

EXAMINING EFFECTS OF NOTE-TAKING AND EXTENDED WRITING
ON THE EXPOSITORY TEXT COMPREHENSION
OF FOURTH GRADE STUDENTS

MICHAEL ARMAND HEBERT

Dissertation under the direction of Professor Steve Graham

The purposes of this study were to examine (1) the extent to which taking notes about text and writing an extended response about text would enhance reading comprehension for fourth grade students, (2) whether note taking was more effective than extended writing for improving reading comprehension across three measures, and (3) whether the effects of the writing tasks were moderated by student writing ability. Students were randomly assigned to a note taking condition which they took notes about an expository text, an extended writing condition in which they compared and contrasted ideas from the text with their own experiences, or a read and study control condition in which they studied the important ideas from the text. Minimal instruction was provided to the students in each treatment group during a single 45 minute session, primarily to ensure they understood their assigned task. The students then met for another 45 minute session, during which they were asked to read an expository passage and complete their assigned task. Students' reading comprehension was tested using three measures. Students in the two writing groups made significantly greater gains than students in the read and study condition on the multiple choice inference measure. However, the results are tempered by low internal consistency found for the measure. No other statistically significant differences were found between the treatment groups, and no significant

moderator effects were found. Implications for future research are framed in terms of the limitations of the study.

EXAMINING EFFECTS OF NOTE-TAKING AND EXTENDED WRITING
ON THE EXPOSITORY TEXT COMPREHENSION
OF FOURTH GRADE STUDENTS

By

Michael Armand Hebert

Dissertation

Submitted to the Faculty of the
Graduate School of Vanderbilt University
in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Special Education

December, 2012

Nashville, TN

Approved:

Professor Steve Graham

Professor Karen Harris

Professor Donald Compton

Professor Bridget Dalton

To my parents, Daniel and Jean, who taught me all the important stuff
and
To Brooke, for reminding me not to neglect the fun stuff

ACKNOWLEDGEMENTS

This work would not have been possible without the financial support of Vanderbilt's predoctoral Experimental Education Research Training (ExpERT) program, funded by the Institute of Education Sciences (grant number R305B040110). I am grateful to Dr. David Cordray, Director of the ExpERT program, who afforded me the opportunity to participate in the program and who also provided valuable instruction and direction in the development of my research skills.

I would like to thank the members of my Dissertation Committee for providing me with the guidance and insight necessary to complete this work. Each of them has taught me a great deal about conducting scientific research, and, perhaps more importantly, have provided models of integrity and excellence to strive for in my career. I am especially indebted to Dr. Steve Graham, my mentor, advisor, teacher, and friend. He is the most generous and supportive person that I have had the privilege to work with, and has become one of the most influential people in my life.

Of course, the people who have made the pursuit of this project worthwhile are my family and friends, who are too numerous to name individually, but collectively bring out my best. I would specifically like to thank my parents, Daniel and Jean, who have encouraged me in every one of my life's pursuits. I would also like to thank Jamie, who, with the help of the other Erving boys, never lets me forget who I really am.

TABLE OF CONTENTS

	Page
DEDICATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF TABLES	vi
 Chapter	
I. INTRODUCTION	1
Purposes of the Current Study	4
Purpose one	4
Purpose two	6
Purpose three	8
Research Questions	8
Hypotheses	8
Research Question 1: Are the Combined Writing Treatments (CW) more effective than the Read and Study Treatment (RS) for improving the reading comprehension of fourth-grade students after controlling for initial writing ability?	9
CW vs. RS: Hypothesis One – Topic Knowledge Measure	9
CW vs. RS: Hypothesis Two – Multiple Choice Inference Measure ..	10
CW vs. RS: Hypothesis Three – Essay Application Measure	11
Research Question 2: Is note taking more effective than extended writing (EW) for improving the reading comprehension of fourth grade students after controlling for initial writing ability?	11
NT vs. EW: Hypothesis Four – Topic Knowledge Measure	12
NT vs. EW: Hypothesis Five – Multiple Choice Inference Measure ..	12
NT vs. EW: Hypothesis Six – Application Essay Measure	12
Research Question 3: Does writing ability moderate the effects of writing the treatment conditions for questions one and two?	13
Does writing ability moderate the effects of treatment for the comparison of CW to RS?	13
Does writing ability moderate the effects of treatment for the comparison of NT to EW?	13
II. METHOD	15
Participants	15
Treatment Conditions and Random Assignment	16
Read and Study	17
Note-taking	17
Extended Writing	18
Experimental Reading Passages	19
Measures	19
Pretest Measure: Wechsler Individual Achievement Test, 3 rd Edition ..	19
Outcome Measures: Topic Knowledge, Multiple Choice Inference, & Application Essay	20
Topic Knowledge (Aligned with NT treatment)	21
Application Essay (Aligned with EW treatment)	22
Multiple Choice Inference Measure (Treatment-independent)	24

	Study Implementation.....	25
	Day 1 – Pretesting	25
	Day 2 – Examples and modeling for each of the treatment groups ...	25
	Day 3 – Students complete their assigned writing or studying tasks.	26
	Day 4 – Posttests	27
III.	ANALYSIS.....	29
	Data Modifications.....	30
	Data Transformations.....	31
	Potential Outliers	31
IV.	RESULTS	33
	Treatment Integrity	33
	Regression Models for Outcome Measures	33
	Multiple Choice Outcome.....	34
	Application Essay	35
	Topic Knowledge-Total Correct.....	36
	Topic Knowledge-Proportion Correct	37
V.	DISCUSSION.....	38
	Is writing more effective than reading and studying for improving the expository text comprehension of fourth graders?.....	38
	Is note-taking more effective than extended writing for improving the expository text comprehension of fourth grade students, after controlling for initial writing ability?	40
	Lack of Power.....	40
	Questionable Measurement.....	40
	Students did not complete the writing tasks as anticipated.....	41
	Treatment contamination	43
	Did writing ability moderate the effects of writing the treatment conditions for questions one and two?	43
VI.	CONCLUSION.....	45
	Appendix	
	A. TABLES (1-5)	47
	B. STUDY TIPS AND EXERCISE SHEETS.....	52
	C. TOPIC KNOWLEDGE MEASURE.....	60
	D. TOPIC KNOWLEDGE ANSWER KEY	63
	E. APPLICATION ESSAY RUBRICS	65
	F. MULTIPLE CHOICE INFERENCE MEASURE	69
	G. NOTE-TAKING INSTRUCTIONS	74
	H. STUDENT WRITING EXAMPLES	76
	REFERENCES	88

LIST OF TABLES

Table	Page
1. Demographic Information of Student by Treatment Condition.....	48
2. Key Activities Completed in Instructional Steps for each Treatment Condition ..	49
3. Intra-class Correlations for the Unconditional and Fully Defined Multilevel Models.....	50
4. Means and Standard Deviations for each Treatment Condition on Three Outcome Measures	50
5. Summary of Regression Analyses for the Effects of Treatment and Writing Skill on Four Reading Outcomes	51

CHAPTER I

INTRODUCTION

Despite large scale efforts to improve students' reading in policy endeavors such as *No Child Left Behind* and *Reading First*, a large number of students in this country are not particularly good readers. The 2009 National Assessment of Educational Progress (NAEP; National Center for Educational Statistics, 2010) reported that only 38% of 12th grade students performed at or above the "proficient" level in reading (defined as solid academic performance). In terms of younger students, only 33% of 4th graders and 32% of 8th graders performed at the proficient level or above (National Center for Educational Statistics, 2009). In contrast, 34%, 43%, and 36% of 4th, 8th, and 12th grade students, respectively, scored at the "basic" level, denoting only partial mastery of the literacy skills needed at their grade-level. The rest of the tested students' scores were below this basic level.

