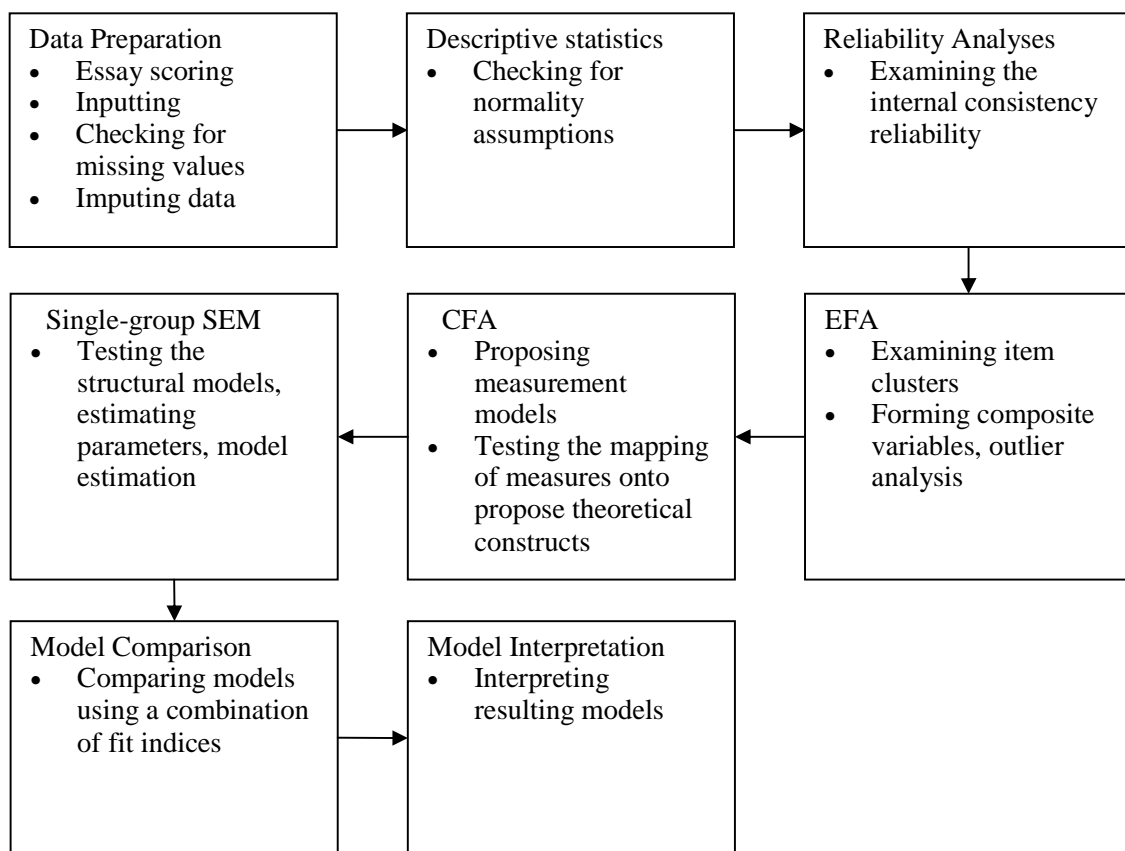
First, a tentative model was drawn based on the problem of interest: how might these hypothesized variables be related? The procedure specifies the existence or types of hypothesized relationships among latent and measured variables. Based on relevant theories and research, a number of variables were selected and the directions of their causal relationships were determined. Given the information, the tentative model was revised and specified. Then, data were collected. Prior to analyzing the data, the identification status (over-, under-, or just-identified) of the model was checked using the formula $[p*(p+1)]/2$ where p is equal to the number of observed variables in the model. At the same time, hypotheses were tested to determine which relationships were significantly different from zero and which were not.

Then, free parameters were estimated using SEM programs. Analysis of Moment Structures (AMOS) was used as a major SEM program for statistical modeling. The step of evaluation aimed to determine whether the estimated relationships among measured and latent variables in the implied model adequately reflected the observed association in the observed data. A multitude of fit statistics or fit indices were used to test whether the model was a good explanation of the data. For the present study, the root mean square error of approximation (RMSEA) was used as the primary measure of fit, supplemented by the standardized root mean square residual (SRMR), comparative fit index (CFI), Akaike Information Criterion (AIC), Tucker-Lewis index (TLI), or some other indices if necessary. AIC is an invaluable tool when comparing models that are not hierarchically related. Smaller AIC values are indicative of a better fitting model.

Using the modification indices, the final goal was to arrive at a best-fitting model that represented the constructs being measured. According to Hu and Bentler

(1999), an acceptable model fit was defined by the following criteria: RMSEA ($\leq .06$), SRMR ($\leq .08$), CFI ($\geq .95$), and TLI ($\geq .95$). Figure 3.1 shows a flow chart of statistical procedures used in this study. A variety of goodness-of-fit indices may be used but in this particular study, Hu and Bentler's (1999) rules of thumb commonly employed in fields of educational measurement were followed.

Figure 3.1 A flow chart of statistical procedures used in this study



Retrospective interviews

Triangulation in data collection techniques and data sources allow researchers to provide an alternate evidence for verifying or validating the information obtained by other methods. Triangulation can be accomplished by integrating quantitative and

qualitative methods, resulting in better interpretation of the issue of interest (Steckler, McLeroy, Goodman, Bird, & McCormick, 1992). The present study was predominantly quantitative. Qualitative results were used to clarify various aspects of quantitative findings.

To probe more deeply into the study participants' strategy use, retrospective interviews were employed in the study to generate qualitative data. This method allowed participants to verbalize their thoughts after they performed the task. Although the data were collected from only a small range participants (only ten high performing students and ten low performing students) as measured by an integrated reading-listening-writing test, the qualitative findings were useful for a deeper understanding of the nature of integrated writing test performance and integrated writing strategy use.

The interviews were conducted in English and lasted about 20 minutes. The interview questions (Appendix E) were open-ended questions designed to triangulate with items on the strategy inventory and explore other strategy use issues in greater depth. The combination and integration of data sources including the strategy inventory, the background questionnaire, test scores, planning sheets, and interviews protocols could enhance the validity of a study.

Participants for retrospective interviews were students in the high- and low-performance groups determined by their averaged writing scores. They were first asked to talk about their strategy use as presented in the questionnaire and additional strategy use during the integrated writing test. This part took about 15 minutes. Following the first part, participants were asked to talk about their general language learning experiences, attitudes, language learning strategies, and general writing strategies. This part took about five minutes. Interviews were held within a week after their test either in a UT computer

lab or in school library. They were audio-recorded digitally and transcribed for further analyses.

SUMMARY

This chapter described the participants in the study, the measurement instruments, the methods used to collect and score data, and statistical procedures and qualitative approach used to analyze the data. The quantitative analyses address all research questions regarding the nature of the strategy inventory as well as the relationship between strategy use and writing test performance. The qualitative data were used to triangulate with the questionnaire items and provide more information about test takers strategy use and experiences during the test.

Chapter 4

RESULTS

This chapter presents the results of quantitative and qualitative analyses in three sections. The first section discusses results of item-level exploratory factor analyses (EFAs) performed on the strategy inventory for integrated writing and the reading-listening-writing test performance. The second section addresses the first and second research questions regarding underlying constructs of integrated writing strategy use and integrated writing test performance. Based on the composite variables generated in EFAs, two sets of measurement models, integrated writing strategy use models and test performance models, were proposed using a series of confirmatory factor analyses (CFAs) to affirm the mapping of measures onto proposed theoretical constructs. The third section attempts to answer the third research question about the relationship between strategy use and test performance. Finally, a detailed discussion about the full latent model will be provided. In addition to quantitative results, each section contains relevant qualitative information to triangulate and properly interpret quantitative results. It also provides insight into patterns of strategy use among test takers in high- and low-performance groups.

ITEM-LEVEL EXPLORATORY FACTOR ANALYSES

A series of item-level analyses were conducted prior to modeling the relationships among the items in the strategy inventory for integrated writing and the integrated writing test performance as measured by an integrated reading-listening-writing test. The purpose of the analyses was to establish the convergent and divergent validity of each instrument. More specifically, the study examined whether items were measuring the same construct

form a factor (convergent) or items were measuring a different construct form a separate factor (divergent). This procedure allowed the validation of the taxonomies proposed in Chapter Three.

First of all, how items in each instrument related to their proposed taxonomies was examined. For example, although items 5, 4, 8, and 17 were developed to measure Connecting, there was no empirical evidence to claim that these items corresponded to test takers' observed behavior patterns. These analyses also provided insight into how to formulate composite variables from individual items. In other words, does the Connecting scale consist of items 5, 4, 8, and 17, or a different set of items? First, a series of preliminary item-level analyses were performed to inspect the internal consistency reliability. More specifically, the reliability of the Connecting scale, including items 5, 4, 8, and 17, was examined. Second, a number of item-level factor analyses were conducted to assess the factorial structure of the strategy inventory and the writing test. Put differently, these analyses tested the clustering of strategy items and how they related to the proposed strategy taxonomies in Chapter Three. For example, do items 5, 4, 8, and 17 cluster to form the Connecting scale or a different strategy scale? These findings provided substantive information regarding the convergent and divergent validity of each measure in the study. They also served as a foundation for the modeling of the factor structures of integrated writing strategy use and integrated writing test performance. The detailed procedures were presented in Chapter Three.

The strategy inventory for integrated writing

The preliminary analysis first examined the normality of the strategy inventory data. A combination of skewness and kurtosis coefficients was used because it allowed for the testing of two types of normality violations separately (Bollen, 1989). In addition, descriptive analyses including mean, standard deviation, and Cronbach's coefficient alpha

for inventory reliability were reported. Subsequently, a series of EFAs were performed to determine whether each strategy item corresponded with its respective scale.

Distributions and reliability

This section presents the item-level data from the strategy inventory for integrated writing. The descriptive statistics for the 34 items are presented in Table 4.1. The table presents distributions for integrated writing strategy use variables. In this table, all strategy-type variables such as CON1 (i.e., the first inventory item categorized as the Connecting strategy) were listed alphabetically and were presented with their corresponding strategy inventory items such as Item 5. All of the strategy scales are positive strategies except for the Test-Wisness Strategy Use scale. The means ranged from 2.75 to 4.45 and the standard deviations from .80 to 1.41. A certain degree of deviation from normality was revealed, as five variables had absolute skewness values exceeding 1.0 and positive kurtosis values. The distributions were slightly negatively skewed. In general, the skewness and kurtosis values were within the acceptable range, values no greater than the absolute value of 3, indicating that the distributions of the variables were not wildly non-normal.

Table 4.1 Distributions for the integrated writing strategy use variables (N = 161)

Variable	Strategy Items	Mean	SD	Skewness	Kurtosis
CON1	5	4.45	.80	-1.68	2.96
CON2	4	4.31	1.03	-1.58	1.95
CON3	8	3.97	1.24	-1.13	.36
CON4	17	3.64	1.24	-.50	-.82
EVAL1	29	2.96	1.21	-.08	-.93
EVAL2	33	3.71	1.10	-.63	-.32
EVAL3	34	3.00	1.22	-.01	-.90
EVAL4	23	3.59	.96	-.46	-.03
EVAL5	31	2.75	1.35	.14	-1.27
MON1	21	3.65	1.12	-.66	-.26
MON2	32	3.73	1.16	-.64	-.42
MON3	24	3.98	1.12	-.84	-.29
MON4	25	3.03	1.16	-.16	-.96
MON5	30	3.31	1.24	-.25	-1.00
MON6	26	3.72	1.09	-.52	-.55
ORG1	11	4.10	.96	-1.01	.58
ORG2	10	3.89	1.00	-.74	.20
ORG3	9	3.72	1.07	-.71	-.09
ORG4	13	3.76	1.07	-.63	-.08
PLAN1	7	3.36	1.16	-.34	-.68
PLAN2	27	3.59	1.11	-.46	-.58
PLAN3	2	3.54	1.14	-.38	-.77
SEL1	12	3.85	.99	-.50	-.61
SEL2	3	4.19	.83	-.70	-.31
SEL3	6	4.03	1.03	-1.10	.86
SEL4	15	3.08	1.29	-.10	-1.08
TM1	1	3.99	1.01	-.99	.56
TM2	19	3.03	1.41	-.10	-1.30
TM3	16	3.49	1.20	-.42	-.79
TW1	14	2.96	1.30	.14	-1.09
TW2	18	2.82	1.26	.34	-.94
TW3	20	3.41	1.24	-.32	-.91
TW4	22	2.75	1.32	.29	-1.02
TW5	28	3.85	1.30	-.83	-.55

Note. CON = Connecting, EVAL = Evaluating, MON = Monitoring, ORG = Organizing, PLAN = Planning, SEL = Selecting, TM = Test-Management, TW = Test-Wisness

Table 4.2 shows the reliability estimates for the eight strategy-type variables, ranging from a low .30 for the Connecting scale to a relatively high .86 for the Organizing scale. In general, the reliability estimates show mixed results: most strategy scales were relatively high (ORG, MON, EVAL, and TW), but two were relatively low (CON and TM), and the rest were moderate (SEL and PLAN).

Table 4.2 Reliability estimates for the integrated writing strategy use variables (34 items)

Variables	No.	Item used	Reliability estimates (α)
1. Rhetorical Strategy Use			
• Organizing (ORG)	4 items	9, 10, 11, 13	.86
• Selecting (SEL)	4 items	3, 6, 12, 15	.64
• Connecting (CON)	4 items	4, 5, 8, 17	.30
Subtotal	12 items		.78
2. Self-regulatory Strategy Use			
• Planning (PLAN)	3 items	2, 7, 27	.54
• Monitoring (MON)	6 items	21, 24, 25, 26, 30, 32	.84
• Evaluating (EVAL)	5 items	23, 29, 31, 33, 34	.73
Subtotal	14 items		.87
3. Test-Taking Strategy Use			
• Test-Management (TM)	3 items	1, 16, 19	.34
• Test-Wiseness (TW)	5 items	14, 18, 20, 22, 28	.74
Subtotal	8 items		.74
Total	34 items		.89

Item-level EFAs for the strategy inventory

In order to investigate how integrated writing strategy items clustered with proposed strategy-type variables, a Pearson correlation coefficient matrix was produced based on all 34 items. Following the initial data screening described in the previous section, a factor analysis (i.e., principal axis factoring) was used to extract factors according to the procedures described in Chapter Three.

To examine whether or not the variables were clustered into a smaller set of underlying factors, the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy was

used. The KMO statistic revealed a high value of 0.81 (close to 1.0), indicating that factor analysis was appropriate for these data. The significant result ($p < .001$) of Bartlett's measure of sphericity indicated that the original correlation matrix was not an identity matrix²¹; therefore, some relationship may be identified between variables, and factor analysis is an appropriate analysis.

Principal axis factoring (PAF) was used to analyze common factor variability, removing the uniqueness or unexplained variability from the model. Loadings with an absolute value less than .4 were suppressed. The results showed that a six-factor oblimin solution was most appropriate in terms of optimizing parsimony and interpretability. Most of the strategy-item variables loaded with their respective strategy scales (e.g., Items MON1, MON2, MON3 loaded with the Monitoring scale). The Planning scale was not included in the clustering as the Planning scale produced two items (PLAN2 and PLAN3) that did not have sufficient loadings on any factor, and one (PLAN1) loaded highly with the Connecting scale. Thus, the Planning scale was dropped, and PLAN2 and PLAN3 were deleted because of low loadings. Item PLAN1 was recoded as a Connecting item. Another scale not accounted for in the clustering was the Test-Management scale, as this scale produced one item (TM1) with low loadings on all factors, one that loaded with the Test-Wiseness scale (TM2), and one that loaded with the Monitoring scale (TM3). It was not surprising that TM2 double-loaded with two factors given the low alpha for the Test-Management scale in the reliability analysis results. Due to the uninterpretable loadings of TM2 on the Test-Wiseness scale, the item was dropped. Item TM3 was recoded as a Monitoring item (MON7). Also, Item TW5 loaded highly on the Evaluating scale. It was finally deleted because its underlying construct was not congruent with the rest of the Evaluating items. Five other items CON4, EVAL4, EVAL5

²¹ The identity matrix is defined as any square matrix whose elements are all zero and the entries along the main diagonal are the value 1.

SEL2, and TW4 were dropped due to their low (i.e., less than .40) and uninterpretable loadings. Table 4.3 shows the strategy items/variables that were dropped and recoded based on the EFA results.

Table 4.3 Variables dropped and recoded from the original strategy inventory

Variable	Item	Action
CON4	I reread the reading passage again to look for main ideas.	Dropped
EVAL4	I used different words or phrases to describe ideas from the reading passage or lecture.	Dropped
EVAL5	I added new points based on the reading or the lecture.	Dropped
PLAN1	I made a writing plan (e.g., outlines, notes, keywords, etc.)	Coded as CON3
PLAN2	I thought about a word, phrase, or sentence before I wrote it down.	Dropped
PLAN3	I thought about the type of essay I wanted to write.	Dropped
SEL2	I tried to summarize overall ideas from the whole reading passage in my mind.	Dropped
TM1	I reread the task requirements carefully.	Dropped
TM2	I thought about mentioning the authors in my essay.	Dropped
TM3	I double checked to see if my writing met the task requirements.	Coded as MON7
TW4	I tried to write about my knowledge or my own experiences in the essay.	Dropped
TW5	I first wrote out a writing template I had memorized before and filled in some ideas from the lecture or the reading passage.	Dropped

Note. CON = Connecting, EVAL = Evaluating, PLAN = Planning, SEL = Selecting, TM = Test-Management, TW = Test-Wiseness

The qualitative data including students' planning sheets, interviews, and responses to an open-ended questionnaire helped explain the problems with some of these items. With respect to Item PLAN1, it was found that many students created a writing plan to connect ideas from the reading passage and the lecture. Many students' planning sheets showed that major ideas from the reading passage were organized chronologically based on the order of their presence in the passage. Similarly, the corresponding ideas in the lecture were listed side by side with the idea notes taken from the reading passage. The interview data also revealed the same pattern. One student, Richard, in the high-performance group described his strategy in the following way, "I noted down the

keywords from the reading first. Then I added some keywords from the lecture to my notes. It helped me compare ideas from the reading three points from the reading and the lecture.” This might explain why Item PLAN1 loaded highly on the Connecting scale rather than the Planning scale.

In terms of Item TM3, the interview data showed that this strategy was mostly used by students in the high-performance group. These students tended to read the task requirements several times or write down the requirements on the planning sheet provided to them as a reminder to themselves during writing. This item was viewed by students as a more general monitoring strategy rather than a specific strategy used to respond to the task at hand. Feng mentioned, “I practiced this type of writing before...I think it’s important to know exactly what they want you to do. I often write down the essay question so that I could check whether my writing meet the requirements.”

The item seemed to represent a component of an overarching metacognitive control over the task management, arrangement and implementation. The test takers’ perspective on how they perceived the essay question provided explanation regarding the high loading of Item TM3 on the Monitoring strategy-type variable.

These changes left 24 items in the strategy inventory for integrated writing. The initial factor extraction yielded six eigenvalues greater than 1.0, accounting for 44.87 % of the variance (Table 4.4). Factor 1 represents the Monitoring scale; factor 2 represents the Test-Wiseness scale; factor 3 represents the Organizing scale; factor 4 represents the Connecting scale; factor 5 represents the Evaluating scale; and factors 6 represents the Selecting scale. These factors were composites of strategy items and they were represented as indicators associated with the proposed latent factors in the next section. For example, Selecting, Connecting, and Organizing were hypothesized to be the indicators of the latent factor Rhetorical Strategy Use.

Table 4.4 EFA results for the Strategy Inventory for Integrated Writing: Oblimin Rotation

Item No.	SELFS MON F1	TWS TW F2	RHES ORG F3	RHES CON F4	SELFS EVAL F5	RHES SEL F6
32 (MON7)	.794					
24 (MON3)	.764					
26 (MON5)	.741					
30 (MON6)	.669					
21 (MON2)	.578					
16 (MON1)	.443					
25 (MON4)	.438					
33 (MON8)	.403					
20 (TW3)		.733				
14 (TW1)		.655				
18 (TW2)		.649				
10 (ORG2)			.866			
9 (ORG1)			.771			
13 (ORG4)			.719			
11 (ORG3)			.631			
4 (CON1)				.687		
5 (CON2)				.603		
7 (CON3)				.472		
8 (CON4)				.401		
15 (EVAL1)					.445	
29 (EVAL2)					.437	
34 (EVAL3)					.436	
12 (SEL2)						-.514
6 (SEL1)						-.463

Note. RHES = Rhetorical Strategy Use, SELFS = Self-Regulatory Strategy Use, TWS = Test-Wisness Strategy Use, ORG = Organizing, SEL = Selecting, CON = Connecting, MON = Monitoring, EVAL = Evaluating, TW = Test-Wisness

As seen in Table 4.4, the oblimin factor solution was used to form six strategy-type composite variables to be used in the succeeding confirmatory factor analyses. This solution produced three latent factors: (1) the Rhetorical Strategy Use factor including strategy-type items related to Selecting, Organizing, and Connecting, (2) the Self-Regulatory Strategy Use factor including strategy-type items dealing with Monitoring and Evaluating, and (3) the Test-Wisness Strategy Use factor including three test-

wiseness strategy items: TW1, TW2 and TW3. Following Kenny’s (1979) “Two-Indicator Rule,” a measurement model can only be identified if every latent factor is associated with at least two indicators. In this regard, the latent factor, Test-Wiseness Strategy Use, was represented by three inventory items (TW1, TW2 and TW3) rather than one composite of strategy items (TW). Table 4.5 shows a summary of the items used in the composite for each strategy-type scale.

Table 4.5 Composites for the integrated writing strategy use variables (24 items)

Variables	No.	Item used
1. Rhetorical Strategy Use		
• Organizing (ORG)	4 items	9, 10, 11, 13
• Selecting (SEL)	2 items	6, 12
• Connecting (CON)	4 items	4, 5, 7, 8
Subtotal	10 items	
2. Self-regulatory Strategy Use		
• Monitoring (MON)	8 items	16, 21, 24, 25, 26, 30, 32,33
• Evaluating (EVAL)	3 items	15, 29, 34
Subtotal	11 items	
3. Test-Wiseness Strategy Use		
• Test-Wiseness 1 (TW1)	1 item	14
• Test-Wiseness 2 (TW2)	1 item	18
• Test-Wiseness 3 (TW3)	1 item	20
Subtotal	3 items	
Total	24 items	

The integrated reading-listening-writing test

Similar to the statistical procedures used for the strategy inventory, the preliminary analysis for the writing test first examined the normality of the data on integrated writing test scores. Skewness and kurtosis coefficients were both used to test two types of normality violations separately. In addition, descriptive analyses including mean, standard deviation and Cronbach's coefficient alpha for score reliability were reported. Following the preliminary analyses, a series of EFAs were performed to

determine whether each item (i.e., Content, Organization, Language Use and Verbatim Source Use) corresponded with its underlying construct.