A potentially powerful and often overlooked tool for improving learning and reading comprehension outcomes is writing. Some theorists have argued that writing about information enhances learning or causes new learning to occur (Klein, 1999; Newell, 2007). Indeed, systematic reviews of experimental and quasi-experimental literature found writing activities to be effective for improving content area learning (Graham & Perin, 2007) and academic outcomes (Bangert-Drowns, Hurley, & Wilkinson, 2004).

Klein (1999) argued that writing may facilitate learning in four ways: 1) writing forces explicitness and structured thinking through semantic and syntactic choices, 2) it creates a permanent product that can be reviewed and transformed when contradictions arise, 3) it requires authors to construct relationships among ideas, and 4) the act of writing forces writers to generate and revise goals for the audience based on new content and ideas. It has further been suggested that the cognitive processes involved in writing

correspond to general modes of learning that can be actively applied through metacognitive and self-regulation strategies by writers to improve their learning (Bangert-Drowns et al, 2004). That is, students who actively engage in thinking about their own thinking during writing are more likely to plan, monitor, evaluate and adapt the strategies they use to elaborate and organize ideas, build conceptual frameworks, and synthesize knowledge. Because these hypotheses address different aspects of writing, it is possible that learning is facilitated through all of them, none of them, or different combinations of them, depending on the student and the situation.

Theories for how writing influences learning also apply more specifically to the influence of writing on reading comprehension. In a recent review, Graham and Hebert (2010, 2011) provided empirical evidence to support the claim that writing about text read improves comprehension of it. In their review, Graham and Hebert used meta-analysis to examine the effects of writing on reading by including studies that compared treatment conditions in which students wrote about text read using various writing activities (e.g., summary writing, answering or generating written questions, note-taking, and extended writing activities) to no-writing control conditions in which students read the same text and participated in non-writing activities such as rereading or oral discussion of the text. Based on the results of the meta-analysis, they concluded that writing about text was effective in improving reading comprehension as measured by both norm-referenced ($ES = 0.40$) and researcher created ($ES = 0.51$) assessments. Most of these analyses involved comparing writing about text to either reading it or reading and studying it.

There is also some empirical evidence to support the contention that different writing tasks result in varying effects on reading comprehension for students. Graham and Hebert (2010, 2011) conducted breakout analyses to examine the effectiveness of specific writing activities (i.e., summary writing, generating and answering questions, note taking, and extended writing activities). Effect sizes ranged from 0.27 (generating

or answering questions) to 0.77 (extended writing activities), suggesting that the magnitude at which writing impacts reading may differ by task. However, they did not compare the effects of these writing activities either one to another, or for writers of different ability levels.

It is further possible that distinct writing tasks may lead to differential effects on reading comprehension depending on the way it is measured. That is, some writing-to-read tasks may facilitate distinct types of learning and thinking. Langer and Applebee (1987) argued that tasks such as note-taking, summary writing, and answering questions focus students' attention on the text as a whole and lead to superficial manipulation of the content, while analytic writing tasks focus the writer on a narrower body of information, but require more in-depth processing and reorganization of that information. Therefore, various writing tasks may cause students to perform differently on assessments of reading comprehension based on the focus of the writing task and the information tapped by an assessment. Some support for this contention was provided in a meta-analysis by Hebert, Gillespie, and Graham. (2012), who found that specific writing tasks were sometimes more effective when comprehension was assessed using treatment-inherent measures (i.e., measures that are highly similar to the writing task, as opposed to treatment-independent measures that are not so tightly tied to the writing activity; Slavin, 2008a, 2008b) . In other words, the impact of a specific writing activity may well depend on how comprehension is assessed.

Finally, it is possible that the impact of specific writing activities on reading comprehension depends on students' abilities. Graham and Hebert (2010, 2011) found that the average weighted effect size for writing about text was 0.63 for lower-achieving students, indicating that writing was also a useful activity for improving the reading comprehension of such students. In fact, this effect size was larger than the effect found for studies involving all students, suggesting that students with learning difficulties may actually benefit more from writing than their normally achieving peers. One potential

explanation is that students with weak reading skills benefit more from using another skill to augment their comprehension, while students with stronger reading skills benefit less from writing as they are better at comprehending text in general.

In contrast, it is also reasonable to expect that students with poor writing skills may be less adept at using writing to improve their reading skills, and may not be able to do so without instruction. Indeed, an average weighted effect size for writing about text was equal to zero when lower-achieving students were not taught how to use the targeted writing activities (Graham and Hebert, 2010; 2011). Moreover, the authors suggest, “Students who do not develop strong writing skills may not be able to take full advantage of the power of writing as a tool to strengthen reading” (p. 29, Graham & Hebert, 2010). Consequently, students’ writing ability may mediate the effectiveness of writing to read activities in general or even for specific activities with regards to particular measures.

Purposes of the Current Study

There were three purposes for the current study. These purposes resulted in the inclusion of three treatment groups. An elaborated description of each of the conditions can be found in the Method section, although they are briefly introduced here to better establish the purposes of this study.

Purpose one. One purpose of the current study was to examine whether writing was more effective than reading and studying (RS) for improving the expository text comprehension of fourth grade students after controlling for students’ initial writing ability. Two writing tasks, note-taking (NT) and extended writing (EW) were included as separate treatments in the study, but statistically combined for comparison to the RS condition. The NT task asked students to take notes on important information in the text, while the EW task asked students to write an essay comparing and contrasting information from the text with something from their personal experience (see the Method section for a more complete description of the treatments). These specific writing tasks were chosen because each theoretically required students to attend to different types of

information in text, providing a broader test of the effects of writing about reading than is possible with just a single writing activity. In addition, these two writing activities represent common approaches for writing about text (Graham & Hebert, 2010).