Distributions and reliability

This section presents the results of item-level²² analyses for the integrated writing test based on all 161 test takers. Table 4.6 presents the descriptive statistics for four items based on the scores given by the two different raters. The means for the Content feature were 3.11 for rater 1 and 2.92 for rater 2, suggesting a high degree of rater agreement between the two raters on this item. The standard deviations for the two raters were 1.44 and 1.28 respectively. The means for the Organization feature were 3.82 for rater 1, and 3.86 for rater 2. This result revealed a high level of inter-rater reliability. The standard deviation was 1.09 for rater 1 and 1.02 for rater 2. In terms of the Language Use feature, rater 1 had a mean of 3.71 and rater 2 had a mean of 3.70. The standard deviation was 1.09 and 1.02 respectively. The means for the Verbatim Source Use feature were 3.02 for rater 1 with a standard deviation of 1.97, and 2.99 for rater 2 with a standard deviation of 1.94. The values for skewness ranged from -.21 to -1.24, and kurtosis ranged from -.40 to 1.35. The skewness and kurtosis values were within the acceptable limits (i.e., all values were lower than the absolute value of 3), suggesting the items were normally distributed.

²² The item here refers to the writing features or writing scale items, such as content, organization and language use.

Table 4.6 Distributions for the integrated writing test (N = 161)

Variable	Mean	SD	Skewness	Kurtosis
R1-CON	3.11	1.44	-.29	-.83
R2-CON	2.92	1.28	-.21	-.72
R1-ORG	3.82	1.09	-1.24	1.15
R2-ORG	3.86	1.02	-.93	.48
R1-LU	3.71	1.09	-.52	-.58
R2-LU	3.70	1.02	-.56	-.40
R1-VSU	3.06	1.97	-.52	-1.35
R2-VSU	2.99	1.94	-.47	-1.35

Note. R1 = Rater 1, R2 = Rater 2, CON = Content, ORG = Organization, LU = Language Use, VSU = Verbatim Source Use.

Table 4.7 presents the inter-rater reliability estimates for the four items using Pearson's product-moment correlation. All these scales produced high reliability estimates, ranging from .93 to .99.

Table 4.7 Reliability estimates for the integrated writing test

Variables	Items used (scores)		Reliability estimates (α)
Content (CON)	R1-CON	R2-CON	.94
Organization (ORG)	R1-ORG	R2-ORG	.93
Language Use (LU)	R1-LU	R2-LU	.97
Verbatim Source Use (VSU)	R1-VSU	R2-VSU	.99

Note. R1 = Rater 1, R2 = Rater 2, CON = Content, ORG = Organization, LU = Language Use, VSU = Verbatim Source Use.

Item-level EFAs for the integrated writing test

A series of item-level EFAs were performed in order to investigate the clustering of the test performance items. The KMO measure of sampling adequacy revealed an adequate value of .72, suggesting that factor analysis was appropriate for these data. The result of Bartlett's measure of sphericity ($p < .001$) indicated that the original correlation matrix was not an identity matrix and thus factor analysis was an appropriate analysis.

A two-factor oblimin solution seemed to maximize parsimony and interpretability. The initial factor extraction yielded two eigenvalues greater than 1.0, accounting for 51.58% of the variance. As seen in Table 4.8, this solution produced (1) a

L2 Writing Ability factor including items dealing with content, organization, and language use and (2) a Verbatim Source Use factor which consisted of an item relating to verbatim source use. However, because the Verbatim Source Use factor contains only one item, Verbatim Source Use, it was then viewed as a single indicator variable rather than a factor in the subsequent analyses. Following Kline (2004), a factor cannot be adequately measured by a single indicator. In sum, integrated writing performance, as measured by the integrated reading-listening-writing test, is a two-dimensional construct composed of general L2 Writing Ability and Verbatim Source Use.

Table 4.8 Results of EFA for integrated writing ability variable: Oblimin Rotation

Variable	L2 Writing Ability F1	Verbatim Source Use Ability F2
Content	.797	
Organization	.786	
Language Use	.818	
Verbatim Source Use		.294

Summary

Using reliability analyses and exploratory factor analyses, this section describes the process used to establish convergent and divergent validity for the strategy inventory and test performance. In other words, these item-level analyses allowed examination of the factorial structures of integrated writing strategy use, as measured by the Strategy Inventory for Integrated Writing, and integrated writing test performance as measured by the integrated reading-listening-writing test. From these exploratory analyses, five strategy-type variables (Organizing, Selecting, Connecting, Monitoring and Evaluating) and three strategy item variables (TW1, TW2 and TW3) were hypothesized to measure three strategy use variables: Rhetorical Strategy Use, Self-Regulatory Strategy Use, and Test-Wisness Strategy Use. Given the design of the integrated writing test used in this

study, four integrated writing test performance variables were identified (Content, Organization, Language Use and Verbatim Source Use), and were hypothesized to measure two underlying subsets of the integrated writing performance: L2 Writing Ability and Verbatim Source Use.

Although EFAs provide information of construct dimensionality, CFAs can be used to determine whether a hypothesized factor model fits or does not fit the sample data. In the following section, two measurement models were constructed to test the hypotheses regarding the underlying constructs of integrated writing strategy use and integrated writing test performance. More specifically, the relationship between observed variables (e.g., Organizing) and latent variables (e.g., Rhetorical Strategy Use) were examined using a series of CFAs. Subsequently, structural models regarding the interrelationship between strategy use variables and test performance variables were proposed based on integrated writing and second language assessment literature.

MODELING INTEGRATED WRITING STRATEGY USE AND TEST PERFORMANCE

The results of item-level factor analyses provide invaluable information about plausible composites for strategy use variables and writing test performance variables. Based on these results and previous literature on integrated writing and second language assessment, a number of competing models regarding the construct of integrated writing strategy use and the integrated reading-listening-writing test performance were proposed in a confirmatory mode of SEM applications. Once the connections between the observed and latent variables in each measurement model were determined, the relationships among the latent variables were examined. This was done by proposing a number of structural models and examining model fit and interpretability. The first, second, and third research questions will be addressed in the following sections.

The strategy inventory for integrated writing

This section addresses the first research question “What is the nature of strategy use in an integrated reading-listening-writing test?” More specifically, this part investigates the relationship between integrated strategy-type indicator variables (e.g., Organizing, Selecting, and Connecting) and their underlying strategy use factors (e.g., Rhetorical Strategy Use). To answer this question, the normality assumptions for each observed variable were examined prior to estimating measurement models.

Distributions and reliabilities

Based on the previously discussed EFAs results, eight strategy-type variables (see Table 4.5) were generated. Table 4.9 shows a summary of descriptive statistics for each strategy type. The means ranged from a low 3.00 for the reported use of the Evaluating strategies to a high 4.00 for the reported use of the Connecting strategies. The standard deviation ranged from .75 to 1.32. The values for variable skewness and kurtosis were all within the limits, suggesting that these strategy-type variables were normally distributed.

Table 4.9 Distributions for the integrated writing strategy use variables (N = 161)

Variable	Mean	SD	Skewness	Kurtosis	Min	Max
Rhetorical Strategy Use						
ORG	3.85	.86	-.70	.11	1.00	5.00
SEL	3.92	.87	-.61	-.31	1.50	5.00
CON	4.00	.75	-.89	.67	1.50	5.00
Self-Regulatory Strategy Use						
MON	3.56	.81	-.36	-.54	1.36	5.00
EVAL	3.00	.90	-.18	-.68	1.00	5.00
Test-Wiseness Strategy Use						
TW1	3.26	1.32	-.31	-1.02	1.00	5.00
TW2	3.06	1.31	-.15	-1.11	1.00	5.00
TW3	3.17	1.27	-.33	-.97	1.00	5.00

Note. ORG = Organizing, SEL = Selecting, CON = Connecting, MON = Monitoring, EVAL = Evaluating, TW = Test-Wiseness.

Table 4.10 presents the reliability estimates for the integrated writing strategy variables. The reliabilities for the six strategy-type variables ranged from .54 for the Evaluating strategy scale to .86 for the Monitoring strategy scale. The reliability estimates for the new composite variables consisted of both high (ORG, MON, and TW) and moderate (SEL, CON, and EVAL) ones.

Table 4.10 Reliability estimates for the integrated writing strategy use variables (24 items)

Variables	No.	Item used	Reliability Estimates (α)
1. Rhetorical Strategy Use			
• Organizing (ORG)	4	9, 10, 11, 13	.85
• Selecting (SEL)	2	6, 12	.62
• Connecting (CON)	4	4, 5, 7, 8	.62
Subtotal	10		.81
2. Self-regulatory Strategy Use			
• Monitoring (MON)	8	16, 21, 24, 25, 26, 30, 32,33	.86
• Evaluating (EVAL)	3	15, 29, 34	.54
Subtotal	11		.85
3. Test-Wiseness Strategy Use			
• Test-Wiseness (TW)	3	14, 18, 20	.72
Subtotal	3		.72
Total	24		.87

Confirming the dimensionality of the strategy inventory for integrated writing

To validate the dimensionality of the Strategy Inventory for Integrated Writing, a series of CFAs were performed. *A priori* for the following CFAs included the number of factors, the factorial structure (i.e., which item that loads on which factor), and the relationships among factors (i.e., whether the factors are correlated). Three measurement models (Model 1.1, Model 1.2, and Model 1.3) were proposed and tested in the following section.

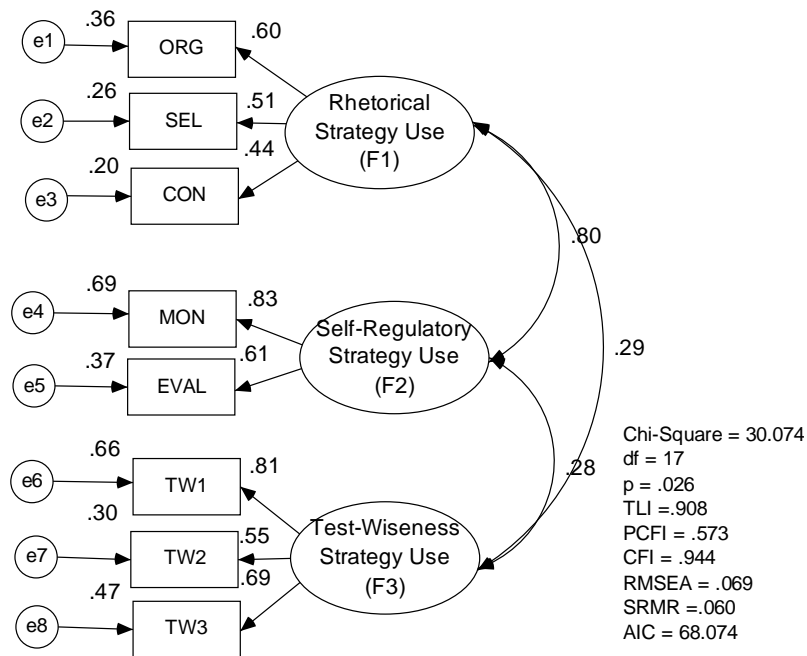
The hypothesized model: Model 1.1

The results of EFAs and reliability analyses indicate that the strategy inventory for integrated writing represents a three-factor model of integrated writing strategy use. This model contains three correlated factors (Rhetorical Strategy Use, Self-Regulatory Strategy Use and Test-Wiseness Strategy Use) with eight observed variables (SEL, ORG,

CON, MON, EVAL, TW1, TW2, and TW3). This model does not include correlations between the error associated with each observed variable.

A first-order confirmatory factor analysis was performed to test the hypothesized multidimensionality of strategy use in the integrated reading-listening-writing task. Specifically, this model was designed to test the hypothesis that integrated writing strategy use is a multifaceted construct composed of (1) Rhetorical Strategy Use consisting of the Selecting, Organizing and Connecting strategies (Spivey, 1990, 1997), (2) Self-Regulatory Strategy Use consisting of the Monitoring and Evaluating strategies (Asención, 2004; Esmaeili, 2002; Stein, 1990), and (3) Test-Wiseness Strategy Use consisting of three Test-Wiseness strategies regarding the inappropriate use of source materials (Cohen, 2006; Cumming et al., 2005). The goal of this statistical procedure is to generate the best-fitting and most meaningful model that supports the theoretical views of integrated writing strategy use. To date, a number of descriptive studies have supported the multidimensionality of integrated writing strategy use during reading-writing tasks (e.g., Asención Delaney, 2008; Plakans, 2008; Spivey, 1984), but no study has modeled and tested integrated writing strategy use during reading-listening-writing tasks. Figure 4.1 presents Model 1.1.

Figure 4.1 Initially-hypothesized model of integrated writing strategy use: Model 1.1



Note. ORG = Organizing, SEL = Selecting, CON = Connecting, MON = Monitoring, EVAL = Evaluating, TW = Test-Wiseness.

Estimation results for Model 1.1

Prior to estimating the measurement model, the normality assumption associated with the maximum likelihood estimation procedures used in SEM was checked. As seen previously in Table 4.9, the skewness and kurtosis values for the individual strategy-type variables were all within the limits, suggesting that these variables were normally distributed.

Once the normality assumption was achieved, the hypothesized Model 1.1 was assessed to determine to what extent the model fit the data. In terms of model adequacy as a whole, the initially-hypothesized three-factor model of integrated writing strategy use produced a chi-square value of 30.074 with 17 degrees of freedom ($p < .05$), indicating a global misfit of the model to the sample data. Other fit indices were also used

to examine model fit, including the Tucker-Lewis index (TLI) and Bentler's comparative fit index (CFI). Model 1.1 produced a TLI of .908 and CFI of .944, suggesting an adequate fitting model. Standardized root-mean square residual (SRMR) value in this model exceeds the value .05, indicating the model is beyond acceptable.

To disentangle the mixed results from the fit indices, significance of individual parameters were examined. The loadings of eight observed variables on the respective factors were highly significant ($z \geq 2$), as were the variances and the covariances among factors. Inspection of standardized residuals and modification indices indicated some localized point of ill fit in this solution (e.g., some standardized residual covariances were larger than the absolute value of 2 and some modification indices had a value larger than the absolute value of 4). Because the model was not a good fit to the data, model parameters were not interpreted.

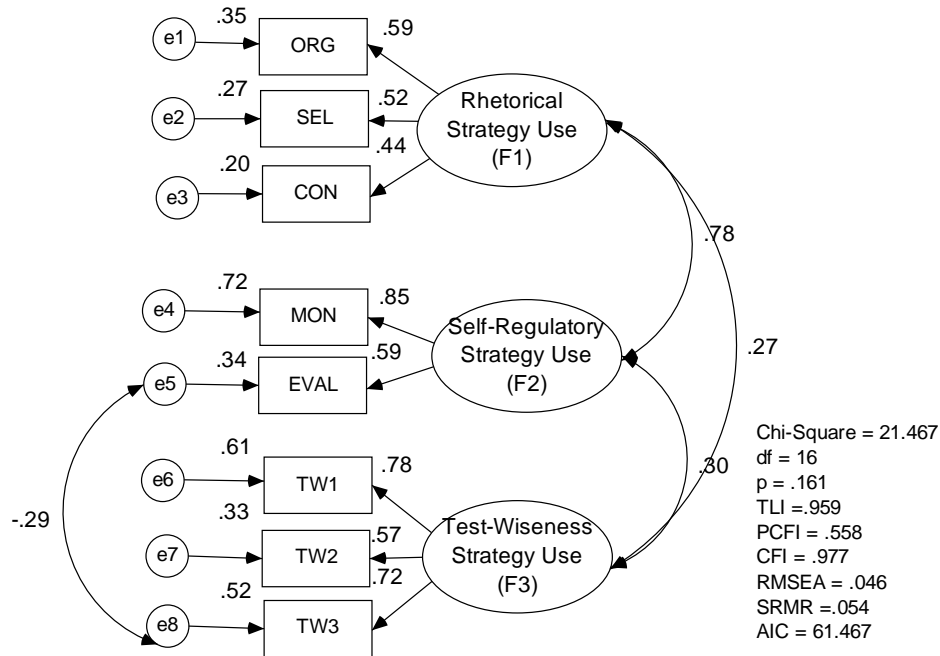
The hypothesized model: Model 1.2

Based on the results of model estimation for Model 1.1, a series of *post-hoc* fitting procedures were performed to examine alternative models. One change made from Model 1.1 was relaxing a constraint to the model by estimating a covariance parameter between the errors associated with the Evaluating strategies (e5) and with the TW3 strategy item (e8). The modification index suggested that freeing this path resulted in a decrease in χ^2 of at least 7.787. This specific change may indirectly indicate that the Evaluating strategies and the Test-Wisness (TW3) strategy item measured something in common beyond their respective factors. To explain, Item TW3 and the Evaluating items shared something in common in addition to the proposed factors. Some of the shared variance was due to the latent factor and some of the shared variance was due to another exogenous common cause.

It made sense when Item TW3 was examined with one of the Evaluating items. Item TW3 “I copied the sentences from the reading passage and revised them” seemed to be negatively related to EVAL 3 “I made changes in phrases to ensure I didn’t copy the exact phrases” in terms of sentence paraphrasing strategy. The difference was that TW3 was often used by test takers who were not skilled in paraphrasing sentences from sources whereas EVAL3 was usually used by test takers who had paraphrased sentences by rewriting them in their own words and checked if they had copied some exact phrases from sources. This paraphrasing strategy was a subset of the Evaluating and Test-Wisness strategy scales that was not hypothesized to measure in the first place. Another source of this correlated error may be due to the similar wording in both items. This information is useful for revising these items and enhancing scale reliabilities.

Similar to Model 1.1, Model 1.2 contains three correlated factors (Rhetorical Strategy Use, Self-Regulatory Strategy Use and Test-Wisness Strategy Use) with eight observed variables (SEL, ORG, CON, MON, EVAL, TW1, TW2, and TW3). The errors associated with the Evaluating strategies and TW3 was postulated to be correlated from a substantive and statistical point of view.

Figure 4.2 Hypothesized model of integrated writing strategy use: Model 1.2



Note. ORG = Organizing, SEL = Selecting, CON = Connecting, MON = Monitoring, EVAL = Evaluating, TW = Test-Wiseness.

Estimation results for Model 1.2

The fit statistics for Model 1.2 indicated a good fit to the data. The chi-square is 21.467, with 16 degree of freedom ($p = .161$), suggesting a good model fit. The RMSEA was below .05 (.046, 90% confidence interval = .000–.506), and the TLI and CFI were all above .95. The SRMR for this model was below .08 (.054), meaning that the matrix implied by the model differed from the actual correlation matrix by only .054. Hu and Bentler (1999) recommended joint criteria to retain a model, and this positive combination of indices provided confirmation of a well-fitting model. All estimated parameters including factor loadings and covariances were statistically significant ($z \geq 2$) except for the covariance between factors of Rhetorical Strategy Use and Test-Wiseness Strategy Use. Some localized point of ill fit indicated in standardized residuals and

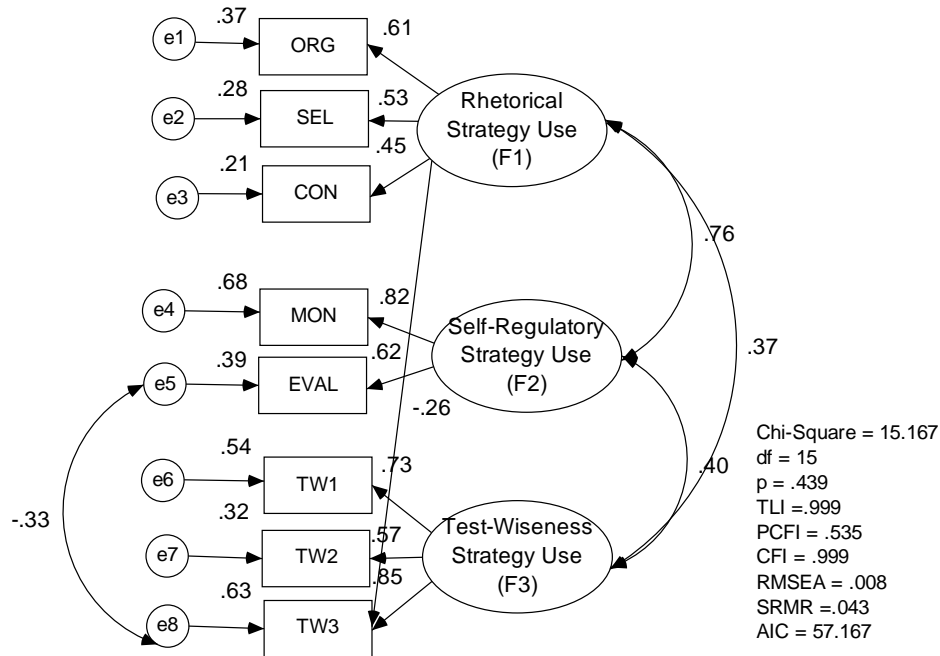
modification indices were found in this solution. All factor loadings were found to be statistically significant at the level of .05.

The hypothesized model: Model 1.3

Based on the results of Model 1.2 and literature in L1 and L2 integrated writing, a series of *post hoc* modification procedures were performed. In recognizing the multifaceted nature of Item TW3, Item TW3 was hypothesized to cross-load with both Test-Wisness Strategy Use and Rhetorical Strategy Use. Previous research exploring writers' ability to use source information suggests that inappropriate use of source texts often results from inadequate summarizing skills to select, organize and paraphrase information from source texts (Brown & Day, 1983; Howard, 1995; Kennedy, 1985).

Model 1.3 is a first-order confirmatory factor analysis designed to test the linkage between the Rhetorical Strategy Use and test takers' reported patchwriting strategy (TW3). Similar to Model 1.2, Model 1.3 contains three correlated factors (Rhetorical Strategy Use, Self-Regulatory Strategy Use and Test-Wisness Strategy Use) with eight observed variables (SEL, ORG, CON, MON, EVAL, TW1, TW2, and TW3). The uniqueness associated with TW3 and "Evaluating" strategies was hypothesized to be correlated. Item TW3 was postulated to cross-load with Test-Wisness Strategy Use and Rhetorical Strategy Use based on substantive and statistical points of view.

Figure 4.3 Hypothesized model of integrated writing strategy use: Model 1.3



Note. ORG = Organizing, SEL = Selecting, CON = Connecting, MON = Monitoring, EVAL = Evaluating, TW = Test-Wiseness.

Estimation results for Model 1.3

As seen in the Figure 4.3, Model 1.3 provided an excellent fit to the data: the chi-square was not significant, RMSEA was below .05 (.008), and the CFI and TLI values were all above .95 (.99). All loadings of eight indicator variables on the respective latent factors were highly significant ($z \geq 2$), as were the variances and covariances among latent factors. No indicator of ill fit was found in this solution (e.g., largest standardized residual = -1.609). The modification indices did not suggest any model revisions.

Comparing competing models

To test the hypothesis regarding the nature of Item TW3, the previously hypothesized model, Model 1.2, Model 1.3, were compared using a chi-square difference test. Table 4.11 shows the comparison of fit indexes for two proposed models of

integrated writing strategy use. Since the $\Delta \chi^2$ was significant for Model 1.3 ($\chi^2_{\text{diff}}(1) = 6.3, p < .05$), the less parsimonious model—the model with lower degree of freedom—was favored. In sum, the statistical results provided strong evidence for acceptance of Model 1.3.

Table 4.11 Comparison of fit indexes for Model 1.2 and Model 1.3

Model	χ^2	<i>df</i>	$\Delta \chi^2$	Δdf	<i>p</i>	AIC	CFI	RMSEA (90% CI)
Model 1.2	21.467	16				61.467	.99	.046 (.000–.506)
Model 1.3	15.167	15	6.3	1	.01	57.167	.99	.008 (.000–.777)

Discussion of quantitative results: Model 1.3

As seen in Figure 4.3, the graphic representation of Model 1.3 indicates that integrated writing strategy use is represented by three intercorrelated underlying factors measured by eight observed variables. These findings provide some valuable information on the nature of strategy use during the integrated writing test. Each factor is well represented by its proposed observed variables—Self-Regulatory Strategy Use is represented by the Monitoring and Evaluating strategies, Rhetorical Strategy Use is represented by the Organizing, Selecting, and Connecting strategies, and Test-Wisness Strategy Use is well measured by three test-wisness strategies (TW1, TW2, and TW3).

The relatively high positive relationship (.76)²³ between Rhetorical Strategy Use and Self-Regulatory Strategy Use suggests that test takers who used the rhetorical strategies more frequently tended to use the self-regulatory strategies also. On the other hand, the relatively low inter-factor correlations (.37 and .40) between Test-Wisness Strategy Use and two other factors suggest that the test-wisness strategies are less associated with the rhetorical strategies and the self-regulatory strategies. In addition to

²³ This high correlation between the “Rhetorical Strategy Use” and the “Self-Regulatory Strategy Use” may be indicative of a higher order factor. However, for the purposes of this study, Model 1.3 was used because the design of separate latent factors allowed for investigation of the relationship between strategy use factors and integrated writing test performance.

the correlations among factors, the strategy item TW3 (i.e., a patchwriting strategy) was found to load both on its proposed strategy factor, Test-Wiseness Strategy Use (.85), and Rhetorical Strategy Use (-.26).

These findings seem to support three notions: (1) the interactive dimensions of the rhetorical strategies and self-regulatory strategies, (2) Test-Wiseness Strategy Use as a relatively independent dimension of Rhetorical Strategy Use and Self-Regulatory Strategy Use, and (3) the complexity of the patchwriting strategy (TW3). First of all, the rhetorical strategies function in concert with the self-regulatory strategies. They should be viewed as two interactive dimensions of the same mental mechanisms that are intricately intertwined with one another. Tsai (2004), Phakiti (2003) and many others claim that most cognitive-related strategies occur in association with metacognitive-related strategies²⁴. For instance, test takers need to activate their metacognition that regulates their performance on cognitively demanding tasks. Similarly in this study, when test takers are asked to perform the writing task, such as summarizing key points from sources, they need to evaluate their task effectiveness and fulfillment of task requirements before, during and after they write.

The second notion supported by the findings was that Test-Wiseness Strategy Use can be seen as a relative independent facet of integrated writing strategy use because the use of test-wiseness strategies had little relationship with the self-regulatory and rhetorical strategies.

The third notion is concerned with the negative linkage between the patchwriting strategy and Rhetorical Strategy Use. The negative loading of the patchwriting (TW3) strategy on Rhetorical Strategy Use suggests that the more test takers use the rhetorical

²⁴ Although self-regulatory strategy is usually defined as a metacognitive strategy in the literature, in this data its use often combine with cognitive tactics. Also, rhetorical strategy is more concerned with the fine-grained cognitive operations that test takers used to complete the integrated writing test task.

strategies (i.e., organizing, selecting, and connecting), the less likely they are to use the patchwriting strategy. This finding suggests that Item TW3 might be a behavior that is strongly associated with the Test-Wiseness strategies, and is counterproductive during the use of rhetorical strategies. In addition to the cross-loading, the uniqueness associated with Item TW3 (e8) was found to correlate with uniqueness that associated with the Evaluating strategies (e5). These results in Model 1.3 indicate that the TW3 strategy item is a complex one in that (1) it is related to more than one underlying factor and this relationship is both positive and negative and (2) its uniqueness is correlated with another uniqueness associated with another strategy type. To provide further explanation of this model, qualitative results are reported in the following section.

Discussion of qualitative results

Following the coding procedures presented in Chapter Three, all qualitative data were first read for patterns. These patterns were then coded and examined in terms of their connections with different phases and subphases of completing the integrated writing task. The qualitative data collected from interviews, test takers' planning sheets, and the open-ended section of the strategy inventory were found to support these two notions shown in the quantitative data and add further insight into how test takers in the high- and low-performance groups used different strategies in response to the integrated writing task.

Pre-composing phase: Preparation, reading and listening

When comparing test takers' strategy use in the two performance groups, the test takers in the high-performance group reported a wider range of variance across writers than the test takers in the low-performance group. In general, it was found that higher language proficiency allows strategies to be used more effectively and flexibly whereas lower language proficiency hampers effectiveness of all types of strategy use. The

qualitative data also shows some counter-examples. The following section addresses how writers in different groups used strategies to interact with two texts in terms of three subphases: preparation, reading and listening.

Subphase 1: Preparation

The first subphase, preparation, revealed some differences in strategy use between the test takers from the high- and low-performance groups. The test takers from the high-performance group reported reading the writing directions carefully and subsequently constructing a mental blueprint about how they could manage time to approach the task. Five students—Feng, Dana, Haley, Richard and Fulcher—reported knowing exactly what they were required to do and were ready to process information from the reading and the lecture in a timely fashion although they were not sure about the relationship between the passage and the lecture (e.g., one is agreeing or disagreeing with the other). Richard mentioned how he reacted to the writing directions by setting goals, planning actions and carrying out his plans:

Because the task required me to compare and contrast ideas in the reading and the lecture, I tried to organize the ideas the way they were presented in the reading and the lecture. I folded the planning sheet in half and noted down three pros on one side and three cons on the other. This really helped me a lot.

For the test takers in the low-performance group, some of them remembered not paying much attention to the writing directions. Although all students were told about the procedures of the test and what they were expected to do prior to the test, they were encouraged to read through the writing directions for better comprehension. Four students—Kim, Lin, Kate, and Sara—reported skipping the writing directions in the first place and considered that part as unimportant. Kate declared, “I already know what I need to do. I don’t need to waste my time.” Two students, Jimmy and Wokie, mentioned that they expected the two texts to support each other, and this false expectation impeded

their level of comprehension for the lecture. For example, Jimmy pointed out, “I expect that the listening will be the same ideas of the reading, but after hearing it, I see oh... What’s this? [surprised face]. That’s why I don’t hear the first part.”

Subphase 2: Reading

In this phase, test takers had time to read a passage on an academic topic for approximately two minutes. They were allowed to take notes as they read.

High-performance group

Two types of strategies—organizing and selecting—were reported extensively by the test takers in the high-performance group. However, these strategies were used in different ways. For example, four students—Feng, Tonnie, Dana, Richard, Lossie, Liu and Yvonne—reported skimming and scanning for major points quickly and then going back to read for details. They attended to the organization of the passage and focused only on the main ideas. Organizing strategies were reportedly used to guide them through the reading and to help them comprehend major ideas of the reading. For instance, Yvonne mentioned how she read the passage:

I paid special attention to first one or two sentences because I need to know, ok what is the topic? Ok it’s about fish farming. It’s not a topic that I am familiar with. And I kind of skim the topic sentence and I read the first sentence of each paragraph because I know each paragraph because I know it’s important to know the basic idea about each paragraph.

In addition to organizing and selecting strategies, two interesting patterns emerged: different note taking styles and different ways for dealing with unknown words. In terms of note-taking, some students in this group found it helpful while others, Yawen and Haley, found it “a waste of time” and “unnecessary” respectively. Further inspection of Yawen’s planning sheet suggested that she spent time writing down full sentences from the reading and it might have affected her efficiency for the task. Haley mentioned that she did not need to take notes because she could “always go back to the reading after

the lecture.” With regard to dealing with unknown words, two students, Feng and Liu, both pointed out that they tried to either guess the meaning of those words from their contexts or just ignored them as long as they could comprehend the overall ideas of the passage. For example, Liu mentioned, “Though I don’t know all the words in the reading, I still can understand the text. I just skip those I don’t know.”

During reading, the high performing students used their schemata of the typical relationships among information units to identify the major points in the reading passage. Also, almost every student in this group took notes to improve reading comprehension or retain information for later use. Some of them jotted down only key words and others wrote down full sentences. When they had difficulties understanding some vocabulary words, they were able to tolerate a little incomprehension and move on without comprehending everything.

Low-performance group

Most of the test takers in the low-performance group reported using the selecting strategies extensively and few reported using the organizing strategies. Although some of them mentioned the use of both strategies, they also talked about some problems they had when using them. Five students—MD, Chih, LJ, Tori, and Kim—mentioned that they learned how to select key points for writing based on the organization of the passage when they were in school while they were not able to do so in this test. For example, MD said,

I read all of them very fast and catch the main idea. But I think this way is not so suitable for me because sometimes if this topic is the architecture or something more difficult to me. The topic this time... I read all of them... but maybe I didn’t catch the point.”

The rest of the students mentioned reading through the whole passage from the beginning to the end. As many students read the passage, they reported taking notes from

the reading. Similar to students in the high-performance group, some of these students reported writing down keywords while others reported writing down whole sentences from the reading. One student, Sara, mentioned that when she did not understand some words in the passage, she tried to understand them in their contexts. Two students, Elle and Kim, found themselves stuck in the reading when they did not understand some words. Elle said, “Like this word [pointing the word “jeopardize”], I don’t understand. I am wasting my time because I don’t understand some words.” These students also said they noted down unknown words from the reading in case those words were key ideas in the reading passage.

In the reading phase, the low performing students looked for major ideas and paid little attention to the structure of the passage. Similar to students in the high-performance group, most of these students took notes during reading but the ways they took notes were less systematic and organized. When they found unknown words in the reading, they mentioned not being able to leap over those words and finish reading the sentence. These students also reported writing down unknown words for later use in the writing.

Subphase 3: Listening

In this phase, the test takers listened to a lecture on an academic topic closely related to the reading passage for two minutes. They were allowed to take notes as they listened to the lecture. All three types of rhetorical strategies were reported in this subphase. However, one’s language proficiency, again, played a major role in terms of the quality and effectiveness of strategy use.

High-performance group

The selecting and connecting strategies were most widely reported in this subphase by the high performing test takers. The selecting strategies were reportedly used in concert with the connecting strategies. The note-taking strategies, on the other hand,

were found to bridge the use of these two strategy types. For example, Feng emphasized the important role of note taking in completing the task:

The purpose of note-taking is to help me memorize ideas in the lecture and also help me compare the reading and the listening. The way I did it was to grasp a keyword from the reading, and from the listening, and then compare them. Of course I took notes for both reading and listening. I couldn't organize everything in my mind, so the notes helped me do the comparison, and the notes come in handy for my writing.

Only one student, Liu, found that the organizing strategies helped her concentrate when listening to the lecture. Liu reported paying extra attention when she heard words such as “illustrate” or “explain.” She mentioned predicting some key points or concluding points after those transition words.

In this phase, most of the high performing students mentioned using the selecting and connecting strategies via note-taking. Only one student mentioned using the organizing strategies to increase her listening comprehension. It is possible that the lecture was not perceived difficult to comprehend by high performing students and thereby students were able to understand the text without using the organizing strategies.

Low-performance group

The listening section of the integrated writing test posed the most difficulties for the test takers in the low-performance group. Their English listening comprehension was not adequate enough for the selecting and connecting strategies to be used effectively. However, most low performing students reported using the organizing strategies while listening to the lecture to improve their comprehension. One student, Kim, viewed transition words as red flags reminding him to start paying attention to the subsequent key points. As he stressed, “there are some arguments used such as *however*, *contrary*, conjunction words. I listen to things after those.” He relied on the organizing strategies for better comprehension of the lecture.

Seven students—Kate, Wokie, Jimmy, Chih, Elle, MD, and Tori—found the lecture to be difficult. For example, Wokie said he tried to “catch the main point” while at the same time pointed out that he only understood “very few” of the lecture. Elle reported spending most of her time “guessing” what the speaker was talking about, and writing down something she did not understand because she was afraid that those were the key points of the lecture. Tori described how she tried to figure out the lecture, “I just guessed some information from the other part, the reading part.” One of the students, Chih, expressed his feelings when he could not understand the lecture, “I skipped all lecture. That’s my habit, if I don’t understand something I will lose my patience I feel angry ...I feel angry about why I can’t understand.” Most of the other test takers simply stated their incomprehension of the lecture.

Because these test takers mentioned having problem comprehending the lecture, they also found it difficult to take notes for what they heard as well as to connect the ideas in the lecture with those in the reading passage. For instance, Kate described how she could not take good notes of the lecture, which in turn inhibited her from connecting information in both sources:

I took notes about this topic when I wrote. Just fish-farming. That’s all. I just remember why fish farming...the commercial...I don’t understand the listening parts... I didn’t include... add lecture content. It’s related to my fist problem... I didn’t think. I put my idea... Because I didn’t listen the lecture....That’s my mistake.

In terms of the potential source of incomprehension, Kim and Sara thought the speaker’s voice distracted them from paying attention to the lecture. For example, Kim explained how the voice affected him, “The adaptation to the voice. I need to get accustomed to the voice after a while.” More specifically, Sara made note of the speaker’s voice quality by saying, “It reminds me the TOEFL. I don’t like the TOEFL. I don’t like the voice, too low and monotone.”

In the listening subphase, most of the low performing students tried to comprehend the lecture using the organizing strategies. However, due to their lack of listening comprehension, they mentioned having difficulties using the selecting and connecting strategies. These results corroborate other reading-to-write research (e.g., Grabe, 2001) that has emphasized that threshold comprehension ability is critical for integrated writing tasks to be completed.

Composing phase: Rereading and writing

In this phase, the reading passage was accessible to all test takers. They could choose to reread it or not during writing. They had 20 minutes to complete their writing.

High-performance group

When the test takers could get access to the passage again, four of them—Fulcher, Richard, Haley and Yawen—reported going back to reread it either for main points or for phrases that could be used in their writing. Fulcher briefly read it and “tried to organize the passage and the lecture format” in his mind so that he could “compare individual points.” Yawen went back to the passage to look for words that could be used in her own writing. “When I wrote the first point, I went back to see what he wrote. I forgot which word he used, so I went back to check. I would go back to see which are the words the author used, and used those words to write my own essay.”

The rest of the test takers chose not to reread the passage, and simply wrote based on their notes. They usually had a plan either on the planning sheet or in their minds about how they were to write their essays. For example, Yvonne said,

I didn't read it again while I wrote my writing. I think I've already got important information from that passage so I didn't. I just used notes for my writing. I start from the beginning to the end because before I started writing my article I know what should be included in each paragraph. If it's a long sentence I probably will write topic sentence for each. And then the details.

Several of the high performing test takers applied the monitoring and revising strategies when composing their essays. They reported checking the effectiveness of their performance by examining task goals and their written texts. Liu said, “I examined my essay after I finished writing. I wanted to know whether I was clear about my ideas and also if I made grammatical mistakes.”

When asked about the most effective strategies these test takers used to complete the task, individual test takers seemed to have different points of view. Yvonne thought reading strategies were most important. Feng believed note-taking helped him complete the task. Fulcher thought monitoring was the most effective strategy that guided him through writing:

I think...uh it's monitoring, including the content and also my language while I am writing probably. I might have the pre-constructed concepts in my mind about how I am going to write this paper...because it's clear about the comparison between the individual points so first you have to mention the ideas in the reading passage and then you have to prepare to listen to the notes lecture.

Low-performance group

Most of the test takers in this group mentioned rereading for main points before and during their writing. Kim and Elle thought that they wanted to save time for writing so they started writing as soon as the lecture was over. The rest of the students skimmed for main points as they wrote their essays. MD went back to read some parts of the reading passage again, “Sometimes when I was writing...I will go back to see a few of the reading part and that will help me to organize my writing better.” Tori read most part of the reading passage for the second time for better comprehension. “Because I don't understand what the reading part say,” she said.

Some of these students—Jimmy, Kate, Tori, MD—remembered reading the passage in order to look for words to be used in their writing. Jimmy described his process and purpose of rereading the passage, “I stared some, stared some, and write.

When I don't know how to write I tried to copy some places of the passage... and some just not with the whole part, just the key part." Kate felt that using phrases from the reading could help her "improve making a sentence." MD responded to how she used the information from the passage, "...because some specific words I can't memorize so I just go back to find the words and use it."

When it comes to writing the whole essay, two students, Kate and MD, who were preparing for the TOEFL at the time they took the test, said they applied previously memorized writing templates in their writing. Kate showed an example about how she used a memorized structure in her writing, "I wrote the listening *cast doubts* on the reading. That's the good phrase I memorized to present opposite."

Each student in this group was fairly aware of their task goal—to connect the reading passage with the lecture. When many of them found that they were not able to compare ideas from the two sources because of reading or listening comprehension problems, they were left no choice but to complete the essay based on their knowledge or personal experiences. Elle said, "I knew that the listening was talking about something opposite, something like...no evidence, but I don't get it. So I just made up something and fill in that part of my writing."

Summary

In summary, the quantitative results show that Model 1.3 provides a good explanation of integrated strategy use in this study. With its three intercorrelated latent factors and eight measured variables, this model supports the hypothesis that integrated writing strategy use is a multidimensional construct composed of Rhetorical Strategy Use, Self-Regulatory Strategy Use and Test-Wisness Strategy Use. Rhetorical Strategy Use is represented by inventory items associated with selecting, organizing, and connecting pieces of information from sources. Self-Regulatory Strategy Use is

represented by inventory items related to monitoring and evaluating. Test-Wisness Strategy Use is represented by a set of test-wisness strategy items. The inter-factor correlations also indicate that these three strategy variables cannot be seen to be purely independent from each other, but should be viewed as complicated behaviors related to one another.

The qualitative results support the concept of the interactive nature of Rhetorical Strategy Use and Self-Regulatory Strategy Use. Also, they complement quantitative results by indicating that the frequency of Test-Wisness Strategy Use, including the patchwriting strategy, seems to be slightly related to test takers' writing proficiency but largely dependent on test takers' confidence in expressing their ideas in their own words and their epistemological idea of using the exact phrases from the "authoritative" text to ensure their writing was "correct."

The qualitative results also provide insights about how test takers used different strategies at different phases and subphases. In the pre-composing preparation phase, the test-takers in the high-performance group were able to recognize their task goals and set up a plan to carry out the task. For the test takers in the low-performance group, half of them were very aware of their goals but did not have a plan in mind about how they could complete the task. The other half of these test takers only had some vague ideas about what they were asked to do, let alone planning ahead for action.

In the reading subphase, high performing students reported using all rhetorical strategies while low performing students mentioned focusing on the selecting and organizing strategies. In the listening subphase, most test takers in the high-performance group reported their use of the selecting and connecting strategies, and only one mentioned her use of the organizing strategies. However, those in the low-performance group found it difficult to use the selecting or connecting strategies because of their

limited listening comprehension skills. In the composing phase, writing/rereading writing, the use of all rhetorical strategies and self-regulatory strategies was salient in the interviews with the test takers from both groups. Since the test takers could only get access to the reading passage in the last subphase, they reported applying all the rhetorical strategies to reread the passage or their notes but not listen to the lecture.

The integrated reading-listening-writing test

This section addresses the second research question “What is the nature of test performance as measured by an integrated reading-listening-writing test?” In other words, this section investigates the relationship between analytic scores on different aspects of essays (e.g., Content) and their underlying factors (e.g., L2 Writing Ability). To answer this question, the normality assumptions for each observed variable were examined prior to estimating measurement models.

Distributions and reliabilities

Table 4.12 presents a summary of descriptive statistics for these four variables. The means ranged from 3.00 for the variable Content to 3.84 for the variable Organization. The standard deviation ranged from 1.02 for Content to 1.94 for Verbatim Source Use. The lowest mean scores of Content and Verbatim Source Use indicated that the task was challenging for test takers in terms of presenting comprehensive arguments from the reading passage and the lecture and avoiding the use of verbatim strings of words from source materials. The organization of writing posed the least problems with a mean score close to 4 (3.84). All skewness and kurtosis were within the acceptable limits, suggesting that the integrated writing test performance variables were normally distributed.

Table 4.12 Distributions of the integrated writing performance variables

Variable	Mean	SD	Skewness	Kurtosis	Min.	Max
Content	3.00	1.34	-.321	-.759	.00	5.00
Organization	3.84	1.02	-1.182	.979	.50	5.00
Language Use	3.70	1.04	-.571	-.488	1.00	5.00
Verbatim Source Use	3.02	1.94	-.509	-1.337	.00	5.00

Table 4.13 displays the reliability estimates for internal consistency for all integrated writing performance variables. The reliabilities for the four variables ranged from .93 for Organization to .99 for Verbatim Source Use. All the reliability estimates were fairly high.

Table 4.13 Reliability estimates for the integrated writing performance variables

Variable	No. of items ²⁵	Items used		Reliability estimates (α)
A. L2 Writing Ability				
Content	2	R1-CON	R2-CON	.94
Organization	2	R1-ORG	R2-ORG	.93
Language Use	2	R1-LU	R2-LU	.97
Subtotal	6			.92
B. Verbatim Source Use				
Verbatim Source Use	2	R1-SU	R2-SU	.99
Subtotal	2			.99

Note. R1 = Rater 1, R2 = Rater 2

Confirming the dimensionality of the integrated reading-listening-writing test

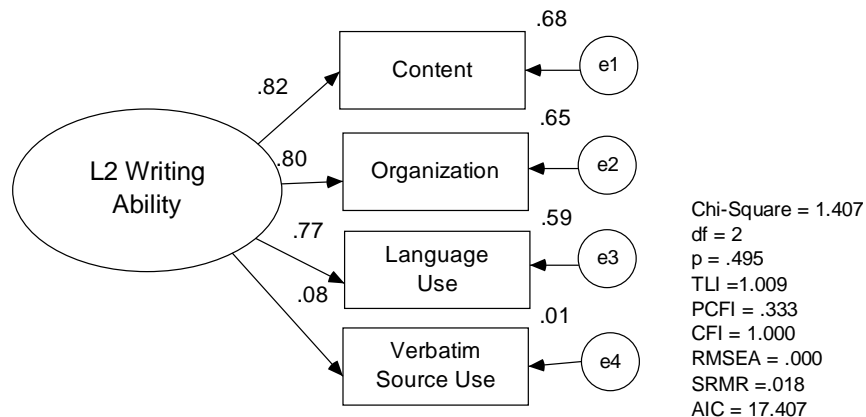
The dimensionality of the integrated reading-listening-writing test was further affirmed by a series of CFAs with a *priori* of the number of factors, the factor-item relationships, and the relationships among factors. Two measurement models regarding the facets of the integrated reading-listening-writing test were proposed in the following section.

²⁵ The items here refer to the ratings for each essay.

The hypothesized model: Model 2.1

One may hypothesize that a test takers' L2 writing ability can be presented by the measure of verbatim source use. Cumming et al. (2005) found fewer instances of verbatim phrases as test takers' proficiency increased for reading-writing tasks, while higher instances of verbatim phrases as test takers' proficiency level occurred for listening-writing tasks. However, no previous studies have looked at the relations between verbatim phrases and test takers' writing proficiency in a reading-listening-writing task. To test how test takers' writing abilities relate to their use of verbatim phrases from sources, the observed variable, Verbatim Source Use, was hypothesized to load on the latent factor, L2 Writing Ability, in Model 2.1. This model contains one factor loaded with four observed variables: Content, Organization, Language Use and Verbatim Source Use.