Taking notes allows students to identify and record important information across text, while not requiring students to write connected text. Kiewra (1989) suggests that note-taking is an effective tool for writing about text because it is an encoding function that helps increase attention to text and allows for surface organization of text material. Langer and Applebee (1987) further assert that note-taking focuses student attention on the content of passage specific ideas and allows students to read in small segments, but typically results in little integration of these ideas. Note-taking may be well suited to the writing skills of weaker writers as it allows for brevity, while not requiring much organization or elaboration of ideas.

Extended writing tasks on the other hand, require students to reformulate and extend ideas (Kiewra, 1989), and focuses attention on generating, integrating, evaluating, combining, and recombining ideas, resulting in a deeper level of processing (Langer & Applebee, 1987). Such tasks often focus students' attention on specific aspects of text (versus the whole text in general). However, the length and complexity of these tasks may make it more difficult for some students to take advantage of these tasks to learn from reading. There is very little experimental research examining the effectiveness of extended writing to improve reading outcomes, and almost no research examining the effects of such tasks with weaker writers. While these tasks have been particularly powerful in enhancing reading comprehension of students in general (Graham & Hebert, 2010), we need to know whether student writing ability significantly impacts the effectiveness of such tasks.

As previously stated, the EW and NT treatments were statistically combined and compared to a control condition in which students were asked to simply read and study (RS). The combined treatments are designated as combined writing (CW) throughout the

remainder of this report for ease of discussion. It was decided to combine the writing groups for this comparison to obtain a more complete comparison of the effects of writing to RS, while keeping the comparisons in the study orthogonal.

Fourth grade was selected as the ideal grade level for examining these writing treatments. There were three reasons for this. One, it is widely agreed that fourth grade is the grade-level at which students make the transition from learning to read to reading to learn (Chall, 1983, 1996). The decision to use fourth grade students as participants in the current study was aimed at examining whether writing can facilitate this transition. Two, studies identified in the reviews conducted by Graham and Hebert (2010) and Hebert et al. (2012) examining the effects of unstructured note-taking on reading comprehension included students in fifth grade or later. Therefore, the proposed study extends the research base on the use of unstructured note-taking with slightly younger students. Three, only three studies were found examining the effects of extended writing tasks on reading for students in fourth grade or younger in the Graham and Hebert (2010) review, and all of those studies involved writing about story text; two asked students to write personal responses, and one asked students to summarize and make inferences. In this study, students were asked to write a compare and contrast essay involving analysis and interpretation of expository text.

Purpose two. A second purpose of this study was to compare whether NT was more effective than EW for improving the reading comprehension of fourth grade writers, after controlling for initial writing ability. As previously stated, NT and EW activities should result in students attending to different aspects of the text, which, in turn, should lead to differential effects on reading comprehension. Additionally, fourth grade writers may perform better or worse on a particular writing activity, which may allow them to take better advantage of one task over another. They may, however, perform equally well, or poorly, on both activities. Comparing the effectiveness of these tasks was aimed at determining whether one writing task was more effective than

another, or if the tasks were equally effective or ineffective for improving reading outcomes.

Because the effectiveness of these two writing activities may be dependent on the measures used to assess reading comprehension, three measures were used to assess reading comprehension. Two of the measures were designed to be “treatment-inherent” measures, one aligned with the NT treatment and one aligned with the EW treatment, and one measure was designed to be a “treatment-independent” measure (Slavin, 2008a, 2008b). First, a topic knowledge measure was used to evaluate how much information students remember across the whole text. The topic knowledge measure required students to write as much as they could remember about topics across the text, and it was selected because it is closely aligned to the NT activity. That is, the method of response was similar to the expected writing that occurred in the NT treatment condition, and the activity and assessment were designed to focus students’ attention on the same content.

Next, a multiple choice measure was used to assess students’ ability to make inferences about information across the whole text. The multiple choice measure focused on information presented across the whole text, but forced students to think beyond the information directly presented. As the measure involved no writing and required students to process information that was not the focus of any of the conditions, the measure should not have overly favored either of the two writing conditions, providing a “treatment independent” measure of comprehension.

Finally, students were asked to write an essay involving the application of some of the ideas presented in the text. Specifically, students were asked to apply underlying concepts that were the focus of the compare and contrast essay in the EW condition to a new situation. The essay measure required students to analyze information about the concepts from the text, deeply process the information to determine how the concepts fit the new situation, and reorganize the information to fit the writing task. This task was chosen because it aligned with the task employed in the EW condition.

Purpose three. A final purpose of this study was to examine whether the treatment comparisons made in purposes one and two were moderated by students writing ability. That is, the homogeneity of regression lines assumption was examined to determine if there was consistency of effects for treatment across different levels of student writing ability. To investigate this, students' initial writing ability was assessed prior to the experiment and included as a covariate in the final regression models. Interactions were then created between writing ability and the treatment comparison variables, and also included in the statistical models as a test for homogeneity. That is, the interaction acted as a test to determine whether the effects of treatment were consistent across levels of student writing ability.

Research Questions

This study was designed in terms of its three purposes to answer three research questions: (1) Are the combined writing treatments more effective than reading and studying for improving expository text comprehension for fourth grade students, after controlling for initial writing ability? (2) Is note-taking more effective than extended writing for improving the expository text comprehension of fourth grade students, after controlling for initial writing ability? (3) Does writing ability moderate the effects of the treatment conditions for questions one and two? The following sections describe the experimental design and hypotheses related to each question.

Hypotheses

The research questions were examined using a true experiment, with students randomly assigned to treatments across multiple sites. Hypotheses regarding the results of the study are presented for research questions one and two for each of the three measures. The hypotheses for research question three are presented in relation to each of the first two research questions.

Research Question 1: Are the Combined Writing Treatments (CW) more effective than the Read and Study Treatment (RS) for improving the reading comprehension of fourth-grade students after controlling for initial writing ability?

The hypotheses for the comparison between CW and RS were based on the notion that the two writing treatments (NT & EW) would cause students to think differently about text, but that the combined groups would provide an indication of an overall effect for writing, as they involve different but typical ways of writing about text. Because CW was a combination of NT and EW, each hypothesis includes a discussion of both of these writing treatments in comparison to RS.

CW vs. RS: Hypothesis One – Topic Knowledge Measure. Students in the CW treatments were expected to outperform students in RS treatment on a measure of topic knowledge, as both NT and EW were expected to outperform RS for different reasons. First, while both the NT and RS tasks were designed to focus students' attention on surface ideas across the whole text, students in the NT condition were expected to reap the benefits of taking notes. The NT task was expected to allow students to increase their attention to text, forcing them to be formal and explicit about the information they chose to take notes on, providing an external record of their ideas that was available for their review, and resulting in better elaboration or organization of ideas.