Figure 4.4 Initially-hypothesized model of integrated writing performance: Model 2.1



Estimation results for Model 2.1

As seen in Table 4.12, all individual variables in integrated writing performance were univariately normally distributed. Each of the overall goodness-of-fit indices suggested that Model 2.2 fit the data well ($\chi^2(2) = 1.407$, $p = .495$, SRMR = .018,

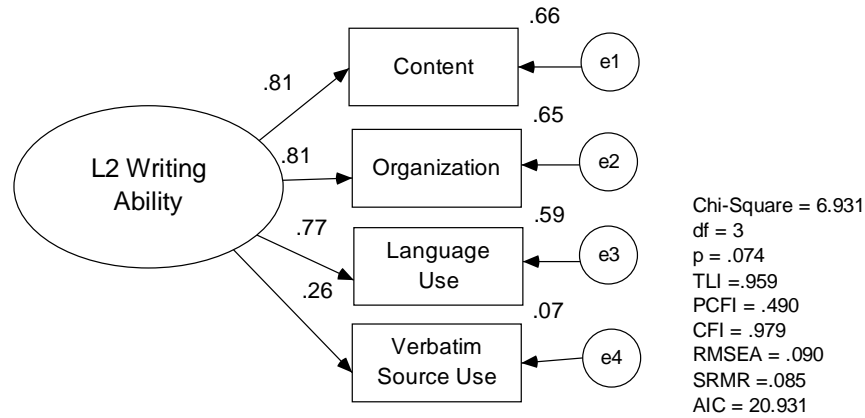
RMSEA < .001, TLI = 1.009, CFI = 1.000). Inspection of standardized residuals and modification indices showed no localized points of ill fit in this solution (e.g., largest standardized residual was -.579 and no modification index was listed). Standardized parameter estimates from this solution are presented in Figure 4.4. All freely estimated unstandardized parameters were statistically significant ($p < .001$) except for the path from L2 Writing ability to Verbatim Source Use. Because the correlation between the factor and Verbatim Source Use variable was not statistically significant, the parameter could be removed from this model, presumably without any noticeable loss of fit.

The hypothesized model: Model 2.2

The results of the item-level exploratory factor analyses (see Table 4.8) and reliability analysis suggested that integrated writing performance should be presented as a one-factor model. To test whether Verbatim Source Use was a stand-alone observed variable or a component of L2 Writing Ability, this variable was hypothesized to load on the L2 writing ability factor by constraining the loading to .5²⁶ and examining whether the chi-square of this model was significantly worse than that of Model 2.1. Model 2.2 was designed to test the hypothesis that integrated writing test performance includes two underlying components: L2 writing ability and verbatim source use. The hypothesis was derived from various studies that indicate the causes of verbatim source use may be reasons other than language proficiency (Howard, 1995; Kirkland & Saunders, 1991; Pennycook, 1996). A first-order confirmatory factor analysis was performed to test the multidimensionality of integrated writing test performance as measured by the integrated reading-listening-writing test.

²⁶ The loading of .5 is considered a small but reasonable value for a component hypothesized to be a part of a factor, given other loadings were relatively high at around .8. Therefore, the loading was fixed to be .5 to test if Verbatim Source Use can be considered an element of L2 Writing Ability.

Figure 4.5 Hypothesized model of integrated writing performance: Model 2.2



Estimation results for Model 2.2

To explore the trait structure of integrated writing test performance, the statistical assumptions underlying the maximum likelihood estimation procedures were examined. The univariate values for skewness and kurtosis for individual observed variables were all within acceptable range. No significant violations of normality were identified.

The initial model was a good fit to the data. The chi-square in the present model was 6.931, with 3 degrees of freedom and a probability of .074. Other measures of fit included RMSEA, CFI and SRMR. In this model, RMSEA was larger than .06, suggesting a reasonable fit. CFI value was .98 and is larger than the cutoff value of .95, indicating a good fit. The SRMR value was .085, suggesting a reasonable fit of the model to the data. By these criteria, it appeared that the one-factor model fit the data reasonably. Standardized residuals and modification indices revealed a localized point of ill fit in this solution (e.g., the largest modification index = 5.334). Figure 4.5 presents the standardized parameter estimates from this solution. Regarding the model itself, most subsets provided relatively strong measures of the ability except for Verbatim Source Use.

The individual parameter estimates from Content, Organization and Language Use to L2 Writing Ability were found to be statistically significant at .001 level, implying that the underlying factor, L2 Writing Ability, is well measured by the observed variables—Content, Organization, and Language Use—and that these variables are measuring L2 writing ability. Except for Verbatim Source Use, the factor loadings for the L2 Writing Ability were .77 or higher. The results suggest that Verbatim Source Use should be seen as a measure independent of the measures of Content, Organization and Language Use.

Comparing Competing Models

To test the hypothesis that the Verbatim Source Use scale is measuring something independent of Content, Organization and Language Use, Model 2.1 and Model 2.2 were compared using AIC and a chi-square difference test. Both test results indicate that Model 2.1 was preferred. The hypothesis that Verbatim Source Use should be treated as a variable independent of L2 Writing Ability was supported.

Table 4.14 Comparison of fit indexes for Model 2.1 and Model 2.2

Model	χ^2	<i>df</i>	$\Delta\chi^2$	Δdf	<i>p</i>	AIC	CFI	RMSEA (90% CI)
Model 2.1	1.407	2				17.407	1.00	.000 (.000–.617)
Model 2.2	6.931	3	5.52	1	.018	20.931	.98	.090 (.000–.174)

Discussion of quantitative results: Model 2.1

Model 2.1 supports one-factor and single-indicator measurement solution of the integrated writing performance. This solution confirms that integrated writing test performance consists of two components: L2 Writing Ability and Verbatim Source Use. The factor L2 Writing Ability is represented by test takers' scores on the Content, Organization, and Language Use scales. Verbatim Source Use is represented by test takers' scores on the Verbatim Source Use scale. The small but meaningful correlation

(.08) between the L2 Writing Ability factor and Verbatim Source Use suggests that the integrated writing test performance variables measured by the integrated reading-listening-writing test can be seen as two related but distinct measures.

The revised model 2.1 contains one factor (L2 Writing Ability) with three observed variables (Content, Organization, and Language Use) and an uncorrelated observed variable (Verbatim Source Use). As explained earlier in this chapter, Kline (2004, p. 71) warns that a hypothetical construct cannot be adequately measured by any single indicator for two reasons: (1) indicators are not perfectly reliable, and (2) their scores may not be perfectly valid. Because multiple indicators were not available for a hypothesized Verbatim Source Use factor, it was included in the model as a single-indicator variable.

The results were in line with the concept proposed by previous researchers (Cumming et al., 2005) that the extent of verbatim phrases in written responses relates not only to test takers' proficiency levels but many other factors such as medium of comprehension of stimulus materials, task conditions/characteristics, and memory factors. To delve into the intricacies of source use by test takers in the high- and low-performance groups, I pulled all segments related to source use from interview protocols and individuals' responses to the open-ended questions from the strategy inventory and analyzed them for patterns. The following section addresses the issue of source use from a qualitative perspective.

Test taker's discussion of source use

The qualitative data also revealed the dynamic interactions of test takers and their use of verbatim phrases. Generally speaking, it was found that there is a non-linear relationship between test takers' writing proficiency levels and their inappropriate use of source materials. The major differences in source use across the high- and low-

performance group depend on test takers' underlying rationales for using verbatim phrases in their writing.

High-performance Group

The test takers in the high-performance group reported they used verbatim phrases for three reasons. First, they mentioned using verbatim phrases when they needed to write some domain-specific or technical terms. One student said, "I wrote down some key phrases, something that I read or heard about. Those are really helpful, especially some technical terms like growth-inducing chemicals, something like that." Also, the reading-listening-writing task led these students to consider the expressions and usages from the sources as authoritative, and thus they thought these phrases should be presented in their writing exactly as they were in the reading passage. Lossie explained, "I copied and pasted some phrases because in testing situation like this, I wanted to make sure I used correct expressions from the passage. For creative writing task, I'll use my own words."

Finally, they wanted to use verbatim phrases from the reading and the lecture to improve their writing quality. They mentioned a pressing need to improve their writing by using a sequence of phrases from source materials even though they were very much aware that using verbatim phrases was inappropriate. Liu, for example, responded to the question about her use of verbatim phrases in the writing:

When I wrote ideas from the reading passage, I copied some phrases from that. He definitely wrote better than I did. I think it's not good, but I couldn't think of better expressions in a short time. I'll try to avoid that if I got a chance to redo my test.

Low-performance group

The test takers from the low-performance group showed two major reasons for copying verbatim phrases from sources. The issue of comprehension is one of them.

Some students found the reading passage or the lecture difficult to understand. They were likely to be stuck when they encountered some incomprehensible words or phrases. For example, MD mentioned, “I used some sentences from the reading and some words from the lecture. I just really try to understand what they are talking about.” Another interesting point brought up by a student from the low-performance group, Elle, was that she noted down phrases she did not understand from the listening passage and used those in her writing. Another reason for using verbatim phrases was to save time. In the interview, Tori said she used verbatim phrases “because I want to summarize the reading part and I think using the reading part is more convenient and easy way.”

Summary

In summary, the test takers used verbatim phrases in their writing regardless of their writing proficiency; however, they used the same strings of phrases from the sources for different purposes and reasons. For the test takers in the high-performance group, they borrowed string of words from sources because they wanted to improve their writing and to make sure they used the “correct” wording from the sources. The test takers in the low-performance group were likely to retain verbatim phrases when they had difficulties comprehending source texts. They also reported that pulling phrases directly from sources was easier than expressing ideas using their own words since they had limited time to complete the task.

Testing latent variable models

This section addresses the third research question “What is the relationship between strategy use and test performance on an integrated reading-listening-writing test?” Specifically, a number of full latent variable models were constructed to examine the relationship between strategy use and test performance.

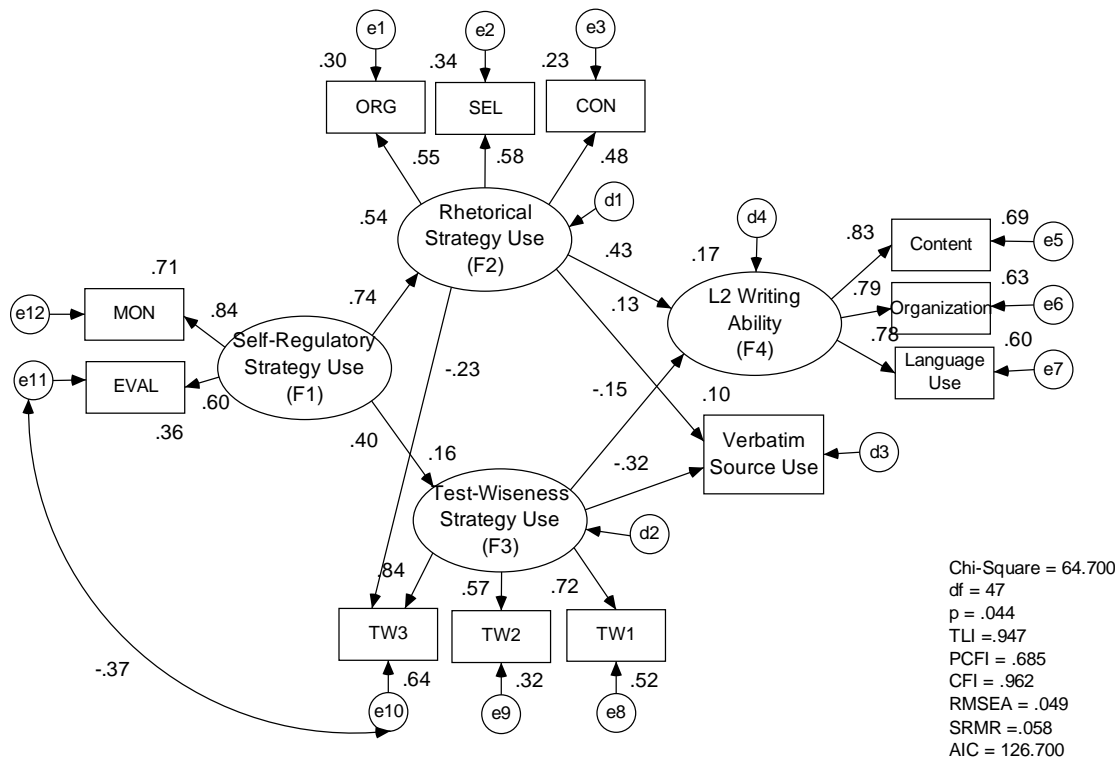
The hypothesized model: Model 3.1

Several full latent variable models of the relationship between strategy use and test performance were formulated and examined based on previous measurement models and literature on second language writing and integrated writing. Many of them were found to be misfitting or meaningless in terms of interpretation. However, one model which seemed to represent the sample data adequately from statistical and substantive viewpoints was Model 3.1.

Figure 4.6 presents a combination of two measurement models: Model 1.3 and Model 2.1. Integrated strategy use was depicted by Model 1.3, the three-factor model with three inter-correlated underlying factors and eight observed variables. Integrated writing test performance was characterized by Model 2.1, a one-factor and single-indicator measurement model with four observed variables.

Cohen (1998, 2006) has suggested that studies in test-taking strategies are useful in examining test validity. To test this hypothesis, Rhetorical Strategy Use and Test-Wisness Strategy Use were assumed to have a direct effect on L2 Writing Ability and Verbatim Source Use. The measurement model 1.3 also contained a correlation between the uniqueness associated with the Evaluating strategies and Item TW3. Moreover, Item TW3 was postulated to cross-load with Test-Wisness Strategy Use and Rhetorical Strategy Use. The model 3.1 is the initially-hypothesized full latent model. Figure 4.6 shows a schematic representation of Model 3.1.

Figure 4.6 Initially-Hypothesized model of the relationship between strategy use and test performance: Model 3.1



Note. SEL = Selecting, ORG = Organizing, CON = Connecting, Mon = Monitoring, EVAL = Evaluating, TW = Test-Wisness

Estimation results for Model 3.1

An examination of statistical assumptions underlying the maximum likelihood estimation procedures indicates the value for skewness and kurtosis for the individual variables were univariately normally distributed. As a result, the normality assumption was not violated. The overall model-fit statistics showed several indications of a misfitting model. Model 3.1 yielded a chi-square value of 64.700 with 47 degrees of freedom and a probability of .044, indicating that the actual and implied matrixes were statistically significantly different from one another. Therefore, the model and the data were not consistent with one another. To ensure model fit, other fit indices were also used

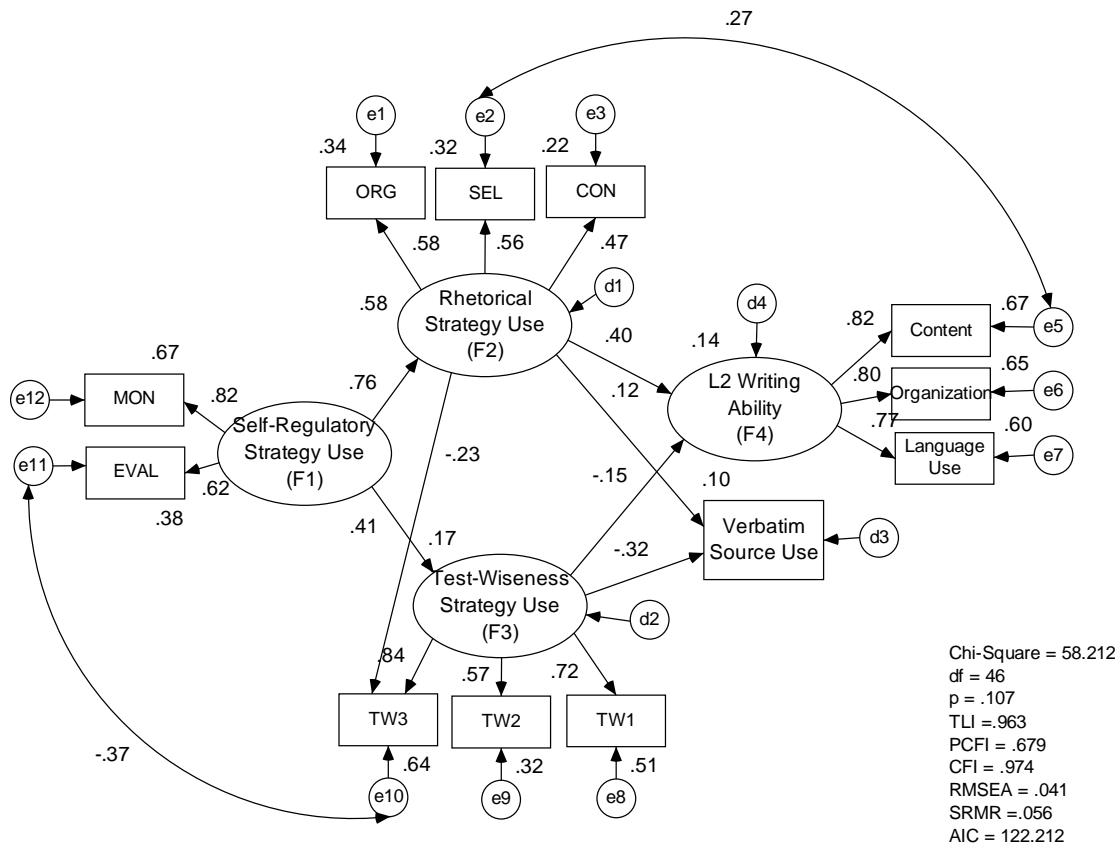
to examine this model. The TLI and CFI values fell around the borderline of .95. Following a more stringent rule of thumb, Model 3.1 showed an adequate fit but not a good fit to the sample data. The SRMR value was below .80 (.058) and the RMSEA value was .049, indicating that the model was an adequate fit to the data.

Some degree of misfit in this model was indicated in its standardized residuals and modification indices (e.g., largest standardized residual = -2.431, largest modification index = 10.066). Based on these results, this model was rejected. Some modifications were needed before conducting further analyses.

The hypothesized model: Model 3.2

Based on standardized residuals and modification indices, another full latent model was constructed to present the relationship between strategy use and test performance. In Model 3.2, Self-Regulatory Strategy Use was hypothesized to have a direct effect on Rhetorical Strategy Use and Test-Wiseness Strategy Use. Self-Regulatory Strategy Use was also hypothesized to have an indirect effect on L2 Writing Ability and Verbatim Source Use. Rhetorical Strategy Use was hypothesized to have a direct effect on L2 Writing Ability and Verbatim Source Use. Test-Wiseness Strategy Use was proposed to have a direct effect on both L2 Writing Ability and Verbatim Source Use. As suggested by the modification index, Model 3.2 also included one more error covariance between errors associated with the Selecting strategies and Content. Relaxing this path resulted in a decrease in χ^2 of at least 6.488. This particular change indicated that the Selecting strategies and test takers' score on Content scale shared additional variance owing to influences other than their respective latent constructs: Self-Regulatory Strategy Use and L2 Writing Ability. Figure 4.7 provides a schematic representation of the model of the relationship between strategy use and test performance.

Figure 4.7 Hypothesized model of the relationship between strategy use and test performance: Model 3.2



Note. SEL = Selecting, ORG = Organizing, CON = Connecting, Mon = Monitoring, EVAL = Evaluating, TW = Test-Wisness

Estimation results for Model 3.2

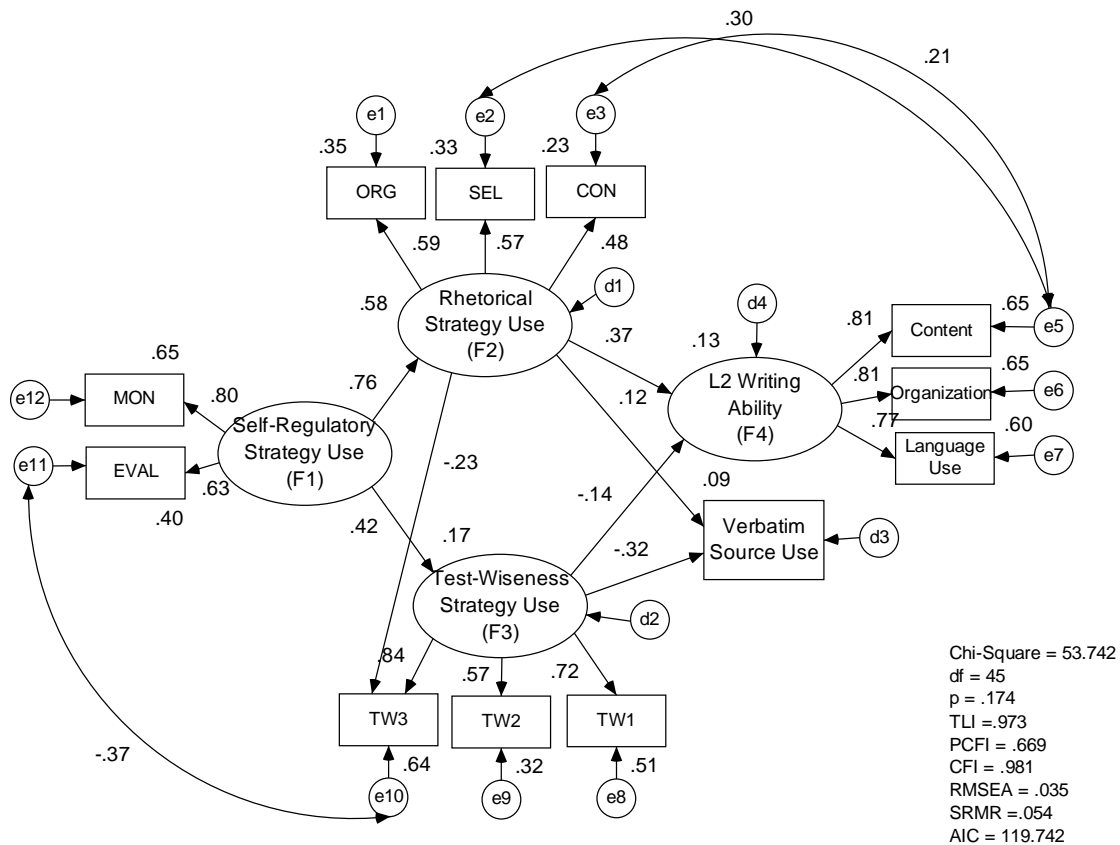
Several measures of fit suggested that Model 3.2 fit the data well. The chi-square in the present model is 58.212, with 46 degree of freedom and a probability of .107. The TLI and CFI were larger than .95, indicating a good fitting model. The SRMR value was .056, suggesting an adequate fit of the model to the data. Also, Model 3.2 produced a RMSEA value smaller than .05 (.041), suggesting a good model fit.

All freely estimated unstandardized parameters were statistically significant. Factor loading estimates revealed that all indicators were strongly related to their purposed latent factors. An inspection of standardized residuals and modification indices showed a small degree of misfit in this solution (e.g., largest standardized residual = -2.339, largest modification index = 8.69). Thus, some adjustments were made to improve the model fit before further analyses.

The hypothesized model: Model 3.3

Based on the estimation results for Model 3.2, one more constraint was relaxed by estimating the error covariance between the Connecting strategies and Content. Like in the Model 3.2, Self-Regulatory Strategy Use was hypothesized to have a direct impact on Rhetorical Strategy Use and Test-Wiseness Strategy Use, and an indirect impact on L2 Writing Ability and Verbatim Source Use. Figure 4.8 shows the diagrammatic representation of Model 3.3.

Figure 4.8 Hypothesized model of the relationship between strategy use and test performance: Model 3.3



Note. SEL = Selecting, ORG = Organizing, CON = Connecting, Mon = Monitoring, EVAL = Evaluating, TW = Test-Wiseness

Estimation results for Model 3.3

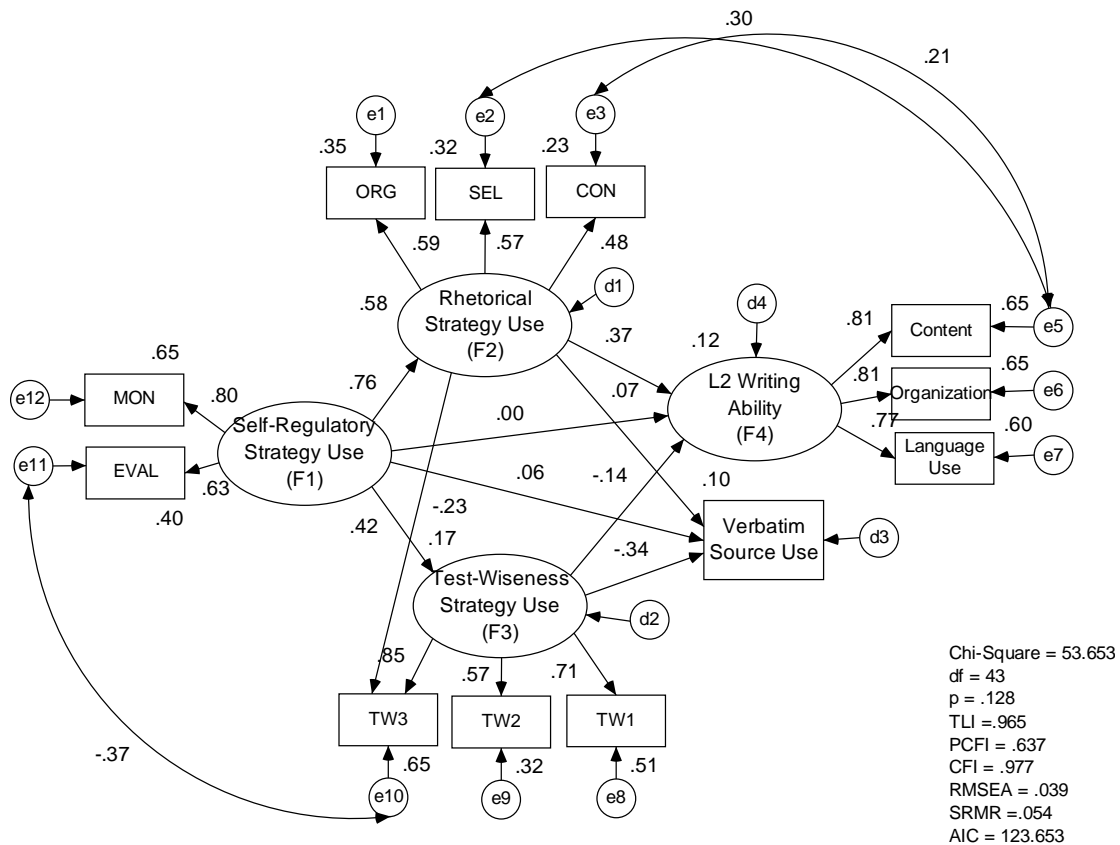
In terms of goodness-of-fit, Model 3.3 yielded a χ^2 statistic of 53.74 with 45 degrees of freedom and a probability of .174. Other overall goodness-of-fit indices suggested that Model 3.3 fit the data well (SRMR = .054, RMSEA = .035, TLI = .973, CFI = .981). The modification index suggested a linkage between the uniqueness associated with Item TW2 and with the Content scale. Because this relationship did not make theoretical sense, the suggestion was dropped. However, all factor loadings,

variances and covariances were found to be statistically significant except for the path between Rhetorical Strategy Use and the Verbatim Source Use variable and the path between Test-Wiseness Strategy Use and L2 Writing Ability. These paths were of interest for the current study, thus they were retained in the model for further analyses.

The hypothesized model: Model 3.4

In Model 3.3, Self-Regulatory Strategy Use was assumed to have no direct effect on test takers' L2 Writing Ability and Verbatim Source Use, but that its effect was indirect through Rhetorical Strategy Use and Test-Wiseness Strategy Use. One can also argue that Self-Regulatory Strategy Use may also have direct effect on L2 Writing Ability and Verbatim Source Use. Thus, whether or not this assumption is supported should be tested by freeing these parameters and studying the change in fit of the model. In Model 3.4, Self-Regulatory Strategy Use was hypothesized to have both direct and indirect effects on L2 Writing Ability and Verbatim Source Use. Rhetorical Strategy Use and Test-Wiseness Strategy Use were still hypothesized to have direct effects on L2 Writing Ability and Verbatim Source Use. Figure 4.9 shows the hypothesized Model 3.4.

Figure 4.9 Hypothesized model of the relationship between strategy use and test performance: Model 3.4



Note. SEL = Selecting, ORG = Organizing, CON = Connecting, Mon = Monitoring, EVAL = Evaluating, TW = Test-Wiseness

Estimation results for Model 3.4

The overall model-fit statistics Model 3.4 was a good fit to the data. This model produced a chi-square value of 53.653 with 43 degrees of freedom and a probability of .511. The TLI and CFI values were above .95. The SRMR value was below .80 (.054) and the RMSEA value was below the borderline value of .05 (.039).

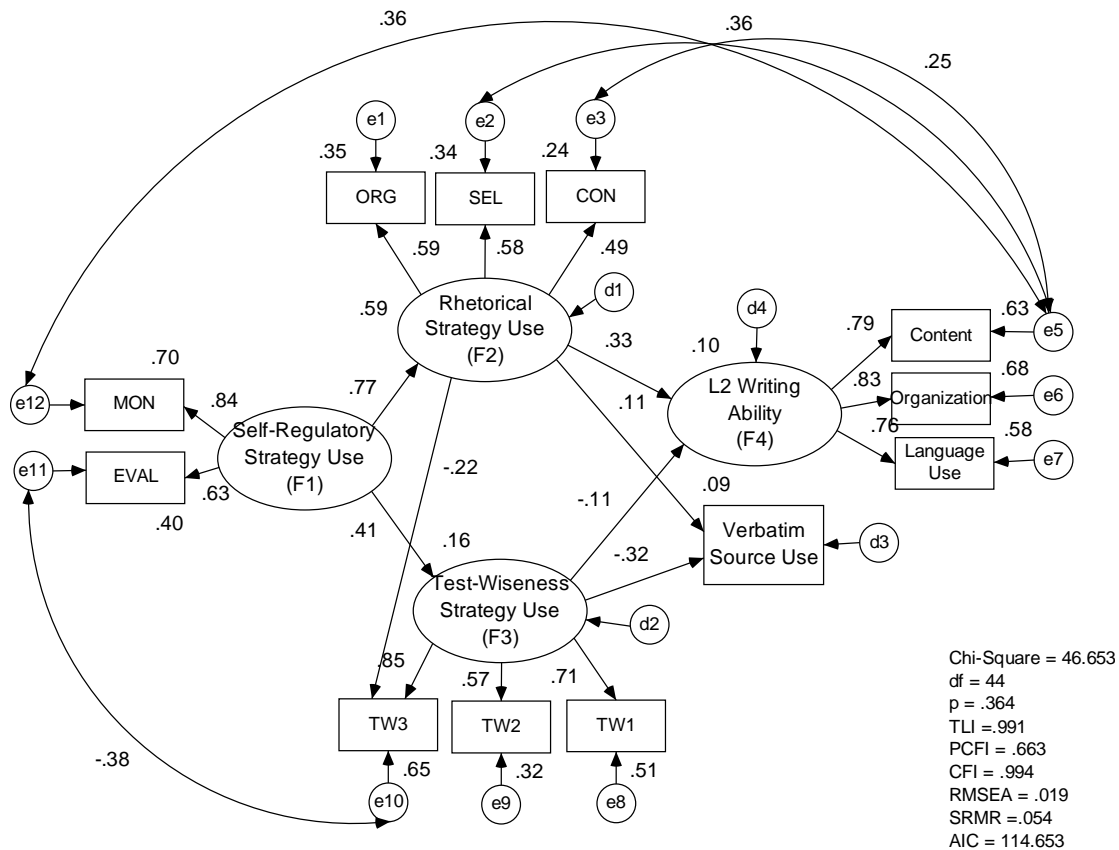
All factor loadings were significant at the level of .05 except for (1) the path from Self-Regulatory Strategy Use to L2 Writing Ability and Verbatim Source Use, (2) the

path from Test-Wiseness Strategy Use to L2 Writing Ability, and (3) the path from Rhetorical Strategy Use to Verbatim Source Use. Some degrees of misfit in this model were indicated in its standardized residual and modification indices (e.g., largest standardized residual = - 2.252, largest modification index = 8.544).

The hypothesized model: Model 3.5

In Model 3.4, the Self-Regulatory Strategy Use was hypothesized to have both direct and indirect effects on L2 Writing Ability. Previous research in writing strategies (Purpura, 1999; Song & Cheng, 2006) have found strategies test takers used to control their writing processes had indirect effects on their performance in language tests. However, to date, no studies were found to investigate how different strategy use directly or indirectly affected test takers' performance in an integrated writing test. To explore this relationship for the first time using a structural equation modeling approach, Model 3.5 was established. Self-Regulatory Strategy Use was hypothesized to have only indirect effect on test takers' L2 Writing Ability and Verbatim Source Use. Based on the modification index, one more path was relaxed between Monitoring strategies and Content. These relationships were presented in Model 3.5.

Figure 4.10 Hypothesized model of the relationship between strategy use and test performance: Model 3.5



Note. SEL = Selecting, ORG = Organizing, CON = Connecting, Mon = Monitoring, EVAL = Evaluating, TW = Test-Wisness

Estimation results for Model 3.5

The overall model-fit statistics Model 3.5 was an excellent fit to the data. This model produced a chi-square value of 46.653 with 44 degrees of freedom and a probability of .364. The TLI and CFI values were above .95. The SRMR value was below .80 (.054) and the RMSEA value was below the borderline value of .05 (.019). All factor loadings were significant at the level of .05 except for two paths: (a) the path from Rhetorical Strategy Use to Verbatim Source Use, and (b) the path from Test-Wisness

Strategy Use to L2 Writing Ability. Some degrees of misfit in this model were indicated in its standardized residual and modification indices (e.g., largest standardized residual = - 2.111, largest modification index = 5.306). However, the addition of these paths did not make substantive sense. Therefore, they were not added to the model.

Comparing Competing Models

Table 4.15 shows the comparison of fit indices for alternative models of the effects of strategy use on test performance. The model shown in Figure 4.7 (Model 3.2) is a more parsimonious, more constrained version of the model shown in Figure 4.8 (Model 3.3) and Figure 4.9 (Model 3.4). This model is nested within Model 3.3 and Model 3.4. A chi-square difference test was used to compare Model 3.2 and Model 3.3. The results indicated that the additional constraint on Model 3.2 resulted in a statistically significant increase in χ^2 , suggesting that not only did Model 3.2 fit worse than Model 3.3, but it fit statistically significantly worse. Because Model 3.3 was chosen, Model 3.4 was then compared to Model 3.3. Since the $\Delta \chi^2$ was not significant for Model 3.4 ($\chi^2_{\text{diff}}(3) = 4.56, p = .21$), the more parsimonious model, Model 3.3, was favored. Based on these results, Model 3.5 was compared to Model 3.3. The chi-square difference results showed that the parsimony came at too great a cost in terms of model fit, and thus the constraint (i.e., an error association between e12 and e5) on Model 3.3 was rejected. Thus, Model 3.5 was preferred. The AIC criterion also shows similar results. Smaller AIC values were indicative of a better fitting model. Considering all models had reasonable fit to the data, Model 3.5 (AIC = 114.653) was favored over other models.

Table 4.15 Comparison of fit indexes for Model 3.2—3.5

Model	χ^2	<i>df</i>	$\Delta\chi^2$	Δdf	<i>p</i>	AIC	CF I	RMSEA (90% CI)
Model 3.2	58.21	46				122.21	.97	.041 (.000–.667)
Model 3.3	53.74	45	1	4.47 ^a	.03	119.74	.98	.035 (.000–.758)
Model 3.4	53.65	43	3	4.56 ^b	.21	123.65	.98	.039 (.000–.684)
Model 3.5	46.65	44	2	11.56 ^c	<.01	114.65	.99	.019 (.000–.890)

Note. ^aCompared to Model 3.2. ^bCompared to Model 3.3. ^cCompared to Model 3.3.

DISCUSSION OF THE FULL LATENT MODEL 3.5

The full latent model presented in Figure 4.10 displays dynamic relationship between integrated writing strategy use and integrated writing test performance. In creating the full latent model, two measurement models (i.e., Model 1.3 and Model 2.1) were combined and the directional influences of one variable or factor upon another were added. In order to examine the interactions between integrated writing strategy use and test performance, this section first discusses two individual measurement models and then focuses on their relationships within the full latent model.

The integrated writing strategy use model

Based on previous literature in integrated writing (Asención, 2004; Esmaeili, 2002; Plakans, 2008; Spivey, 1997) and second language assessment (Cohen, 2002; Purpura, 1997; Tsai, 2004), the strategy inventory for integrated writing was originally designed to measure three underlying factors: Rhetorical Strategy Use, Self-Regulatory Strategy Use, and Test-Taking Strategy Use. Each was represented by a group of strategy types. The results of a series of EFAs and CFAs showed that the strategy items clustered into six strategy types that measured three underlying factors of integrated writing strategy use: Rhetorical Strategy Use, Self-Regulatory Strategy Use and Test-Wisness Strategy Use.

As seen in Figure 4.11, integrated writing strategy use was explained by these three strategy use factors within the full latent model 3.5. Rhetorical Strategy Use was moderately well-explained by the connecting strategies with a factor loading of .49 ($R^2 = .24$), by the organizing strategies with a loading of .59 ($R^2 = .35$), and by the selecting strategies with a loading of .58 ($R^2 = .34$). Self-Regulatory Strategy Use was well-explained by the monitoring strategies with a loading of .84 ($R^2 = .70$) and by the evaluating strategies with a loading of .63 ($R^2 = .40$). Test-Wiseness Strategy Use was explained by TW1 with a loading of .71 ($R^2 = .51$), by TW2 with a loading of .57 ($R^2 = .32$), and by TW3 with a loading of .85 ($R^2 = .65$). All path loadings were statistically significant at the .05 level.

The integrated writing strategy use model also yielded a strategy-type cross-loading. Item TW3 presented a significant positive loading (.85) on Test-Wiseness Strategy Use, and a significant negative loading (-.22) on Rhetorical Strategy Use. This result suggests that copy-and-revise strategy (TW3) should be viewed as a type of test-wiseness strategy (i.e., test takers' respond to test tasks using test format rather than invoking expected cognitive or linguistic processes), and this strategy use is detrimental to the use of rhetorical strategies such as the selecting and connecting strategies. This finding supports Campbell's (1990) depiction that students' inability to integrate sources into their writing plays a role in their decision to copy inappropriately.

The integrated writing strategy use model also presents a significant relationship between error terms associated with the evaluating strategies and TW3. From a substantive perspective, this relationship appears logical, as one Evaluating item measures test takers' strategy of not copying exact phrases and the TW3 measures their copy-and-revise behavior. The common cause of these variables relates to test takers' copying behaviors, which is not accounted for in the present model. It also appears that

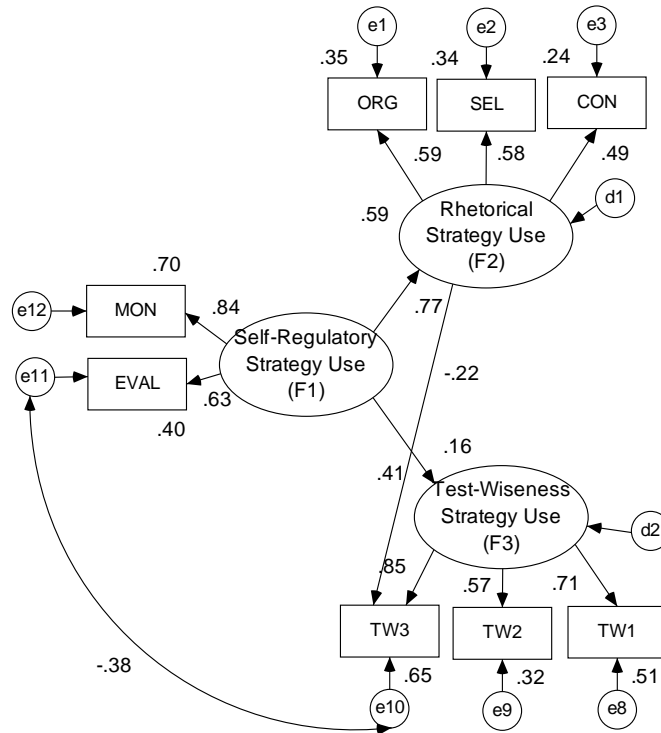
the correlated errors emerge from the items that are similarly worded. One advantage of SEM over other traditional methods such as regression analysis is that it accounts for measurement errors allowing “true” effect to be revealed. This information provides insight into item revisions and modifications.

In terms of the relationships among three strategy use factors, Self-Regulatory Strategy Use had a positive direct effect on Rhetorical Strategy Use ($\beta = .77$). It suggests that self-regulatory mechanisms serve an executive control over other cognitive resources. Only when these mechanisms are invoked can test takers set goals, manage time, process available information and finally complete the required task. Each type of strategy works in conjunction with various other types of strategies concurrently. One supports or complements the others through the process in completing the task within the stipulated time.

Self-Regulatory Strategy Use was also found to have a direct positive effect on Test-Wisness Strategy Use ($\beta = .41$). This positive relationship indicates that many students selectively copied some phrases from the reading passage based on the task requirements. An inspection of test takers essays showed that those exact phrases used by test takers were the key phrases in the reading passage, rather than mere random phrases.

In sum, these results show that Rhetorical Strategy Use, Self-Regulatory Strategy Use and Test-Wisness Strategy Use are closely related but play distinctive roles in the integrated writing test. Figure 4.11 presents the integrated writing strategy model drawn from the full latent Model 3.5.

Figure 4.11 Integrated writing strategy use model (from the full latent Model 3.5)



Note. SEL = Selecting, ORG = Organizing, CON = Connecting, Mon = Monitoring, EVAL = Evaluating, TW = Test-Wiseness

The integrated writing test performance model

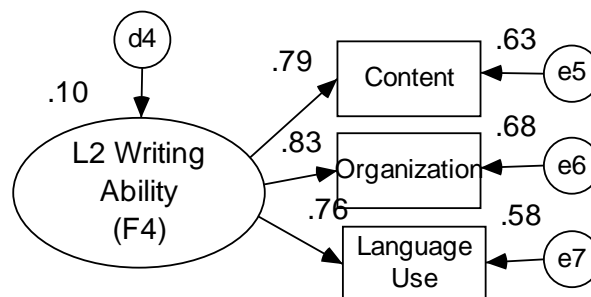
The reading-listening-writing test was originally designed to measure L2 writing ability. However, with the increasing interest in source use in L2 integrated writing tests (Cumming et al., 2005; Lee & Kantor, 2005), the integrated writing test performance model includes a measure of verbatim source use to explore the relationship between test takers' L2 writing ability and their use of verbatim phrases from source materials. Results of a series EFAs and CFAs revealed a low and insignificant relationship ($\beta = .08$) between L2 writing ability and test taker's verbatim source use. Therefore, in this study,

the integrated reading-listening-writing test measures integrated writing performance through the factor of L2 writing ability and the verbatim source use variable.

In terms of the L2 Writing Ability factor, all measures displayed a strong and statistically significant association with L2 Writing Ability. The L2 Writing Ability was well-measured by the test takers' performance on the Content scale with a loading of .79 ($R^2 = .63$), and by the Organization scale with a loading of .83 ($R^2 = .68$), and by the language Use scale with a loading of .76 ($R^2 = .58$). The results suggested that 58% or higher in test takers' performance on these scales are explained by their L2 Writing Ability.

In sum, these results show that the reading-listening-writing test measures test takers' L2 Writing Ability. Moreover, the Verbatim Source Use scale measures a different component of writing which is loosely related to test takers' L2 writing ability.

Figure 4.12 Integrated writing test performance model (from the full latent Model 3.3)



The relationship between integrated writing strategy use and test performance

The full latent model presented in Figure 4.13 shows the relationship between integrated writing strategy use and integrated writing test performance. Some interesting and unexpected findings are highlighted in the following section.

The Rhetorical Strategy Use, Self-Regulatory Strategy Use and Test-Wisness Strategy Use were originally hypothesized to have significant direct impact on L2

Writing Ability. However, only Rhetorical Strategy Use and Test-Wiseness Strategy Use had direct influence on L2 Writing Ability. Self-Regulatory Strategy showed no direct impact on integrated writing test performance, but had a significant indirect effect on L2 Writing Ability ($\beta = .25$) and Verbatim Source Use ($\beta = -.13$) via Rhetorical Strategy Use and Test-Wiseness Strategy Use respectively. Given the strong relationships among Self-Regulatory Strategy Use and the other two strategy use factors, it appeared that the self-regulating strategies served an executive function in managing the use of rhetorical strategies and test-wiseness strategies.

The tendency for the use of self-regulatory strategies to not occur by themselves solely but in combination with other strategies was also revealed in the qualitative data. A student from the high-performance group showed how she approached the writing task. Before she started reading the passage, she had a plan in mind about how she was going to do it. She said, "I first do the scanning and then go to the detail. It helps me get the organization of the listening and the reading." Similarly, a student from the low-performance group said,

When I was reading, I read all of them very fast and catch the main idea. But I think this way is not so suitable for me because sometimes if this topic is the architecture or something more difficult to me, like the topic this time I read all of them but maybe I didn't catch the point. Because I read so fast and I don't know some words.