Students in the EW treatment were also expected to outperform students in the RS condition on this measure. While both groups were anticipated to gain factual knowledge from simply reading the text, the EW treatment was expected to result in students' considering how ideas were related across text by encouraging deeper processing through comparing and contrasting, thus helping students remember more of the ideas. It must be noted though, that this hypothesis is attenuated by the possibility that even though EW was likely to facilitate remembrance of ideas, it may also have restricted such remembrance to a narrower set of specific ideas, whereas the RS group was expected to focus on information across the whole text. Still, even a small effect for EW, when

combined with the expected effect from NT, was predicted to result in CW outperforming RS in the overall comparison.

CW vs. RS: Hypothesis Two – Multiple Choice Inference Measure. Students in the CW treatment groups were predicted to significantly outperform students in the RS treatment group, as both the NT and EW groups were expected to each independently outperform the RS group. The inference measure required students to use information presented in text to make inferences about information not included in the text. Students in the NT condition were expected to outperform students in RS, because, although both the NT and RS tasks were designed to focus students' attention on surface ideas across text and not beyond the text, students in the NT condition were expected to have better recall and remembrance of the information in the text due to writing about it (see Hypothesis One). Therefore, it was predicted that students in the NT treatment would be better able to process and recall the factual information needed to make inferences more quickly and readily, allowing them to use more cognitive resources to identify the correct inference for each question.

Students in the EW condition were also expected to outperform students in the RS condition on the *Multiple Choice Inference* measure, as the EW condition was designed to require students to think about how ideas in text were related one to another, resulting in something more than plain remembrance of information. Although the writing task completed in the EW condition might have restricted students' thinking to a narrower set of ideas, students were expected to make more correct inferences on questions related to information manipulated during the task. Although neither group was expected to have an advantage on questions requiring information not manipulated by the writing group, the advantage of the EW condition on the narrower set of information was predicted to lead to higher scores on the assessment overall.

CW vs. RS: Hypothesis Three – Essay Application Measure. The hypothesis for the application measure was more complex because only one writing treatment group was expected to outperform the RS group. That is, the EW condition was expected to outperform RS, but no significant difference was expected between NT and RS, reducing the effect for the EW group when the treatments were combined. Nevertheless, I hypothesized that there would be a small, but statistically significant difference between CW and RS.

Students in the NT condition were not expected to perform significantly differently from students in the RS condition on the application task. Both conditions were designed to force students to focus on identifying ideas across the whole text with little analysis, organization, or deep processing. While NT was expected to facilitate remembrance of more facts than RS, neither treatment was a particularly good match for the elements of application.

Alternatively, students in the EW condition were expected to perform significantly better than students in the RS condition on the Application Essay. While the RS condition prompted students to pay attention to surface ideas, with the aim of increasing information recall, the EW writing task required students to compare and contrast ideas presented across the text, which was intended to involve deep processing, reorganization, and analysis of the ideas presented. Therefore, students in the EW treatment condition were expected to have a deeper understanding of the ideas that needed to be applied when completing the essay measure, leading to better performance on the essay assessment.

Research Question 2: Is note taking more effective than extended writing for improving the reading comprehension of fourth grade students after controlling for initial writing ability? For the second question, the differential effects of the two writing

treatments (NT and EW) were examined on the three reading comprehension measures. Hypotheses are presented below.

NT vs. EW: Hypothesis Four – Topic Knowledge Measure. Students in the NT condition were expected to perform significantly better than students in the EW condition on the measure of topic knowledge. The recall measure was designed to involve identifying facts and ideas across the text. It was anticipated this would lead to better performance on overall recall of information for the NT treatment, because NT was expected to focus students' attention on specific ideas across the text, whereas EW was designed to focus students' attention more narrowly on specific ideas involving deeper processing and organization.

NT vs. EW: Hypothesis Five – Multiple Choice Inference Measure. No prediction was made for the comparison of NT to EW for the inference measure. The EW condition was designed to help students think about how the ideas in the text relate one to another beyond plain remembrance of the information. Therefore, it was anticipated students in the EW condition would make more correct inferences on any questions related to information students attended to during the EW task. However, the writing task completed in the EW condition was expected restrict students' thinking to a narrower set of ideas, giving them an advantage on only questions related to those ideas (but this is not certain). Conversely, students in the NT condition were expected to have better recall and remembrance of information across the whole text, allowing them to use more cognitive resources to identify correct inferences for questions related to information they took notes about. Because this measure was designed to be independent of both the NT and EW treatments, it was difficult to determine whether the writing involved in either condition would provide an advantage over the other.

NT vs. EW: Hypothesis Six – Application Essay Measure. Students in the EW condition were expected to perform significantly better than students in the NT condition on the application essay. The application essay was designed to require students to apply

information about the text to a new situation, and the information applied to the assessment task involved the same content manipulated by students in the EW treatment. Moreover, the application essay required analysis, reorganization and deeper processing of information. Therefore, it was predicted that the EW treatment would lead to better performance on the application measure because it would cause students in the EW condition to analyze and reorganize information needed for the application measure, while NT was expected to focus on surface information.

Research Question 3: Does writing ability moderate the effects of writing the treatment conditions for questions one and two? The hypotheses for the results regarding this question are presented first in relation to research question one, and then for research question two. In each case, one hypothesis is postulated for all three measures, rather than for each measure individually.

Does writing ability moderate the effects of treatment for the comparison of CW to RS? A significant interaction between writing ability and treatment was expected. In this comparison, it was expected that weaker writers would benefit from CW as compared to RS. However, I hypothesized that stronger writers may also be stronger readers, and they would not necessarily need to use writing as a tool to augment their reading comprehension. Therefore, the effect for CW may be smaller for stronger writers than weaker writers. This was expected across all three measures of reading comprehension.

Does writing ability moderate the effects of treatment for the comparison of NT to EW? A significant writing ability by treatment interaction was also expected for the NT to EW comparison. The NT task involved writing words and short phrases instead of connected text, and the relationships between ideas could be organized by physical arrangement on the page, rather than through text descriptions. The EW task, on the other hand, required students to generate, integrate, evaluate, combine, and recombine ideas in connected text. For these reasons, it was hypothesized that the NT would be an

easier writing task than EW, allowing weaker writers to benefit more from NT than EW. On the other hand, it was expected that stronger writers would be better able to complete the EW task as intended. Although the stronger writers were also expected to complete the NT task without issue, it was expected that the stronger writers would benefit more from the deeper processing the EW task was expected to elicit. Therefore, it was predicted that the stronger writers would benefit more from EW than NT. The expectation that the stronger and weaker writers would benefit differently from different tasks led to the prediction that the interaction would be statistically significant.

CHAPTER II

METHOD

Participants

Participants for the study included students from 13 fourth grade classrooms across three schools from one school district in the south that served rural and suburban schools. All students in the fourth grade in these schools were eligible for participation in the study. Two-hundred and nine students were consented and assented for the study. Using a person-randomized, multi-site design, students were randomly assigned (within-classroom) to one of three treatment conditions: (a) reading and studying with no writing ($n = 69$), (b) note-taking ($n = 70$), and (c) extended writing ($n = 70$). During the course of the study, 12 students were lost to attrition based on lack of attendance, 4 students could not be included due to failure to pick up their posttests (these students were consented and assented after the study began, leading to a miscommunication due to change in their data collection status), and one student moved out of the district. Consequently, 192 students (88 boys, and 104 girls) completed the study: 61 in the NT group, 67 in the EW group, and 64 in the RS group.

The majority of students who completed the study were Caucasian, consistent with the populations in these schools ($n = 158$; 81.9%), and ranged in age from 9.51 to 11.56 years ($M = 10.26$; $SD = 0.38$). Twenty-six students (13.5%) received special education services. Fifty-four students (28.1%) received free or reduced lunch.

Demographic information summarized by treatment group can be found in Table 1. After randomization, categorical data were examined for potential relationships between the demographic variables and treatment groups using the chi-squared test for independence. A significant chi-squared value was found for the relationship between gender and treatment group, suggesting that a disproportionate number of boys and girls

were assigned to each condition ($\chi^2 = 7.09, p = .029$). Follow-up analyses of each cell's contribution to the chi-square statistic in a two-way table showed that the NT group had a disproportionate number of males (59%), while the EW group had a disproportionate number of females (64.2%). Chi square analyses contrasting Group X Race [$\chi^2 = 4.64, p = 0.79$] and Group X Special Education Status [$\chi^2 = 0.03, p = 0.99$], were not statistically significant.

Initial writing performance was measured using the third edition of the Wechsler Individual Achievement Test (WIAT-III, Breaux, 2010; this test is described later); the average standard score on the WIAT-III was 107.76 ($SD = 13.96$). A one-way ANOVA was used to analyze whether there were differences between the treatment groups on the pretest writing measure. No significant difference was found between the groups, $F(2, 189) = 0.45, p = .638$ (see Table 1).

Treatment Conditions and Random Assignment

Students were randomly assigned to one of three treatment conditions within-classrooms. This ensured that a nearly equal number of students from each classroom would be assigned to each treatment, thus limiting the influence of classroom level factors on the outcomes.

For each condition, the experimental tasks were demonstrated to students in the corresponding treatment groups by the researcher and two graduate student research assistants (collectively referred to as “instructors” for ease of explanation when describing implementation procedures) on Day 2 of the study. Students were then asked to read the experimental text and complete the treatment task on Day 3.

The treatment conditions were carefully designed to include similar elements, instructions, and examples wherever possible, so that the conditions only differed in the activity they were expected to complete after reading. Table 2 shows a comparison of the elements of instruction provided to each treatment group. The NT and RS groups were similar in almost every way, differing only in terms of writing, while the EW treatment

differed a bit more due to the nature of the task. These differences are more fully elaborated in the description of the treatments.

Read and Study (RS). Students in the RS condition were instructed to read a passage, and then study the important ideas. To ensure students knew what was expected, the instructors modeled an example of one way to think about and study text after reading. This occurred via a series of interactive think-alouds by the instructors, consisting of identifying important information, using single words and short phrases to represent the big ideas, and repeating the information to aid in memory. The instructor also provided students with general tips and exercise sheets for students to practice studying (see Appendix B). However, students were not told how to identify important information, nor were they told how to organize the information they studied in any systematic way. Instead, instructors emphasized that choices about what constituted important information, and how the notes should be organized, were up to the individual students. Instructors also told students that they could study the text any way they chose, as long as it did not involve writing. The script for the think aloud was designed to be a close representation of the example of written notes provided to the note-taking group (see description of the NT condition), differing only in that they were instructed to think about the ideas without writing.

On Day 3, students were asked to read the passage and study the important information. No writing implements or paper were provided during the experiment, and the instructors monitored the students to ensure that they did not write.

Note-taking (NT). Students assigned to the note-taking condition were instructed to take notes on the important information in the experimental text. To ensure that students knew what was expected, an instructor modeled paraphrasing main idea statements and details in note form on Day 2. The notes in the examples were written in single words and short phrases grouped together in unconnected text. In addition, the instructor provided students with general tips about note taking (see Appendix B).

However, students were not told how to identify the important information, nor were they told how to organize their notes (e.g., outlines or concept maps). Instead, the instructors emphasized that choices about what constituted important information, and how the notes should be organized, were up to the individual students.

On Day 3, students were asked to read the passage and take notes on the important information. The instructor provided with pencils and lined paper on which to take their notes.

Extended-Writing (EW). Students assigned to the EW condition wrote an essay comparing and contrasting two main concepts from the text. To ensure they knew what was expected, the instructors provided an example to the students on Day 2. The example included four paragraphs comparing and contrasting how penguins and people take care of their young, including an introduction, a paragraph about the similarities of how the two animals care of their young, a paragraph about the differences, and a concluding paragraph. The example also included words and phrases indicating whether a comparison or contrast was made (e.g., similarity, same, alike, different, difference, dissimilar), and these were highlighted during the instruction. Additionally, the instructor gave general tips about how to write a compare-contrast essay, and modeled some of the tips (see Appendix B). However, students were not given instruction on how to identify which information was relevant, nor how to organize or order the ideas they chose to write about. Instead, the instructors emphasized that choices about the ideas used for comparisons and contrasts, and how those ideas should be organized, were up to the individual students.

On Day 3, students were asked to read the passage and take notes on the important information. The instructor provided students with pencils and lined paper on which to write their essays.

Experimental Reading Passages

The reading passages used for this study were informational texts previously used by the National Assessment of Educational Progress (NAEP) to test the reading comprehension skills of fourth grade students (National Center for Education Statistics, 2012). These passages were chosen because they were considered to be grade level appropriate informational passages by the National Assessment Governing Board (National Center for Education Statistics, 2011). The first passage, “Daddy Day Care,” contained information about how penguins care for their young, and was used on Day 2 as the example passage for which each treatment group task was modeled. The second passage, “A Brick to Cuddle Up to,” provided information about strategies colonial Americans used to stay warm in the winter, and was used as the experimental passage.

Measures

Five measures were originally used in the experiment, two pretest measures, and three posttest measures. However, a problem occurred during pretesting that led to having to drop one of the pretests, namely the reading comprehension subtest of the Group Reading Assessment and Diagnostic Evaluation (GRADE, Williams, 2001). In one classroom, the test administrator gave the students only 20 minutes to complete the reading assessment, instead of 30 minutes. In addition, 30 minutes was not long enough for some of the other students to complete the test. Consequently, many students did not complete the reading test ($n = 44$, 23%), and the students did not all have the same amount of time to complete the test, compromising the validity. Therefore, scores from the reading test were not used in the analysis.