The interview data revealed that although students from both the high- and low-performance group used similar strategies such as Monitoring, Selecting and Organizing, they seemed to use them differently. The student in the high-performance group used top-down reading strategies to respond to the reading passage. However, the student in the low-performance group wanted to use similar strategies but ended up focusing on word-level meaning rather than the global meaning of the passage. This finding is consistent with previous research that students with higher proficiency tend to use more global

strategies, while students with lower proficiency level tend to process information at the local level (Koda, 2005; Plakans, 2009)

In terms of the impact of Rhetorical Strategy Use and Test-Wisness Strategy Use on L2 Writing Ability, the model showed significant positive direct influence of the Rhetorical Strategy Use on test takers' L2 Writing Ability ($\beta = .46$). In addition to the impact of Rhetorical Strategy Use on L2 Writing Ability, the model pointed to a negative relationship between Test-Wisness Strategy Use and L2 Writing Ability ($\beta = -.11$). Although the impact is not significant, this suggests that test takers' use of test-wisness strategies had a negative influence on their test performance. In other words, the reading-listening-writing test calls for the use of academic writing skills, rather than test-wisness tricks. The more test takers rely on test-wisness strategies, such as patchwriting strategies or applying previously-memorized writing templates, the less likely they will receive a high score on their writing.

Rhetorical Strategy Use and Test-Wisness Strategy Use were found to have direct impact on test takers' level of Verbatim Source Use. Consistent with previous finding, Test-Wisness Strategy Use had a significant negative impact ($\beta = -.32$) on Verbatim Source Use. This means that the more frequently a test taker uses test-wisness strategies, the lower the score on Verbatim Source Use scale a test taker i. This appears logical since one would presume that test takers' textual borrowing practices are likely to result in instances of copied verbatim phrases in their writing. On the contrary, Rhetorical Strategy Use had a positive impact on Verbatim Source Use ($\beta = .11$), suggesting a test taker who used more rhetorical strategies tended to use source materials more appropriately.

The correlations between the uniqueness associated with the monitoring strategies and content features of student writing (.36), the connecting strategies and content (.25),

and the selecting strategies and content (.36) indicated that three strategy-type variables and Content variable were measuring something over and beyond their proposed latent factors. However, the interpretation of these correlated error relationships remains unclear.

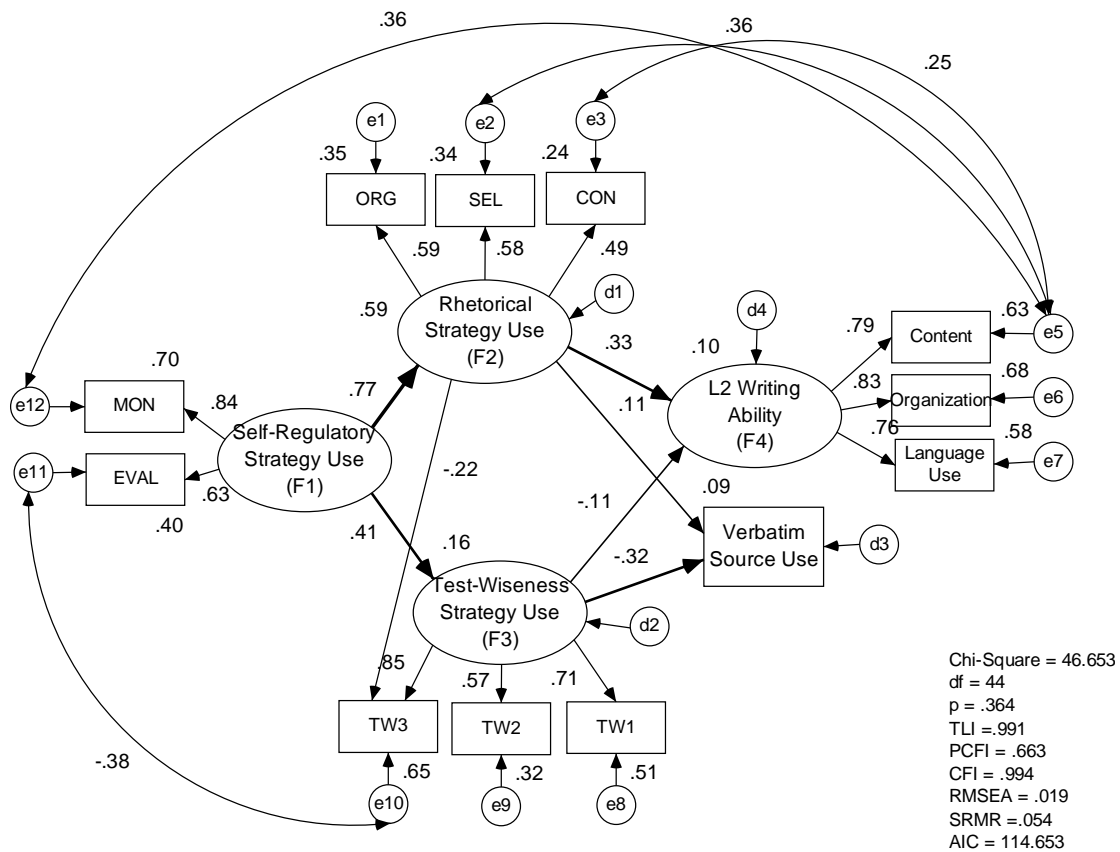
The fact that L2 Writing Ability had no significant direct impact on Verbatim Source Use was interesting since one would easily assume that test takers with higher L2 writing ability should know better about appropriate and inappropriate use of source materials. However, as supported by a number of research studies, plagiarism may be caused by several factors, such as a lack of awareness of authorial identity (Howard, 1995; Scollon, 1995), cultural differences (Sutherland-Smith, 2005; Pennycook, 1996), and task characteristics (Kirkland & Saunders, 1991). In this study, many students were not aware of what type of source use is inappropriate. One student in the high-performance group said, “When I was writing, I would go back to see what exact words or phrases the author used. Then I used those words in my writing.” Another student in the same group said, “I like to use the exact phrases from the reading because those words are more precise.” In sum, most students retained verbatim strings of words from sources with little intention of plagiarizing, but rather the imitation seems to be “a kind of apprenticeship during which they come to understand not only the appropriate rhetorical forms, but also the underlying issues that make writing interesting and arguments effective” (Applebee, 2000, p. 105).

These findings are interesting in two ways. First of all, they enable us to look at the nature of an individual’s strategy use when responding to an integrated writing test. The model revealed that test takers engaged themselves in three sets of strategies: the self-regulatory strategies, the rhetorical strategies, and test-wisness strategies during the test. Each played a different role during the writing task. The self-regulatory strategies

were initiated to regulate techniques for input comprehension and for output production. In the current study, input comprehension refers particularly to the discourse comprehension at varying levels such as an understanding of the word or phrases level or overall ideas of the passage or the lecture. The test-wiseness strategies were often utilized when test takers encountered comprehension issues or when they felt the need to polish their writing by borrowing phrases from the source materials. The model indicates the interrelated nature of these different strategy uses and how they intricately correspond with one another based on the immediate environmental needs, such as task requirements or personal goals.

In addition, the impact of an individual's strategy use on test performance found in this study suggests a dynamic and purposeful nature of strategy use on the integrated writing test. Each set of strategies may either compliment each other or work against each other. For instance, in responding to the integrated writing test, the test takers may activate the self-regulatory strategies to draw from a repertoire of reading and listening strategies (e.g., selecting important information in the reading or the lecture, organizing information available, and connecting information from the reading passage and the lecture). These strategies are viewed as facilitative to their integrated writing test performance. However, they may also apply a series of test-wiseness strategies (e.g., copying sentences from the reading passage and then revising them). These strategies are detrimental to their integrated writing test performance. Given the interactive nature of the use of different strategies and their impact on writing test performance, the very act of strategy use should be seen as a range of choices test takers make to fulfill the requirements of the target task.

Figure 4.13 Model of the relationship between strategy use and test performance



Note. SEL = Selecting, ORG = Organizing, CON = Connecting, Mon = Monitoring, EVAL = Evaluating, TW = Test-Wisness

Summary

This chapter presents the results of EFAs and CFAs. Based on the composite variables generated through a series of EFAs, a series of CFAs were performed to establish two measurement models that explore the relationships among factors. Then, several hypothesized structural models, also known as full latent models, were proposed and tested to establish a substantively- and statistically-sound model that fitted the sample data. The results of these analyses affirmed the multidimensional nature of integrated writing strategy use and integrated test performance. Integrated writing strategy use was

explained by Rhetorical Strategy Use, Self-Regulatory Strategy Use, and Test-Wiseness Strategy Use. Integrated writing test performance was explained by L2 Writing Ability and Verbatim Source Use. Self-Regulatory Strategy Use was found to have no direct impact on L2 Writing Ability or Verbatim Source Use, but to have indirect impact on both of them through Rhetorical Strategy Use and Test-Wiseness Strategy Use. Both Rhetorical Strategy Use and Test-Wiseness Strategy Use had direct impact on L2 Writing Ability and Verbatim Source Use.

Qualitative data including interview protocols, test takers' planning sheets, and answers to an open-ended questionnaire about strategy use were collected to triangulate quantitative results and provide further information about the relationship between strategy use and test performance. These data were read for themes, which evolved into codes for analysis. The qualitative results show that test takers applied different strategies at different composing phases. Test takers in the high- and low-performance groups used similar types of strategies while the frequency and quality of strategy use led to variance in written products.

Chapter 5

DISCUSSION

The goal of this study is to examine test takers' reported integrated writing strategy use, their test performance on an integrated writing task, and the relationship between the two. The results of the study support the original hypothesis that variability in integrated writing test performance is related to test takers' strategy use including their reported rhetorical strategy use and test-wiseness strategy use. The study affirms the complex and multifaceted nature of integrated writing strategies and test performance on the integrated writing task. This chapter first summarizes the study results as they relate to the individual research questions proposed in Chapter Two and the rationale of this study set out in Chapter One. This chapter also presents (1) the theoretical, methodological, and pedagogical implications of the study for second language learning and testing research, (2) the limitations of the present study, and (3) directions for future research.

THE RESEARCH QUESTIONS

Research Question 1

What is the nature of strategy use in an integrated reading-listening-writing test?

In response to this question, a series of CFAs were conducted on the integrated writing strategy use variables. A model with three intercorrelated factors, eight observed variables, one correlated error and one cross-loading offers a reasonable interpretation of the correlations among observed variables. These results support the hypothesis that integrated writing strategy use is a multifaceted construct consisting of three factors: Self-Regulatory Strategy Use, Rhetorical Strategy Use and Test-Wiseness Strategy Use. Each

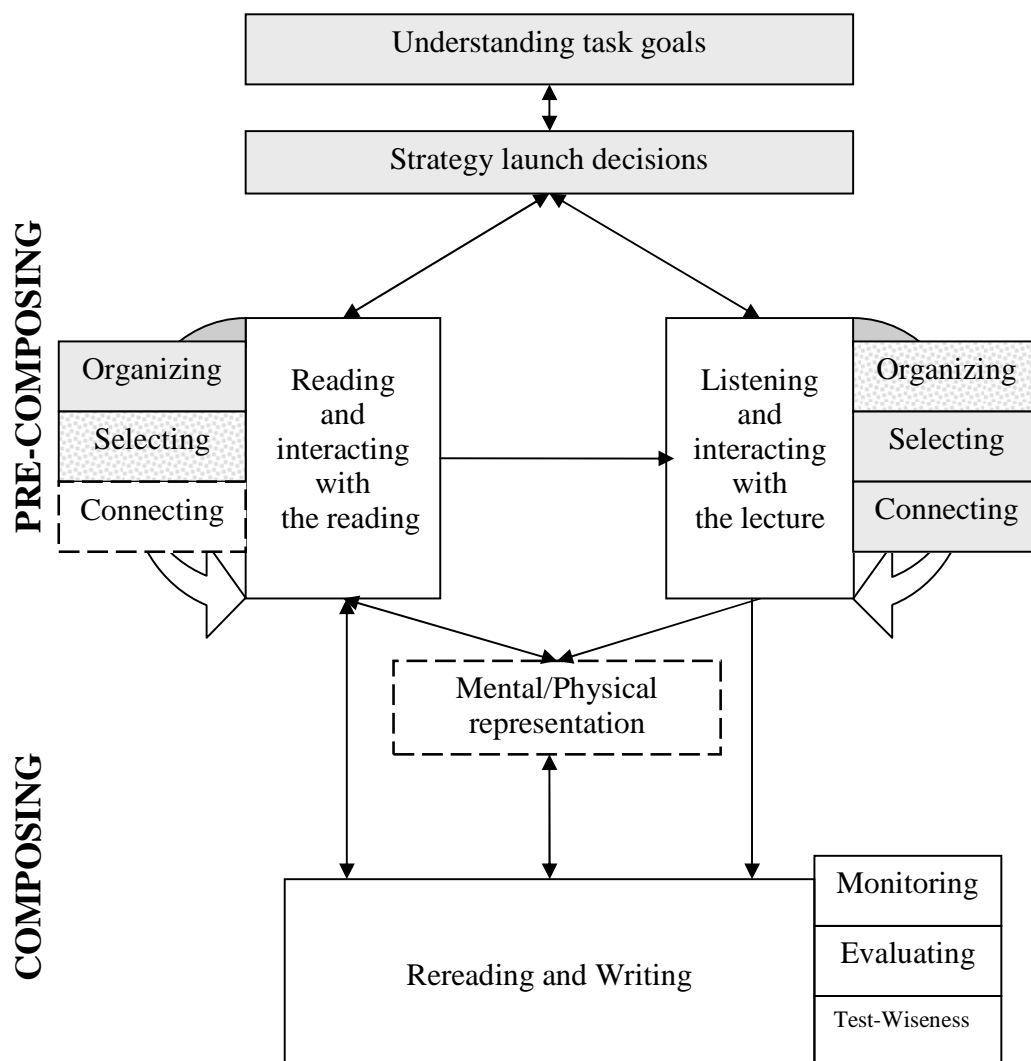
factor was presented by a number of strategy types (e.g., Connecting and Monitoring) or strategy items (e.g., TW1 and TW2). Each factor was also interrelated with other factors. As a result, integrated writing strategy use is seen to be a set of highly complex cognitive and metacognitive processes that are largely related to one another, suggesting that each kind of strategy use has considerable influence on the other.

In addition, the relationships among Rhetorical Strategy Use, Self-Regulatory Strategy Use and Test-Wiseness Strategy Use support a widely accepted notion that cognition is monitored and managed through the use of self-regulating strategies (O'Malley & Chamot, 1990; Wenden, 1991). It also suggests that test takers seem to invoke self-regulating mechanisms in their use of time and cognitive resources to complete a task. For instance, because of time and task constraints, test takers may need to read strategically only selected information from the reading passage, write down only keywords from the lecture, and then work through the task by combining both pieces of information logically. The interactive nature of the integrated writing task requires test takers to use a combination of these strategies simultaneously so that the task can be completed within the time limit.

Based on the interview protocols, a working model of the integrated writing strategies used for completing the reading-listening-writing task was created (Figure 5.1) in order to better understand the general patterns of strategy use across test takers from the high- and low-performance groups. The model presents a view of different integrated writing strategies used in the pre-composing preparation phase and its subphases as well as in the composing phase. It is a tentative model and is not designed to replace the existing integrated writing models (e.g., Flower's *Key elements in discourse construction*; Plakans' *Composing process model for reading-to-write test tasks*) or writing models (e.g., Flower and Hayes's *Cognitive process model of the composing process*). Instead, it

provides more information about the roles of strategy use in integrated reading-listening-writing tasks. As suggested by Weigle (2002), second language writing models address various factors that influence writing and should be context-sensitive. This model is a data and context-specific schematic representation of test takers' reported strategy use when responding to the reading-listening-writing task.

Figure 5.1 Composing strategies for a reading-listening-writing task



This overall pattern of strategy use reoccurred with some individual differences. In general, two phases occurred for all test takers: (a) a pre-composing phase and (b) a composing phase. The pre-composing phase consists of three subphases: preparation, reading, and listening. The composing phase consists of rereading and writing subphases. In this model, the double-headed arrows denote possible interactions within or between two phases. The patterns presented in grey squares occurred in higher frequency in the high-performance group than in the low-performance group. The patterns presented in dotted background squares occurred in higher frequency for the low performing students than for the high performing students. The dashed squares present different patterns for the high and low performing students.

In the pre-composing phase, all test takers reported reading first the task directions, passage, and then listening to the lecture. In the preparation subphase, most of the test takers in the high-performance group reported trying to understand the task goals and purposes by writing down or rereading the task directions. Many of them mentioned having a writing plan in mind before writing. Only a few participants in the low-performance group reported these behaviors.

In the reading subphase, high performing students reported using the organizing strategies as indices or guidelines in their search for major ideas. Low performing students, on the other hand, mentioned selecting key points as they read the passage from the beginning to the end. Also, unknown words seem to hamper reading comprehension for the low performing students more so than for the high performing students. The results show that the high performing students engage in more global reading strategies while the low performing students focus on local reading strategies. In terms of the connecting strategies, although students in both the high and low-performance group mentioned that they tried to connect ideas in the passage and the lecture by taking notes,

the quality of notes varied. Some of them noted down only relevant key words or phrases from the passage while others wrote down whole sentences in their planning page. Various students in the low-performance group reported writing down words or phrases with which they were not familiar.

After the test takers read the passage, they proceeded to listen to the lecture. Because many test takers in the low-performance group struggled with listening comprehension, they said they had difficulties applying the selecting and connecting strategies. However, several students mentioned trying to figure out the main ideas from the lecture by paying attention to the phrases that came after transitional words. Because many of them did not comprehend the lecture, they reported writing about their ideas or personal experiences in their essays. This is similar to the connecting strategies defined by Spivey (1990) (i.e., writers' incorporation of their own knowledge or experience about the topic in their writing), which were not related to the requirements of the current integrated task. Therefore, these students ended up receiving lower essay scores. In contrast, the test takers in the high-performance group were able to comprehend the lecture with little difficulties, thus they did not use the organizing strategies to help them improve their comprehension as often. They were also able to effectively extract important ideas from the lecture and connect those ideas with concepts in the reading passage.

During the listening subphase, some test takers mentioned that they took notes to make associations between sources, while others formulated a mental representation of two strands of ideas in their minds. The notes or mental representations served as a third text test takers referred to when composing their essays. Because high performing test takers have better listening comprehension, they have a clearer picture about arguments from

two sources. This pre-composing phase was mostly linear, and test takers moved from one step to another in generally the same order.

In the composing phase, some test takers reported rereading the source passage, their notes or reflecting on their ideas before composing. As they wrote, they simultaneously monitored their writing progress, evaluated their writing, and figured out effective ways to improve their writing. To do so, they remembered rereading the passage, their notes, reflecting on some ideas in the lecture, or using test-wiseness strategies to manage the task. This second phase was fairly recursive and nonlinear, and test takers moved back and forth to coordinate different strategies for better products.

Research Question 2

What is the nature of test performance as measured by an integrated reading-listening-writing test?

To address this question, a series of CFAs were performed on the integrated writing test performance variables. A one-factor model and a single-indicator of integrated writing test performance provide a reasonable explanation of the relationships among the observed variables. The integrated writing test performance consisted of a factor, L2 Writing Ability, and an observed variable, Verbatim Source Use. Based on the full latent model 3.5, the variance in L2 Writing Ability was well explained by three observed scoring scale variables: Content, Organization, and Language Use. The Verbatim Source Use variable was presented as a single-indicator observed variable independent of the L2 Writing Ability factor. A fairly low correlation was found between the L2 Writing Ability factor and the Verbatim Source Use variable, suggesting that Content, Organization, and Language Use were related to Verbatim Source Use measure at a very low degree.

The interview data protocols also support this quantitative result. The test takers in both the high-performance and low-performance groups reported using verbatim phrases from the reading passage and lecture although they used them for different purposes and reasons. The high performing students reported using verbatim replications of source sentences because they wanted to write better essays. The low performing students, on the other hand, believed retaining verbatim phrases from source texts to be a result of their lack of L2 reading and listening comprehension.

Research Question 3

What is the relationship between strategy use and test performance on an integrated reading-listening-writing test?

In response to research question 3, a number of analyses were performed to construct full latent models plausible for examining the relationship between strategy use and test performance. The analyses follow the procedures proposed by Jöreskog (1993). The strategy use and test performance measurement models were inspected first, followed by the structural models representing their relationships.

The structure of integrated writing strategy use

Within the full latent model, integrated writing strategy use was again explained by three factors: Self-Regulatory Strategy Use, Rhetorical Strategy Use, and Test-Wisness Strategy Use. Each strategy factor was explained by two to three strategy-type variables or strategy items. The self-regulatory strategies were found to have a direct impact on the rhetorical strategies and the test-wisness strategies, and to have an indirect impact on L2 writing ability and verbatim source use via the rhetorical strategies and test-wisness strategies. The interview protocols also revealed that the self-regulatory strategies were used at the same time as and along with rhetorical and test-wisness strategies. These findings support the work of Purpura (1997) and Tsai (2004) in their

claims that metacognition regulates a repertoire of several cognitive behaviors. This finding also complements Bachman and Palmer's (1996) global view that strategic competence is composed of mainly metacognitive operations. This model helps explain how test takers take initiative to deploy strategies in sequences and in clusters, not isolation, to meet the requirements of an integrated writing task.

The model also produced one significant cross-loading: TW3 (i.e., a patchwriting strategy) that was negatively associated with Rhetorical Strategy Use and positively associated with Test-Wiseness Strategy Use. Similar results were revealed in the interview data. This suggests that students who reported using the rhetorical strategies tended to avoid using the patchwriting strategy and vice versa.

The structure of integrated writing test performance

Within the full latent model, integrated writing test performance was explained by the L2 Writing Ability and the Verbatim Source Use variables. L2 Writing Ability was measured by three scales in the analytic scoring rubric: Content, Organization, and Language Use. For the Verbatim Source Use variable, test takers who avoided verbatim source use received higher scores. As observed in the previous CFAs, this model showed that L2 Writing Ability and test takers' Verbatim Source Use were not significantly correlated.

Further inspection of the model and interview protocols indicated that the integrated reading-listening-writing task seemed to require both local and global processing of the reading passage and the lecture. Test takers had to process input at the lexical and syntactical level for comprehension, and then extract thematic and semantic units from loads of source information. Finally, test takers were required to compare and contrast source information and avoid using verbatim strings from source materials by expressing ideas in their own words.

The relationship between strategy use and test performance

The full latent model was constructed to explore the relationship between strategy use and test performance. It was originally hypothesized that all three strategy factors would have a direct impact on integrated writing test performance. However, this relationship was not found in the data. Rather, the model indicated that Self-Regulatory Strategy Use had a significant, direct positive impact on Rhetorical Strategy Use and Test-Wiseness Strategy Use. Moreover, Self-Regulatory Strategy Use had an indirect positive impact on test performance via Rhetorical Strategy Use and an indirect negative impact on test performance via Test-Wiseness Strategy Use. The interview protocols also show that both high and low performing students remembered modifying their strategies as they went from reading, listening to writing based on their awareness of task goals and effectiveness. These results suggest that the self-regulatory strategies serve an executive function over other strategies. Whether Self-Regulatory Strategy Use improves or deteriorates integrated writing performance depends on the type of strategies test takers consciously decide to apply.

Rhetorical Strategy Use was found to have positive direct impact on L2 Writing Ability and Verbatim Source Use. This result suggests that test takers' use of selecting, connecting and organizing strategies helped them improve their overall integrated writing test performance including content, organization and language use aspects of their writing, and helped them lower their use of verbatim phrases from source texts. This result provides empirical evidence in support of the concept put forth by a number of researchers (Mateos, Villalon, de Dios, & Martin, 2007; Spivey, 1997) that writing from sources requires a higher degree of knowledge processing that includes selecting, organizing and connecting.