The writing pretest will be described, followed by the three outcome measures. Because the GRADE test was dropped, it will not be described here.

Pretest Measure: Wechsler Individual Achievement Test, 3rd Edition (WIAT-III). On Day 1 of the study, students were pretested for initial writing ability using the expository paragraph writing subtest of the Wechsler Individual Achievement Test, Third

Edition (WIAT-III, Breaux, 2010). The test was administered to all of the students in each of the 13 classrooms by the researcher and two researcher assistants. Students were given a pencil *without* an eraser, blank paper for planning, and lined paper with the following writing prompt: ‘*Write about your favorite game. Include at least 3 reasons why you like it.*’ Test administrators read aloud the instructions and prompt to the students. Students were given 10 minutes to write an essay response. If a student stopped writing before the time was up and they had written less than 30 words, the test administrator said, “Try to write more.”

The paragraph writing subtest of the WIAT-III was reported by its authors as reliable across grade levels for Theme Development and Text Organization ($r = .92$) and Word Count ($r = .99$), as well as test-retest stability for fourth grade (.86) (Breaux, 2010). In the current study, all of the compositions were scored by one graduate student researcher, with a random sample of the essays (33%) scored by the first author for reliability purposes. Interrater reliability of the scoring was .91 for the sample. Only the scores given by the graduate student researcher were used in the analyses.

Outcome Measures: Topic Knowledge, Multiple Choice Inference, & Application Essay. Outcome assessments were given on Day 4 of the experiment. Reading comprehension was assessed using three distinct measures. Two “treatment-inherent” measures were designed to be closely aligned with each of the two writing treatment conditions: 1) a *Topic Knowledge* measure designed to be aligned with the note-taking treatment, and 2) an *Application Essay* designed to be aligned with the extended writing treatment. The third measure, *Multiple Choice*, was designed to be “treatment-independent.” The measures were conceptualized and designed by adapting outcome measures used in previous studies examining similar writing task. The description of each measure includes details about the studies from which the measures were adapted and the alignment of measures to treatment conditions.

Topic Knowledge (Aligned with NT treatment). A measure of passage specific knowledge was adapted from Langer and Applebee (1987) and used to measure students' memory of factual information explicitly presented in the text. Students were asked to write free-association responses to four key topics from the passage: 1) The center of family life in the colonial home, 2) foot stoves, 3) bathing in colonial times, and 4) keeping warm at bedtime (see Appendix C). Students were instructed to write everything they could remember about each of the topics using single words, short phrases, or complete sentences. Before passing out the text, the instructor modeled an example response for the students using a topic unrelated to the passage (i.e., dogs). The instructor modeled responding to the "dogs" prompt by writing single words, short phrases and complete sentences, such as /animals/, /four legs/, /like to chase cats/, and /Dalmatians have spots./, emphasizing that students should write down everything that comes to mind. Ample space was provided between each of the topics to allow the students to provide as many associations as possible.

Prior to scoring, text related to each of the four response topics was reduced to independent facts introduced by the author of the passage. Each fact was then listed on a scoring sheet, by topic (see Appendix D). Next, each student response was parsed into propositions by two raters, and each proposition was compared with the propositions included on the scoring sheet.

Based on the comparison with propositions in the passage, each of the students' propositions were placed into one of the following categories adapted from scoring systems used by Hayes (1987) and Konopak, Martin and Martin (1990): a) text reproductions; b) incorrect information; or c) irrelevant information. *Text Reproductions* were defined as each instance of a match between a proposition in the reading and a proposition in the students' free associations, although they were not required to match verbatim. *Incorrect information* was defined as instances in which propositions provided untrue information about information directly presented in the text. *Irrelevant*

Information was broadly defined as information that was not directly referenced in the passage, regardless of whether the information was true, untrue, fact, or opinion. Each response was parsed and categorized by two raters, the first author and a graduate student research assistant. Interrater reliability for categorizing propositions was .93.

After parsing and categorizing the propositions for each response, two scores were created. First, each instance of a *text reproduction* was totaled across all four response categories to create a “total correct” score. The total correct scores of the two raters were then averaged. The purpose of the total correct score was to capture all of the information students remembered specifically from the passage.

However, some students’ responses included long lists of irrelevant and or incorrect information with only a few sporadic correct answers. In those instances, it appeared that some students may have “stumbled across a correct answer” when they responded without being precise about what they remembered. To contend with this, a proportion score was also calculated by dividing the number of *Text Reproductions* by the total number of propositions (i.e., *text reproductions* plus *incorrect* propositions and *irrelevant* propositions). The proportion scores for each rater were averaged.

The *Topic Knowledge* measure was considered “*treatment inherent*” because it was designed to align with the NT condition. Much like note-taking, the free association response allowed students to write short words or phrases, not requiring the ideas to be presented in connected text. Moreover, the four response topics required students to recall factual information across the whole text, which aligned with the goals of NT.

Application Essay (Aligned with EW treatment). The *Application* measure was designed to align with the EW treatment condition. Similar to the compare and contrast writing in the EW treatment, this essay measure required students to process the ideas presented in text, analyze how those ideas relate to a new situation, and reorganize and elaborate on those ideas in an extended response. Additionally, the question was

designed to elicit responses related to the ideas about staying warm that were compared and contrasted by students in the EW condition on Day 3.

The assessment required students to write an extended response to a question asking them to apply concepts presented in the text to a new situation. The question read:

“Imagine that it is a very cold winter. After a bad snowstorm, the electricity goes out in the whole city and it is going to take about a week to fix it. Because of that, you will have no heat in your house. Describe what you and your family could do to stay warm at home and elsewhere?”

Students were provided 20 minutes to construct their response to the question. The directions for the task were read aloud to the students and they were given an opportunity to ask questions about the directions. After students’ questions were answered, the question was read aloud to the students and they began writing.

The essays were scored on three dimensions: 1) application of the concepts presented in the reading; 2) elaboration on the ideas presented in the reading; and 3) coherence of the response (including how well the essay stays addressed and stayed on the topic of keeping warm). Rubrics were created for each of the three dimensions and student responses were scored holistically on a scale of 0-5 for all of the essays (see Appendix E for rubrics). Scores for the three dimensions were then summed to create a total score.

All of the essays were scored by a professor from another university who had previous experience with holistic scoring, but was unfamiliar with the design of the study and treatment conditions of the participants. Thirty-three percent of the papers were also randomly selected to be scored by the first author for reliability. Interrater reliability between the two raters for the *Application Essay* total score was .90. Only the scores from the primary rater were used in the analyses.

Multiple Choice Inference Measure (Treatment-independent). This measure included 15 multiple choice questions that required students to make inferences based on information provided in the reading passage (see Appendix F). Each item had four possible answers for students to choose from, consisting of one correct answer and three distractors. Each question was scored as either correct or incorrect, and the number of correct answers was summed to create a total score for the measure. A total score of 15 points was possible.