The interview protocols provide more information about how the test takers in the high- and low-performance groups used different strategies to approach the task. All test takers reported the use of selecting and organizing strategies during reading and listening. However, the high performing students mentioned using more organizing and connecting strategies such as predicting the content of the lecture and taking notes on main ideas as they read. In terms of listening to the lecture, the low performing students' reported use of the selecting and connecting strategies were impeded by their lack of threshold listening comprehension ability.

Test-Wiseness Strategy Use was shown to have a significant negative effect on both L2 Writing Ability and Verbatim Source Use. This finding indicates that the more students report their use of the test-wiseness strategies in a writing test situation, the worse they performed. In other words, behaviors such as copying, patchwriting, and applying previously-memorized templates to writing worsened test takers' integrated writing performance in all aspects. These results became even more negative if a test taker was pressured by the time constraints inherent to the test or was unaware of the appropriate use of source information. As indicated in the interviews, many test takers in the high-performance group said they were not able to come up with good expressions to use in their essays within such a short time. While experiencing the same problems, test takers in the low-performance group tended to copy phrase from the source texts, and did not see this as inappropriate.

In addition to major relationships between strategy use and test performance, the qualitative results suggest that the use of the selecting and connecting strategies were facilitative in terms of improving the content feature in writing. The interview protocols and the planning sheets of the high performing test takers show that many main points selected from the reading and the lecture were used to enrich their essay contents. Put

differently, students who extracted key ideas or arguments from the reading passage and the lecture and then made systematic connections between two strands of ideas had improved content quality in their written responses.

With regard to the linkage between the uniqueness associated with the evaluating strategies and TW3, the more the students consciously reconsidered task goals, planned thoughts, and the quality of written texts, the less likely they were to undertake the TW3 strategy. Similar results were found in the interviews protocols. Test takers who reported being aware of appropriate use of source materials were less likely to use the TW3 strategy.

IMPLICATIONS

The results show that an integrated reading-listening-writing task requires more than comprehension and production abilities. Coordination among reading, listening and writing abilities is essential for examinees to complete the task successfully. The current study has theoretical, methodological and pedagogical implications for second language writing and language testing research.

Theoretical implications

With regard to theoretical implications, this study drew on research in integrated writing, second language writing, and test-taking strategy research to explore the construct of academic writing. By examining the relationship between selected integrated writing strategy use and test performance on a reading-listening-writing task through quantitative and qualitative lenses, this study proposes a model that identifies close linkages between process and product and a model that addresses connections among reading, listening and writing.

The study provides evidence related to the substantive aspects of construct validity (Messick, 1989) because the reported strategies reflect the fact that test takers

were engaged with the writing test task in the manner desired by test developers. The frequent use of the rhetorical strategies shows that test takers were actively interacting with the source texts by identifying, extracting and connecting information in order to complete the task. The positive effect of these rhetorical strategies and negative effect of the test-wisness strategies on test performance not only help confirm the validity of the construct of an integrated reading-listening-writing test but also to refute a concern that test takers might receive a higher score on this test by relying on construct-irrelevant strategies such as test-wisness tricks.

Based on the qualitative results, the working model (Figure 5.1) for integrated writing strategy use was constructed to address the nature of the academic writing tasks required to make connections among reading, listening and writing. In previous research studies by Plakans (2008), a process model that includes reading and writing presents how writers interchangeably move between source texts and their writing. This model provides a detailed comparison of the process involved in independent writing and reading-to-write test tasks. However, little, if any, research has proposed a writing strategy model that demonstrates interrelationships among all three language modalities: reading, listening and writing. The model developed in the present study is useful for examining the tri-directional interactions among reading, listening and test takers' written texts, but further refinement is needed for this model to present a construct for second language reading-listening-to-write.

This study also found that individual test taker's strategy use differed across the high- and low-performance group in terms of the choices and the effectiveness of strategy use. In the reading phase, all test takers applied the selecting and organizing strategies extensively. However, in the listening phase, when the lecture became a source for reference, many low performing students who suffered from a lack of listening

comprehension found it difficult to apply any rhetorical strategies. This information-deprived situation directly impeded their writing performance. These results show that listening comprehension plays a crucial role in distinguishing high and low performing students. More importantly, individual test takers need to have some threshold L2 listening comprehension ability in order to apply the rhetorical strategies effectively in this phase.

Another possible interpretation of test takers' low performance is that low performance may be due to ineffective coordinating the use of different strategies. Two test takers from the low-performance group, Kim and Sara, could understand most of the lecture but still had some difficulties handling this type of task. The interview results suggest that the integrated writing task required test takers to not only have an ability to comprehend source information presented in different modes but to sort out a collection of messages and present this in an organized manner in a limited time frame.

The study also found that test takers' L2 writing ability was not proportionate to their ability to use sources appropriately. Test takers who received high scores on content, organization, and language use features of the writing could receive low scores on verbatim source use, and vice versa. As pointed out by research on plagiarism (Howard, 1995; Kirkland & Saunders, 1991; Pennycook, 1996), inappropriate use of source materials may result from various factors, including task characteristics, cultural differences, personal beliefs and epistemology. In sum, a model of second language writing may need to consider these issues in order to gain further understanding of the relationship between source use and proficiency.

Methodological implications

Most integrated writing research explores the processes or strategy use using descriptive statistics and qualitative approaches. The present study examined the

underlying construct of integrated writing from a different perspective. The study used a structural equation modeling approach as the major analytic tool which allowed measurement errors to be taken into account in parameter estimations. In the practice of many testing researchers (e.g., Bachman, 2002; Kunnan, 1998), the use of SEM is encouraged to investigate the construct validity issue in language assessment.

Also, to explore the types of strategies used to respond to an integrated reading-listening-writing task, the strategy inventory for integrated writing was developed. After a series of EFAs and CFAs, the final inventory contains several strategy types: organizing, selecting, connecting, monitoring, evaluating and test-wiseness. The present investigation touches on several core strategy types that were employed by many test takers. However, the strategy types contained in the inventory are still far from comprehensive, and more study on integrated writing strategy use would be encouraged to complement the current inventory.

Pedagogical implications

The findings give rise to several pedagogical implications for strategy instruction related to integrated reading-listening-writing tasks. In ESL or EFL classrooms, four English language skills (listening, speaking, reading, and writing) are usually taught independently of each other and separately from actual language use. However, since the current study shows that the construct of academic writing involves multiple and complex literacy abilities, teachers should begin to consider teaching reading, listening and writing together rather than separately. Incorporating authentic integrated reading-listening-writing tasks in academic writing classes is the first step to raise recognition of the roles of reading and listening as input in writing.

With regard to the instruction provided in the coursework and assignments, this study offers some insights into the types of strategies reportedly used to perform well on

integrated writing test tasks. Based on the resulting model, the rhetorical strategies improve writing performance. The qualitative results show that the organizing strategies are extensively used by high performing students during reading. These strategies contribute to better reading comprehension and promote better retention of information for writing. Global reading strategies like organizing strategies are therefore more effective strategies and need to be instructed in class. To teach students use writerly reading strategies, teachers should help students establish a purpose for reading (Alexander, Graham, & Harris, 1998; Koda, 2005), such as considering text structure, argument patterns, and links among information units in order to more effectively grasp major concepts for later use.

This study also shows that high performing students took notes while listening to the lecture. These notes serve as an important text for reviewing for source content and distilling major concepts. In academic settings, students are usually required to understand class lectures and be able to apply the knowledge flexibly afterwards. To prepare students for the source-based writing expected in English-medium institutions, teachers should provide ample sample lectures for students to practice systematic note-taking strategies, such as concentrating on the main points and tolerating less than word-by-word comprehension, using abbreviations or symbols, using outline forms or numbering systems, recalling the content from notes, and integrate key concepts in writing. It is essential for students to not only comprehend the lecture but to be able to retain important concepts and data to develop effective writing.

As pointed out by Cohen (1998), strategies are teachable and they promote learner autonomy in language learning and use. Therefore, it is important to guide students through an interactive process of reading and listening for writing. Teachers should provide class activities such as reading and listening for clues based on text organization,

selecting key information from sources, and synthesizing ideas from multiple sources for writing. Students should also be explicitly instructed to use source information appropriately. Moreover, students need to be aware of their goals for writing and be able to apply strategies flexibly as they see fit.

In addition, the complexity of integrated writing tasks requires students to have the “strategic competence” (Canale & Swain, 1980) to handle multifaceted problems. To improve students’ integrated writing performance, teachers should raise learners’ awareness of strategy use, provide a set of strategies that can be used for coping with language tasks requiring several skills, and offer individualized strategy instruction appropriate for learners of different proficiency levels.

LIMITATIONS

Several limitations need to be acknowledged in the interpretation of these results. First of all, this research involved a simulated test, rather than a real testing condition, thus students may have different motivation and exhibit different behaviors from those experienced during an actual administration of the TOEFL. Moreover, this research is limited to one task from a TOEFL iBT data set. Students might demonstrate a varied range or different types of strategies when responding to different tasks in terms of source texts (e.g., the modality, number, topics, lengths, rhetorical organization, and linguistic features of source texts) and the extent to which source information are required to be integrated into student essays.

Another limitation relates to that the reliability of the selecting, connecting, and evaluating scales. These scales were moderately reliable (around .60), but not highly reliable compared to the rest of the scales (around .80) in the strategy inventory. The findings about test takers’ strategy use may be different when using inventories with higher reliability.

A further limitation in the research is that the study focuses on only a fairly small number of potential test takers so that parameter estimates in the model may be less stable. Also, as no interview protocols were analyzed for intermediate performing students, there is no evidence on the strategies this group of students engages with when completing an integrated reading-listening-writing task. Therefore, generalization of claims about L2 integrated reading-listening-writing test performance cannot be extended to a full range of test takers who take this type of test and to all performance groups.

Finally, the data in this study were collected primarily through self-reporting using questionnaires, a method with recognized limitation including a possible lack of internal validity. Questionnaire methods allow specific and large amount of data to be gathered in a reasonably economical fashion, but are by no means a comprehensive representation of test takers' mental processes.

FUTURE RESEARCH

The findings show that integrated writing tasks require complex cognitive and linguistic abilities for comprehension and production as well as systematic coordination among reading, listening and writing strategies. This supports their use as a measure of multiliteracy skills. However, the mutually interdependent structure of the measure (e.g., production is dependent on comprehension) has raised many questions. More work is warranted to further examine these construct-related issues.

The resulting model outlines the relationship between strategy use and integrated writing test performance. In order to understand this relationship further, more research needs to study the processes in which test takers engage to complete integrated tasks. To this end, more process research could usefully be undertaken (e.g., using think-aloud protocols or computer keystroke tracing logs) to describe types of strategies test takers at each performance level use to comprehend different types of source materials, extract

information from these sources, make connection between two sources, incorporate source information in the writing, and eventually present an effective essay. The analysis of process-tracing data will improve our understanding of how test takers' composing processes influence their discourse features and writing performance. This information could also be useful in refining scoring rubrics for integrated writing tasks and in developing automated writing scoring.

In this study, test takers' strategy use was found to be factors explaining some of the variances in integrated reading-listening-writing test scores. However, previous studies have indicated that there are other variables that could also contribute to writing performance. Future research should examine other factors such as writing anxiety (Cheng, Horwitz, & Schallert, 1999), topical knowledge (Lee & Anderson, 2007), and L2 writing experience (Cumming, 1989). The inclusion of other variables in the model would help illuminate the nature of integrated writing proficiency and offer more comprehensive picture about how different factors interact with integrated writing test performance.

This study focuses on students' test performance on the content, organization, language use and verbatim source use features, but more efforts are needed to clarify how the use of different strategies may affect other features of writing, including the discourse-level features of essays in particular. Revisiting the discourse features of each individual test taker's essays could reveal more information on the impact of individuals' effective and ineffective strategy use on different aspects of the writing such as grammatical accuracy, lexical and syntactic complexity, and argument structure.

Given that the inferences made from test scores depend upon the construct of the measure, more should be known about how differences in task characteristics might relate to different aspects of test takers' essays. Future studies should investigate how

task characteristics affect how test takers incorporate sources into the essay, including how they use effective argumentation style, ways of paraphrasing source materials, and voice to reference evidence from a particular source. Content analyses of stimulus materials and test takers' essays would be informative so that test developers could redesign tasks in order to minimize test takers' tendencies to use source materials verbatim or help raters produce more reliable ratings.

The strategy taxonomies in this study have no doubt been examined in many writing studies, but a further investigation of social and affective strategies may be useful for examining the role of motivation and emotion in completing integrated writing tasks. The level of interest in writing, the level of writing or listening anxiety and beliefs about writing from sources could also be considered important factors that may be facilitative or detrimental to test takers' performance.

CONCLUSIONS

This study set out to investigate the relationship between strategy use and L2 writing ability as measured by an integrated reading-listening-writing task. The present findings show that the task requires test takers to deploy an array of comprehension strategies in coordination with production strategies, which supports the use of such tasks in assessing academic writing ability. In order to complete the task successfully, test takers must use several rhetorical strategies expected in university-level academic contexts in their responses, rather than being able to rely on test-wisness strategies.

The second issue explored in this study was whether different strategy use affected verbatim source use. The resulting SEM model shows that the test-wisness strategies—including copying, patchwriting, and applying previously memorized writing templates—had a detrimental impact on test takers' verbatim source use performance.

The model also indicates little relationship between test takers' verbatim source use and on other measures of writing ability.

As a part of the test validation process, this study presents an effort to further understand the premise, promises, and problems of using integrated writing tasks as a measure of academic writing ability. The study provides information relating to strategy use, source use, and integrated writing test performance. It also presents a working model to explore the underlying construct of integrated reading-listening-writing tasks. The results could be valuable in making informed decisions about test development, test use, and academic writing instruction.

Appendix A Recruitment Flyer



PARTICIPANTS WANTED!



Make \$ 10 in 45-60 minutes

FREE WRITING ASSESSMENT!
FREE WRITING EVALUATION!

IF YOU ARE:

- **A UT student**
- **A nonnative English speaker (English is not your first language.)**

You are welcome to participate in a research study on English writing.

Purpose of Study: To investigate the relationship between learner characteristics and their writing performances in English.

Participants will be asked to

1. Read a passage, listen to a lecture and write a response essay.
2. Fill out a questionnaire about your background and thoughts during the writing task.

If you are interested, please contact:

Hui-chun Yang

eslwriting101@gmail.com

→

(603) 866- 1980

The study will require each participant to come in for 1 visit for 45 to 60 minutes depending on your speed. Participants will receive a gift certificate worth **\$10** for 1 complete visit. Also, you may request **a copy of your own writing scores** assigned by professional ESL teachers in terms of different aspects of your writing so you can know more about your writing ability.

Investigators:

Hui-chun Yang, Doctoral candidate, Foreign Language Education, UT Austin

Professor Lia Plakans, Ph. D., Foreign Language Education, UT Austin

Appendix B Integrated Writing Scoring Rubrics

Content
5 – A response is effective in selecting principal information from the lecture to support ideas and accurately connecting key points in the lecture and the reading.
4 – A response is effective in selecting principal information from the lecture to support ideas and accurately presenting information from the lecture in relation to the writing. Only minor points from the source content are misrepresented, omitted or imprecisely connected.
3 – A response contains some key points from the lecture and connects those points with the reading while the points are imprecisely, incorrectly, or incompletely presented or connected. One major key point made in the lecture may be omitted.
2 – A response contains only few relevant points made in the lecture while they are significantly misrepresented. No, little or incorrect connections are drawn from the reading and the lecture.
1 – A response fails to use any meaningful or relevant content from the lecture.
0 – A response is not connected to the topic, written in a foreign language or left blank.
Organization
5 – A response is well-organized and coherent at the paragraph and essay level.
4 – A response is generally-organized and coherent at the paragraph and essay level.
3 – A response is inconsistently- and incoherently- organized.
2 – A response is poorly-organized and lacks of paragraph and essay cohesion.
1 – A response has insufficient length to determine organization.
0 – A response is left blank or written in a foreign language.
Language Use
5 – A response displays accurate and appropriate sentence formation and word forms although minor usage and grammatical errors may occur occasionally.
4 – A response generally displays accurate and appropriate sentence formation and word forms. Although some noticeable minor usage and grammatical errors may occur occasionally, the clarity of the response is not affected.
3 – A response contains frequent usage and grammatical errors that may result in noticeable obscured meanings and expressions or unclear connection of ideas.
2 – A response contains usage and grammatical errors that may largely obscure meanings or connections of ideas.
1 – A response is almost incomprehensible due to language errors. Little meaning can be derived from the response.
0 – A response is totally incomprehensible due to language errors, or because the response is left blank or is written in a foreign language.
Verbatim Source Use
5 – A response displays no verbatim source use.
4 – A response displays a minimal verbatim source use.
3 – A response displays few instances of verbatim source use due to unsuccessful attempts to paraphrase.

2 – A response displays some instances of verbatim source use.
1 – A response displays many instances of verbatim source use.
0 – A response displays wholesale copying from the reading for most parts.

Appendix C Strategy Inventory for Integrated Writing

What did you do to complete the writing task?

Directions:

Read each statement and indicate how often you used each strategy when writing your essay. Check the appropriate box: 1 (very rarely), 2 (rarely), 3 (occasionally), 4 (often), and 5 (very often)

EXAMPLE:

Very rarely Very often

I ate popcorn while I watched the movie <i>Batman</i> .	1	2	3	4	5
Explanation: If you ate popcorn very often when you watched the movie <i>Batman</i> , check 5.					√

	1	2	3	4	5
A. BEFORE I STARTED WRITING THE ESSAY...					
1. I reread the task requirements carefully.					
2. I thought about the type of essay I wanted to write.					
3. I tried to summarize overall ideas from the whole reading passage in my mind.					
4. I wrote down keywords from what I heard in the lecture.					
5. I tried to understand the relationship between the ideas of the reading and lecture.					
6. I tried to memorize some ideas from the lecture.					
7. I made a writing plan (e.g. outlines, notes, keywords, etc).					
8. I wrote down main ideas and important points from the reading passage.					
9. I searched for connections among sentences.					
10. I tried to understand the content according to how information is organized in each paragraph.					
11. I tried to understand the organization of the reading passage or the lecture.					
12. I summarized ideas from the lecture in my mind.					
13. I searched for connections among paragraphs.					
14. I planned to copy good sentences from the reading or lecture in my writing.					
15. I predicted the content of the lecture after reading the passage.					

B. WHILE I WAS WRITING MY ESSAY...	1	2	3	4	5
16. I double checked to see if my writing met the task requirements.					
17. I reread the reading passage again to look for main ideas.					
18. I wrote some phrases based on a writing template I had memorized before the test.					
19. I thought about mentioning the authors in my essay.					
20. I copied the sentences from the reading passage and revised them.					
21. I reread what I had written to see if my writing met the task requirements.					
22. I tried to write about my knowledge or my own experiences in the essay.					
23. I used different words or phrases to describe ideas from the reading passage or lecture.					
24. I reread what I had written to see if I was using correct English (e.g. grammar or spelling).					
25. I checked if I used the same phrases or sentences as the authors.					
26. I revised the sentences to make my writing clearer.					
27. I thought about a word, phrase, or sentence before I wrote it down.					
28. I first wrote out a writing template I had memorized before and filled in some ideas from the lecture or the reading passage.					

C. AFTER I HAD FINISHED A DRAFT OF MY WRITING...	1	2	3	4	5
29. I checked if I used examples to support my main ideas.					
30. I reread my essay and changed the content that didn't express what I meant.					
31. I added new points based on the reading or the lecture.					
32. I reread my essay and made sure my English was correct.					
33. I checked if I had connected the ideas from the lecture to the ideas from the reading passage.					
34. I made changes in phrases to ensure I didn't copy the exact phrases.					

D. WHAT DO YOU THINK? I NEED YOUR OPINIONS!!

1. What difficulties did you have with listening, reading or writing in this task?
2. Please describe how you used the lecture and the reading passage for your writing.
3. Would you prefer writing with lecture and reading (like this test) to a writing test without lecture and reading? Why or why not?
4. Please describe your motivation to complete this writing test.
5. Are there any other strategies you used that did not appear in this strategy questionnaire?
6. How did you feel when you wrote this essay?

☺ **THANK YOU VERY MUCH**

Appendix D Background Questionnaire

1. Background Information

- Home Country _____
- 1st language _____
- University major _____
- Graduate _____ Undergraduate _____ (please check)
- Male _____ Female _____ (please check)

2. Academic Status

- How many semesters have you studied at the University of Texas? _____
- How long have you studied in the U.S.? _____
- Have you graduated from high school in the U.S.? _____

3. Writing Courses and Experiences

- Have you had previous writing courses in your home country and in the U.S.? _____
If yes, which kinds of writing courses have you taken? (choose all that apply)

- a. English composition course at the UT or your undergraduate university
- b. ESL writing course in an ESL program in the U.S.
- c. TOEFL writing test preparation course in the U.S. or in your home country
- d. English composition courses in high school
- e. Writing courses in my first language
- f. Others _____ (please write)
- What types of writing have you done in your academic courses? (choose all that apply)

- a. expository essays (e.g. comparison and contrast, and cause and effect, etc.)
- b. descriptive essays (e.g. describe an object, person, place, experience, etc.)
- c. narrative essays (e.g. telling a story)
- d. argumentative essays (e.g. establish a position on a topic based on evidence)
- e. lab reports
- f. summaries
- g. research paper (e.g. lit review, methods, results, etc.)

4. Opinions about writing, reading and listening in English:

Check the appropriate box 1 (Never), 2 (Sometimes), 3 (Often), 4 (Usually) and 5 (Always)

	1	2	3	4	5
• I enjoy writing in English.					
• I enjoy reading in English.					
• I enjoy listening in English.					
• I have strong skills in writing in English.					
• I have strong skills in reading in English.					
• I have strong skills in listening in English.					

5. English test reports

- If you have taken TOEFL, which year did you take it last? _____
- Which form of TOEFL did you take? paper-based/computer-based/Internet-based) circle one
- What was your most recent TOEFL score _____
- TOEFL sub-scores: writing _____ grammar _____ listening _____ reading _____
speaking _____

(If you are not sure about your score, please make your best guess of the score)

Appendix E Interview Questions

1. How did you feel about the writing test?
2. Could you talk about the strategies you used to read the passage?
3. Could you talk about the strategies you used to listen to the lecture?
4. Have you done similar tasks like this before? Writing tests? Class assignments?
5. How did you use reading passage in your writing?
6. How did you use listening material in your writing?
7. Is there any strategy you used that did not appear in the strategy questionnaire?
8. Could you tell me a little bit about your experiences of learning English?
9. Could you talk about your educational background?
10. Could you talk about your experiences in writing? In first language? In English?
11. Talk about your experiences learning to read? In your first language? In English?
12. Talk about your experiences learning to listen in English?
13. How do you use writing in your (graduate/undergraduate) courses?
14. What kind of writing assignments do you have in your classes? Do you use reading or listening materials in writing?