The multiple choice measure was considered to be independent of the treatments for two reasons. First, the items did not require written responses, which might have favored one or both of the writing treatments due to the mode of response. Second, the multiple choice items required students to make inferences from text using clues from the content presented, which did not align with any of the tasks. That is, students in the RS and NT conditions completed tasks requiring them to study or take notes on information presented in the text, but neither of the tasks required the students to think beyond the text. On the other hand, students in the EW group were asked to complete a task requiring them to examine how the ideas in the text related to one another. Although the EW task required deeper processing and reorganization of ideas presented in the text, it did not require students specifically to make inferences about information beyond the text.

The multiple choice measure was scored by a graduate student, with 30 percent of the items scored by the researcher for reliability purposes. As expected, reliability of scoring was high ($r = 0.97$), with only three errors found due to mistakes in coding. However, a problem was identified with the internal consistency of the measure. The Kuder-Richardson coefficient of reliability (KR20) was calculated for the measure at 0.36, suggesting the questions in the test may not all measure the intended construct (e.g., ability to make inferences), or may measure multiple constructs. This finding compromises the results for this measure, making interpretation difficult.

Study Implementation

The study included four sessions and took place over four consecutive school days in early May of 2012. The experimental procedures were conducted by the first author and two graduate students simultaneously in three classrooms at a time, in two of the schools. The third school had four classrooms, which led to dividing the students from one of the classrooms amongst the other three classrooms.

Day 1 - Pretesting. On the first day of the study, the students were provided an overview of the study schedule and procedures, sans details about differences in the treatment conditions. The instructors then a pretest the WIAT-III paragraph writing subtest to assess student writing ability. Students were provided 10 minutes to complete the writing test.

Day 2 occurred the next day, leaving no time for a make-up test prior to implementing the study. However, because the intervention occurred over four consecutive days, and growth on standardized measures was expected to be minimal over that time frame, a make-up day was included following the study for students who missed Day 1 (n = 6).

Day 2 – Examples and modeling for each of the treatment groups. The modeling and examples for each treatment condition occurred in separate classrooms, reducing the possibility of treatment contamination. Students were randomly assigned to treatment groups within classrooms, requiring them to be regrouped into appropriate treatment groups for Day 2. The purpose of Day 2 was to introduce students to the activity they were expected to employ after reading the text on Day 3.

The instructional assignments for the researcher and research assistants were counterbalanced across conditions and classrooms to control for potential teacher effects. More specifically, each research assistant was assigned to model the examples for all of the treatment conditions at least once, and one of the treatment conditions twice (assigned randomly). A written script was created for the instructor in each experimental condition

to follow, which included modeling and think aloud examples aimed at helping students understand the task they were expected to complete the following day.

During training, the instructors read the scripts word for word to become familiarized with the protocol for each treatment. However, short versions of the scripts were used during the actual intervention allow for the modeling to become more natural. For each condition, the purpose of the modeled task was discussed with the students, followed by an example of one way to complete the task. Students were given an opportunity to ask questions at the end. The research assistant explained that the script is only an example of how they might complete the task assigned to them, and that they could complete the task in any manner they chose on Day 3.

To limit the time between the task instructions for each group and the employment of the task by the students, Day 3 occurred the day immediately following Day 2. Therefore, no make-up day was available for students who missed Day 2, and those students were dropped from the study based on missing the critical instructions for the task.

Day 3 – Students complete their assigned writing or studying tasks. On Day 3, students were again grouped by treatment condition to complete the task assigned to them. For consistency within each classroom, the research assistants worked with the same students they modeled the task with on Day 2. Again, the treatment groups worked in separate classrooms to avoid potential treatment contamination.

Students in each treatment condition were given a sheet of paper with written instructions for the task they were expected to complete, which included the tips they were shown the previous day (see Appendix G for the NT example). The researcher and research assistants read the instructions and tips aloud to the students, and then instructed them to read the passage and complete their assigned task. Students in the NT and EW conditions were provided with writing materials (i.e., pencils, erasers, and paper) for

completing their tasks. Students in the RS condition were not provided with writing materials, and they were monitored to ensure they did not write during the session.

Students read the experimental passage titled “A Brick to Cuddle Up to.” They were told that they could ask the research assistant to read single words to them if they get stuck, but the research assistant could not read phrases or sentences to them. Student completed their NT, EW, or RS task immediately following the reading. Research assistants monitored the students to ensure they completed the assigned task, providing prompts for the students to keep working if they were off task. In the NT condition, the instructor prompted the students to take more notes or study their notes. In the EW condition, the instructor prompted the students to look for additional comparisons or contrasts and write more. For the RS condition, the instructor encouraged students to study or reread the passage.

Day 4 immediately followed Day 3 to ensure that the information is as fresh as possible in the minds of the students for the posttests. Therefore, there was no time for a make-up day for students who missed day three, and those were dropped from the study.

Day 4 - Posttests. Because there were three posttests, the order of the tests was counterbalanced to control for any potential order effects of the assessment. The researcher and research assistants were randomly assigned to give the assessment in six counterbalanced orders. Students were then randomly assigned, within each treatment condition, to take the assessments with one of the three instructors. They were then regrouped and changed classrooms so that the instructions for each of the test orders could be given to the entire group at once. Students were not given the opportunity to review the reading passage prior to taking the tests.

The instructions and items for the assessments were read aloud to the students to reduce the possibility of differences in the outcome due to students’ ability to read the test. For the *Topic Knowledge* measure, an interactive example was provided and completed orally as a class (see measure description earlier in the Method section). The

researcher then read the prompt for each item and provided the students with 15 minutes to complete the test.

For the *Multiple Choice Inference* measure, the researchers read each of the test items and the four possible answers for each item, repeating each question and answer before moving on to the next question. The multiple choice measure took about 10 minutes to complete.

For the *Application* essay, the researcher read the instructions aloud to the students. The researcher then read the question and provided students with 20 minutes to construct a response. The researcher repeated the question and directions to students as necessary.

Due to the schedules of the three schools, the only time a make-up could be scheduled was the following week (3-4 days after students read the experimental passage and completed the task assigned to their group). Therefore, it was decided that a make-up day would not be provided for students who were absent on Day 4 of the study, and these students were dropped from the study.