References

- Alexander, P. A., Graham, S., & Harris, K. R. A. (1998). A perspective on strategy research: Progress and prospects. *Educational Psychology Review*, 10(2), 129–154.
- Allan, A. (1992). Development and validation of a scale to measure test-wisness in EFL/ESL reading test takers. *Language Testing*, 9(2), 101–122.
- Anderson, N. J. (1991). Individual differences in strategy use in second language reading and testing. *The Modern Language Journal*, 75(4), 460–472.
- Anderson, N. J., Bachman, L. F., Perkins, K., & Cohen, A. D. (1991). An exploratory study into the construct validity of a reading comprehension test: triangulation of data sources. *Language Testing*, 8(1), 41–66.
- Asención Delaney, Y. (2008). Investigating the reading-to-write construct. *Journal of English for Academic Purposes*, 7(3), 140–150.
- Asención, Y. (2004). *Validation of reading-to-write assessment tasks performed by second language learners*. Unpublished doctoral dissertation, Northern Arizona University, Flagstaff, AZ.
- Bachman, L. F. (1990). *Fundamental Consideration in Language Testing*. Oxford: Oxford University Press.
- Bachman, L. F. (1991). What does language testing have to offer? *TESOL Quarterly*, 25(4), 671–704.
- Bachman, L. F. (2002). Some reflections on task-based language performance assessment. *Language Testing*, 19(4), 453–476.
- Bachman, L. F. (2004). *Statistical analysis for language assessment*. Cambridge: Cambridge University Press.
- Bachman, L. F., & Palmer, A. S. (1996). *Language testing in practice*. Oxford: Oxford University Press.
- Baddeley, A. D. (1986). *Working memory*. New York: Oxford University Press.
- Bedell, D. A., & Oxford, R. L. (1996). Cross-cultural comparison of language learning strategies in the People's Republic of China and other countries. In Rebecca L. Oxford (Ed.), *Language learning strategies around the world: Cross-cultural perspective* (pp. 47–60). Honolulu: University of Hawaii Press.
- Bereiter, & Scardamalia. (1987). *The psychology of written composition*. Hillsdale, NJ: Erlbaum.
- Bollen, K. A. (1989). *Structural equations with latent variables*. New York, NY: Wiley.

- Bracewell, R. J., Frederiksen, C. H., & Frederiksen, J. F. (1982). Cognitive processes in composing and comprehending discourse. *Educational Psychologist*, 17, 146–174.
- Braine, G. (2001). When an exit test fails. *System*, 29(2), 221–234.
- Brown, A. L., & Day, J. D. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior*, 22, 1–14.
- Byrne, B. M. (1994). *Structural equation modeling with EQS and EQS/Windows: Basic concepts, applications, and programming*. Thousand Oaks, CA: Sage Publications.
- Campbell, C. (1990). Writing with others' words: Using background reading text in academic compositions. In B. Kroll (Ed.), *Second Language Writing* (pp. 211–230). Cambridge, UK: Cambridge University Press.
- Canale, M., & Swain, M. (1980). Theoretical bases of communicative approaches to second language teaching and testing. *Applied Linguistics*, 1(1), 1–47.
- Chamot, A. U. (1987). The learning strategies of ESL students. In A. L. Wenden & J. Rubin (Eds.), *Learner strategies in language learning* (pp. 71–84). Englewood Cliffs, NJ: Prentice Hall.
- Chamot, A. U., Kupper, L., & Impink-Hernandez, M. (1988). *A study of learning strategies in foreign language instruction: Findings of the longitudinal study*. McLean, VA: Interstate Research Associates.
- Chapelle, C. A. (1999). Validity in language assessment. *Annual Review of Applied Linguistics*, 19, 254–272.
- Charge, N., & Taylor, L. B. (1997). Recent developments in IELTS. *ELT J*, 51(4), 374–380.
- Charmaz, K. (2004). Grounded theory. In N. Hesse-Biber & P. Leavy (Eds.), *Approaches to qualitative research: A reader on theory and practice* (pp. 496–521). Oxford, UK: Oxford University Press.
- Charmaz, K. (2006). *Constructing grounded theory*. Thousand Oaks, CA: Sage.
- Cheng, Y. S., Horwitz, E. K., & Schallert, D. L. (1999). Language anxiety: Differentiating writing and speaking components. *Language Learning*, 49(3), 417–446.
- Cho, Y. (2003). Assessing writing: Are we bound by only one method? *Assessing Writing*, 8(3), 165–191.
- Cohen, A. D. (1998). Strategies and processes in test taking and SLA. In Michael H. Long & Jack C. Richards (Eds.), *Interfaces between second language acquisition and language testing research* (pp. 90–111). Cambridge: Cambridge University Press.
- Cohen, A. D. (2006). The coming of age of research on test-taking strategies. *Language Assessment Quarterly*, 3(4), 307–331.

- Cohen, A. D., & Upton, T. A. (2007). 'I want to go back to the text': Response strategies on the reading subtest of the new TOEFL. *Language Testing*, 24(2), 209–250.
- Connor, U. M. (1987). Research frontiers in writing analysis. *TESOL Quarterly*, 21(4), 677–696.
- Connor, U. M., & Kramer, M. G. (1995). Writing from sources: Case studies of graduate students in business management. In Diane Belcher, George Braine & Diane Dewhurst Belcher (Eds.), *Academic writing in a second language* (pp. 155–182). Norwood, NJ: Ablex Publishing Corporation.
- Cooper, C., Cherry, R., Copley, B., Fleischer, S., Pollard, R., & Startisky, M. (1984). Studying the writing abilities of a university freshman class: Strategies from a case study. In R. Beach & L. Bridwell (Eds.), *New Directions in Composition Research* (pp. 19–52). New York: Guilford.
- Crocker, L. (2006). Preparing examinees for test taking: Guidelines for test developers. In Steven M. Downing & Thomas M. Haladyna (Eds.), *Handbook of test development* (pp. 115–130). Mahwah, NJ: Lawrence Erlbaum Association.
- Cumming, A. (1989). Writing expertise and second language proficiency. *Language Learning*, 39, 81–141.
- Cumming, A. (1998). Theoretical perspectives on writing. *Annual Review of Applied Linguistics*, 18, 61–78.
- Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. (2004). A teacher-verification study of speaking and writing prototype tasks for a new TOEFL. *Language Testing*, 21(2), 159–197.
- Cumming, A., Grant, L., Mulcahy-Ernt, P., & Powers, D. E. (2005). *A teacher-verification study of speaking and writing prototype tasks for a new TOEFL* (No. RR-04-05). Princeton, NJ: Educational Testing Service.
- Cumming, A., Kantor, R., Baba, K., Erdosy, U., Eouanzoui, K., & James, M. (2005). Difference in written discourse in independent and integrated prototype tasks for next generation TOEFL. *Assessing Writing*, 10, 5–43.
- Cumming, A., Kantor, R., Powers, D., Santos, T., & Taylor, C. (2000). *TOEFL 2000 Writing framework: A working paper*. Princeton, NJ: Educational Testing Service.
- Currie, P. (1998). Staying out of trouble: Apparent plagiarism and academic survival. *Journal of Second Language Writing*, 7(1), 1–18.
- D'oronyei, Z. (2003). *Questionnaires in second language research: Construction, administration and processing*. Mahwah, NJ: Erlbaum.
- de Larios, J. R., Manchon, R. M., Murphy, L., & Marin, J. (2008). The foreign language writer's strategic behavior in the allocation of time to writing processes. *Journal of Second Language Writing*, 17, 30–47.

- de Larios, J. R., Murphy, L., & Manchon, R. M. (1999). The use of restructuring strategies in EFL writing: A study of Spanish learners of English as a foreign language. *Journal of Second Language Writing*, 8(1), 13–44.
- Deckert, G. D. (1993). Perspectives on plagiarism from ESL students in Hong Kong. *Journal of Second Language Writing*, 2(2), 131–148.
- Ellis, R. (1994). *The study of second language acquisition*. Oxford: Oxford University Press.
- Enright, M. K., Bridgeman, B., & Cline, F. (2002). *Prototyping a test design for a new TOEFL*. Paper presented at the National Council on Measurement in Education, New Orleans, LA.
- Esmaili, H. (2002). Integrated reading and writing tasks and ESL students' reading and writing performance in an English language test. *The Canadian Modern Language Review* 58(4), 599–622.
- Feak, C., & Dobson, B. (1996). Building on the impromptu: A source-based academic writing assessment. *College ESL*, 6(1), 73–84.
- Flower, L. (1994). *The construction of negotiated meaning: A sociocognitive theory of writing*. Carbondale, IL: Southern Illinois Press.
- Flower, L., & Hayes, J. R. (1980). The Cognition of Discovery: Defining a Rhetorical Problem. *College Composition and Communication*, 31(1), 21–32.
- Flower, L., & Hayes, J. R. (1981). A Cognitive Process Theory of Writing. *College Composition and Communication*, 32(4), 365–387.
- Flower, L., Stein, B., Ackerman, J., Kantz, M. J., McCormick, K., & Peck, W. C. (1990). *Reading to write: Exploring a cognitive and social process*. Oxford, UK: Oxford University Press.
- Gilham, B. (2000). *Developing a questionnaire*. London: Continuum.
- Glaser, R. (1984). Education and thinking: The role of knowledge. *American Psychologist*, 40(5), 573–574.
- Grabe, W. (2001). Reading-writing relations: Theoretical perspectives and instructional practices. In D. Belcher & A. Hirvela (Eds.), *Linking literacies. Perspectives on L2 reading-writing connections* (pp. 15–47). Ann Arbor, MI: University of Michigan Press.
- Grabe, W., & Kaplan, R. B. (1996). *Theory and practice of writing*. New York: Longman.
- Haladyna, T. M., & Downing, S. M. (2004). Construct-irrelevant variance in high-stakes testing. *Educational Measurement*, 23(1), 17–27.
- Hamp-Lyons, L. (1991). Scoring procedures for ESL contexts. In Liz Hamp-Lyons (Ed.), *Assessing second language writing in academic contexts*. Norwood, NJ: Ablex.

- Hamp-Lyons, L. (2001). Fourth generation writing assessment. In Tony Silva & Paul Kei Matsuda (Eds.), *On second language writing* (pp. 117–128). Mahwah, NJ: Lawrence Erlbaum Associates.
- Hayes, J. R. (1996). A new framework for understanding cognition and affect in writing. In C. M. Levy & S. Ransdell (Eds.), *The Science of Writing: Theories, Methods, Individual Differences, and Applications* (pp. 1–27). Mahwah, NJ: Erlbaum.
- Herche, J., & Engelland, B. (1996). Reversed-polarity items and scale unidimensionality. *Journal of Academy of Marketing Science*, 24, 366–374.
- Homburg, T. J., & Spaan, M. C. (1981). ESL Reading proficiency assessment: Testing strategies. In M. Hines & W. Rutherford (Eds.), *On TESOL '81* (pp. 25–33). Washington, DC: TESOL.
- Howard, R. M. (1995). Plagiarisms, authorships, and the academic death penalty. *College English*, 57(7), 788–806.
- Hsiao, T.-Y., & Oxford, R. L. (2002). Comparing Theories of Language Learning Strategies: A Confirmatory Factor Analysis. *The Modern Language Journal*, 86(3), 368–383.
- Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling*, 6, 1–55.
- Hyland, T. A. (2005). *A study of the practices of EL1 and EL2 students in reporting information from sources in a timed writing proficiency assessment*. Unpublished Ed.D., University of Toronto (Canada), Canada.
- Jones, R. L. (1985). Second language performance testing: An overview. In P. Hauptman, R. LeBlanc & M. Wesche (Eds.), *Second language performance testing* (pp. 15–24). Ottawa, CA: University of Ottawa Press.
- Jöreskog, K. G., & Sörbom, D. (1989). *LISREL 7: A guide to the program and applications* (2nd ed.). Chicago, IL: SPSS.
- Keith, T. Z. (2006). *Multiple regression and beyond*. Boston, MA: Pearson.
- Kennedy, M. (1985). The composing process of college students writing from sources. *Written Communication*, 2, 434–456.
- Kenny, D. A. (1979). *Correlation and causality*. New York, NY: Wiley.
- Kim, J. O., & Mueller, C. W. (1978). *Introduction to factor analysis: What it is and how to do it*. Newbury Park, CA: Sage University Press.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. New York: Cambridge.
- Kintsch, W., & van Dijk, T. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85(5), 363–394.
- Kirkland, M. R., & Saunders, M. A. P. (1991). Maximizing student performance in summary writing: Managing cognitive load. *TESOL Quarterly*, 25(1), 105–121.

- Kline, R. B. (2004). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Publications.
- Kline, T. J. B. (2005). *Psychological testing*. Thousand Oaks, CA: Sage Publications.
- Koda, K. (2005). *Insights into second language reading*. Cambridge, UK: Cambridge University Press.
- Kunnan, A. J. (1998). An introduction to structural equation modeling for language assessment research. *Language Testing*, 15(3), 295–332.
- Lee, H., –K., & Anderson, C. (2007). Validity and topic generality of a writing performance test. *Language Testing*, 24(3), 307–330.
- Lee, Y.-W., & Kantor, R. (2005). *Dependability of new ESL writing test scores: Evaluating prototype tasks and alternative rating schemes* (No. MS-31). Princeton, NJ: Educational Testing Service.
- Lewkowicz, J. (1994). *Writing from Sources: Does Source Material Help or Hinder Students' Performance?*
- Lewkowicz, J. (1997). The integrated testing of a second language. In C. Clapham & D. Corson (Eds.), *Encyclopedia of language and education* (Vol. 7, pp. 121–130). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Manchon, R. M., & de Larios, J. R. (2007). On the Temporal Nature of Planning in L1 and L2 Composing. *Language Learning*, 57(4), 549–593.
- Mateos, M., Villalon, R., de Dios, M. J., & Martin, E. (2007). Reading and writing tasks on different university degree courses: What do the students say they do? *Studies in Higher Education*, 32(4), 489–450.
- McNamara, T. (1996). *Measuring second language writing performance*. New York: Addison Wesley Longman.
- Messer, S. D. (1997). *Evaluating ESL written summaries: An investigation of the ESL Integrated Summary Profile (ISP) as a measure of the summary writing ability of ESL students*. Unpublished doctoral dissertation, Florida State University, Tallahassee, FL.
- Messick, S. (1989). Validity. In *Educational Measurement* (3rd ed., pp. 13–104). New York: Macmillan.
- Moore, T. (1997). From test to note: Cultural variation in summarization practices. *Prospect*, 12, 54–36.
- Myford, C., & Wolfe, E. (2002). When raters disagree, then what: Examining a third-rating discrepancy resolution procedure and its utility for identifying unusual patterns of ratings. *Journal of Applied Measurement*(3), 300–324.
- O'Malley, J. M., & Chamot, A. U. (1990). *Learning strategies in second language acquisition*. Cambridge: Cambridge University Press.

- Oxford, R. L. (1990). *Language learning strategies: What every teacher should know*. Boston: Heinle.
- Oxford, R. L. (1996). Preface: Why is culture important for language learning strategies? In Rebecca L. Oxford (Ed.), *Language learning strategies around the world: Cross-cultural perspective* (pp. ix-xv). Honolulu: University of Hawaii Press.
- Pennington, M. C., & So, S. (1993). Comparing writing process and product across two languages: A study of 6 Singaporean university student writers. *Journal of Second Language Writing*, 2(1), 41–63.
- Pennycook, A. (1994). The complex contexts of plagiarism: A reply to Deckert. *Journal of Second Language Writing*, 3(3), 277–284.
- Pennycook, A. (1996). Borrowing other's words: Texts, ownership, memory, and plagiarism. *TESOL Quarterly*, 30, 210–230.
- Phakiti, A. (2003). A closer look at the relationship of cognitive and metacognitive strategy use to EFL reading achievement test performance. *Language Testing*, 20, 26–56.
- Plakans, L. (2008). Comparing composing processes in writing-only and reading-to-write test tasks. *Assessing Writing*, 13, 111–129.
- Plakans, L. (2009). The role of reading strategies in integrated L2 writing tasks. Unpublished manuscript.
- Plakans, L. M. (2007). *Second language writing and reading-to-write assessment tasks: A process study*. Unpublished doctoral dissertation, The University of Iowa, Iowa City, IA.
- Pressley, M., Goodchild, F., Fleet, J., Zajchowski, R., & Evans, E. D. (1989). The challenges of classroom strategy instruction. *Elementary School Journal*, 89(3), 301–342.
- Purpura, J. E. (1997). An analysis of the relationships between test takers' cognitive and metacognitive strategy use and second language test performance. *Language Learning*, 47(2), 289–325.
- Purpura, J. E. (1999). *Learner strategy use and performance on language tests: A structural equation modeling approach*. Cambridge: University of Cambridge Local Examinations Syndicate and Cambridge University Press.
- Raimes, A. (1985). What unskilled ESL students do as they write: A classroom study of composition. *TESOL Quarterly*, 19(2), 229–258.
- Raimes, A. (1990). The TOEFL test of written English: Causes for concern. *TESOL Quarterly*, 24, 427–442.
- Read, J. (1990). Providing relevant content in an EAP writing test. *English for Specific Purposes*, 9, 109–121.
- Sasaki, M. (2000). Toward an Empirical Model of EFL Writing Processes: An Exploratory Study. *Journal of Second Language Writing*, 9(3), 259–291.

- Sasaki, M., & Hirose, K. (1996). Explanatory variables for EFL students' expository writing. *Language Learning*, 46(1), 137–174.
- Sasaki, M., & Hirose, K. (1999). Development of an analytic rating scale for Japanese L1 writing. *Language Testing*, 16(4), 457–478.
- Schallert, D. L. (1987). Thought and language, content and structure in language communication. In J. R. Squire (Ed.), *The dynamics of language learning* (pp. 65–79). Urbana, IL: National Council of Teachers of English.
- Schmidt, R. (1994). Deconstructing consciousness in search of useful definitions for applied linguistics. *AILA Review*, 11, 11–16.
- Schumacker, R. E., & Lomax, R. G. (1996). *A beginner's guide to structural equation modeling*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Scollon, R. (1994). As a matter of fact: The changing ideology of authorship and responsibility in discourse. *World Englishes*, 13(1), 33–46.
- Song, X., & Cheng, L. (2006). Language learner strategy use and test performance of Chinese learners of English. *Language Assessment Quarterly*, 3(3), 243–266.
- Spivey, N. (1984). *Discourse synthesis: Constructing texts in reading and writing*. Newark, DE: International Reading Association.
- Spivey, N. (1990). Transforming texts: Constructive processes in reading and writing. *Written Communication*, 7, 256–287.
- Spivey, N. (1997). *The constructivist metaphor: Reading, writing, and the making of meaning*. New York: Academic Press.
- Spivey, N. N., & King, J. R. (1989). Readers as writers composing from sources. *Reading Research Quarterly*, 24, 7–26.
- Steckler, A., McLeroy, K. R., Goodman, R. M., Bird, S. T., & McCormick, L. (1992). Toward integrating qualitative and quantitative methods: An introduction. *Health Education Quarterly*, 19(1), 1–8.
- Stein, V. (1990). Exploring the cognition of reading-to-write. In L. Flower, V. Stein, J. Ackerman, M.J. Kantz, K. McCormick & W.C. Peck (Eds.), *Reading-to-write: Exploring a cognitive and social process* (pp. 119–143). New York, NY: Oxford University Press.
- Stevens, J. (1992). *Applied multivariate statistics for the social science* (2nd ed.). Hillsdale, NJ: Erlbaum.
- TOEFL iBT Data Set 3: Writing*. (2008). Princeton, NJ: Educational Testing Service.
- Torrance, M., Thomas, G. V., & Robinson, E. J. (1994). The Writing Strategies of Graduate Research Students in the Social Sciences. *Higher Education*, 27(3), 379–392.

- Trites, L. J. (2000). *Beyond basic comprehension: Reading to learn and reading to integrate for native and nonnative speakers*. Unpublished doctoral dissertation, Northern Arizona University, Flagstaff, AZ.
- Tsai, C. H. L. (2004). *Investigating the relationships between ESL writers' strategy use and their second language writing ability*. Unpublished doctoral dissertation, Columbia University, New York, NY.
- Wall, D., & Horák, T. (2006). *The impact of changes in the TOEFL examination on teaching and learning in central and eastern Europe: Phase 1, the baseline study (TOEFL Monograph No. MS-34)*. Princeton, NJ: Educational Testing Service.
- Watanabe, Y. (2001). *Read-to-write tasks for the assessment of second language academic writing skills: Investigating text features and rater reactions*. Unpublished doctoral dissertation, University of Hawaii-Manoa, Honolulu, HI.
- Weigle, S. C. (1999). Investigating rater/prompt interactions in writing assessment: Quantitative and qualitative approaches. *Assessing Writing*, 6(2), 145–178.
- Weigle, S. C. (2002). *Assessing writing*. Cambridge: Cambridge University Press.
- Weigle, S. C. (2004). Integrating reading and writing in a competency test for non-native speakers of English. *Assessing Writing*, 9, 27–55.
- Weir, C. J. (1993). *Understanding and developing language tests*. Hemel, Hamstead, England: Prentice Hall.
- Wenden, A. L. (1991). *Learner Strategies for Learner Autonomy*. Englewood Cliffs, NJ: Prentice-Hall.
- Wesche, B. (1987). Second language performance testing: The Ontario test of ESL as an example. *Language Testing*, 4(1), 28–47.
- Weston, R., & Gore, P. A. (2006). A brief guide to structural equation modeling. *The Counseling Psychologist*, 34(5), 719–751.
- Whalen, K., & Menard, N. (1995). L1 and L2 writers' strategic and linguistic knowledge: A model of multiple-level discourse. *Language Learning*, 45(3), 381.
- Yang, L., & Shi, L. (2003). Exploring six MBA students' summary writing by introspection. *Journal of English for Academic Purposes*, 2, 165–192.

Vita

In 1995, Hui-Chun Yang attended Municipal Chien Chen Senior High School, Kaohsiung, Taiwan. In 1998, she entered Tunghai University in Taichung, Taiwan, where she received the degree of Bachelor of Arts in Foreign Language and Literature in May 2002. For the following year, she worked as an intern English teacher in the Affiliated High School of Tunghai University, Taichung, Taiwan. In September 2004 she entered the University of Texas at Austin. She completed the requirements for the degree of Master of Arts in Foreign Language Education in December 2005. In January 2006 she entered the doctoral program in Foreign Language Education at the University of Texas at Austin. In summer 2007 she completed a research internship at Educational Testing Service. From 2007 until 2008 she was employed as a research assistant in Foreign Language Education at the University of Texas at Austin. While she was in the program, she was awarded a Continuing Student Fellowship and TOEFL Small Grant for Doctoral Research in Second or Foreign Language Assessment. She is interested in second language writing, language testing, computer-assisted language learning, assessment-based learning, language anxiety and interlanguage pragmatics. She has presented research at various venues including the American Association for Applied Linguistics Conference, the American Educational Research Association Meeting, the Language Testing Research Colloquium, the Texas TESOL State Conference, the Hawaii International Conference on Education and the Rocky Mountain Modern Language Association Meeting.

Permanent Address: 64, LN 174, Meng-Zih Rd., Kaohsiung City, 813, Taiwan

This dissertation was typed by Hui-Chun Yang.