CHAPTER III

ANALYSIS

As a starting point for each of the analyses, an unconditional two-level mixed-effects model was examined to determine the portion of variance due to classroom differences, as compared to individual differences. The interclass correlation coefficients (ICCs) were calculated for each outcome measure based on the following model for student i in classroom j :

$$Y_{ij} = \beta_{0j} + \varepsilon_{ij}$$

$$\beta_{0j} = \gamma_{0j} + \delta_{ij}$$

where β_{0j} is the mean score of each school, and γ_{0j} is the grand mean. The models were estimated using Stata's `xtmixed` command, using the following syntax with the multiple choice measure used as an example:

```
xtmixed mc || teacher:, var
```

where *mc* was the multiple choice outcome. The ICCs calculated for each of the outcome measures indicated that two percent or less of the variance was attributable to classrooms for all of the outcome measures. This indicated that a multilevel analysis may not be necessary. However, Roberts (2007) cautioned against assuming no group dependence based on a small ICC, arguing that the degree of dependence may actually depend on the covariates included in the model. Therefore, I estimated the full model for each of the outcome measures, including all of the covariates and interactions chosen for the analyses, and then recalculated the ICCs. The ICCs dropped to zero in all of the models. Table 3 shows the ICCs calculated for each of the outcome measures in the unconditional model and fully defined models.

Furthermore, likelihood ratio tests comparing the multilevel models to simple linear regression models were statistically non-significant in each instance, indicating that

simple a simple linear regression was appropriate for all three outcomes. Therefore, single level regression analyses were conducted and reported in the results section.

Data Modifications

Data were examined prior to and during the analyses to be sure that the models met the regression assumptions. During this process, it was necessary to modify the data due to missing values and non-normal data patterns. The data and regression models were also examined for potential outliers.

Missing Data. Despite providing a make-up session for students who missed the pretest, I was unable to obtain pretest writing scores for two participants who participated in all other aspects of the study and completed all of the posttests. To avoid losing these participants to attrition through listwise deletion, values for their pretest writing scores were imputed using the *mi impute mvn* procedure in STATA/SE 11. The *mi impute mvn* employed a Markov Chain Monte Carlo (MCMC) method using data augmentation to generate missing values, assuming a multivariate normal model (StataCorp, 2009).

Ten imputations were created. Imputations created with *mi impute mvn* are usually used in conjunction with Stata's *mi estimate* function, which adjusts the coefficients and standard errors of the estimated models for the variability between imputation. However, *mi estimate* cannot be paired with the *xtmixed* command used for estimating multi-level models in Stata/SE 11. Therefore, an average of the ten imputations was calculated and substituted for the missing values in the two missing cases.

While the MCMC method assumes multivariate normality, the inferences made based on multiple imputations using MCMC are robust if the amounts of missing data are not large (Yuan, 1990). In this case, the amount of missing data imputed was only 1.03% of the pretest writing data, and less than 0.1% of the overall data used in the regression models.

Data Transformations

The assumption of normality was checked for each of the regression models prior to making inferences. The models for each of the outcome variables were constructed with all of the variables in their original metric. Heteroskedasticity was then examined using the Breusch-Pagan/Cook-Weisburg test. The models returned Chi-square values of 0.06 ($p = 0.80$), 0.03 ($p = 0.87$), 9.91 ($p = .002$), and 5.50 ($p = .02$) for the multiple-choice, essay, topic knowledge (total correct), and topic knowledge (adjusted) outcomes, respectively. These results indicated that heteroskedasticity was not a concern for the multiple choice and essay outcomes. However, there was significant heteroskedasticity in the models for both of the topic knowledge outcomes.

Further examination revealed scores for the TK_{TC} and TK_{ADJ} outcomes were not normally distributed. Box and Cox (1964) suggested that transformation of the dependent variable may be desirable for satisfying the assumptions of multiple regression, and to produce the simplest possible regression model. Further, fitting a linear model to transformed variables often leads to a clearer analysis than positing a non-linear model (Singer & Willett, 2003). Likelihood-ratio tests of Box-Cox regression models for both outcomes allowed for the rejection of the null hypothesis that no transformation was needed, TK_{TC} ($\chi^2 = 30.03$, $p < .001$) and TK_{ADJ} ($\chi^2 = 9.37$, $p = .002$). Examination of quantile-normal plots based on the ladder of powers indicated that taking the square root was the most appropriate transformation for both variables.

Following the transformations, recalculated Breusch-Pagan/Cook-Weisburg tests of heteroskedasticity were not statistically significant for the TK_{TC} ($\chi^2 = 0.05$, $p = .83$) or TK_{ADJ} ($\chi^2 = 0.47$, $p = .49$) measures, indicating that the transformations were successful in eliminating the heteroskedasticity in these models.

Potential Outliers

Casewise diagnostics were obtained to identify possible outliers in each of the regression models. Cases with standardized residuals more than two standard deviations

from their predicted scores were examined using DFFITS and DFBETAS statistics. Subjects with values outside of the acceptable ranges for these statistics were examined as potential outliers. Several potential outliers were identified in each model.

Visual methods of examination were also used, including the examination of boxplots and leverage versus squared-residuals plots. As an example, one participant not identified through Casewise diagnostics was identified with a value outside of two times the interquartile range on the boxplot and an unusually high leverage point on the leverage versus squared residuals plot. Closer examination of this subject revealed a writing pretest standard score of 55 (more than 3 standard deviations below the mean). However, despite having a very high leverage, this participant has a small squared-residual in each model, and was therefore not overly influential.

In all cases, the models were run with and without the potential outliers included. Elimination of the outliers did not result in significant changes to the models, nor interpretations of any of the results. Therefore, all potential outliers were included in each of the final models.

CHAPTER IV

RESULTS

Results are presented first for treatment integrity, and then for the outcome measures. All parametric statistical analyses were conducted with the variables transformed as previously specified.

Treatment Integrity

The instructional steps included in the modeling and examples provided on Day 2 of the interventions were examined for implementation fidelity. To determine what percentage of steps were applied as intended, all instructional sessions were tape recorded and reviewed by a graduate assistant who was not involved with the intervention and was blind to the hypotheses. The sessions included four lessons for each of the treatment groups, twelve lessons overall. Treatment fidelity was high, with more than 90% of the steps completed as intended in all three of the instructional conditions, including a mean score of 96.00% ($SD = 3.28$) for RS, 93.27% ($SD = 3.68$.) for NT, and 96.67% ($SD = 1.28$) for EW. A one-way ANOVA was used to analyze whether there were differences between the treatment conditions for percentage of steps completed. No statistically significant difference was found between the groups, $F(2, 11) = 1.50, p = .274$.

Regression Models for Outcome Measures

The results for the treatment comparisons are organized by outcome measure, with the research questions addressed for each measure. For the topic knowledge measure, two scores were created and analyzed as outcomes in separate models: 1) the total number of correct propositions, and 2) the proportion of correct answers to the total number of propositions. Thus, although there were only three outcome measures, four regression models were created to accommodate the two topic knowledge scores. The

means and standard deviations for the measures are presented for each of the treatment groups in Table 4.

All of the research questions were addressed using one single-level regression model per outcome, with the same independent variables used in the examination of each of the measures. First, students' pretest writing scores from the WIAT-III were included as a covariate in each of models, as more skilled writers were expected to perform better on the reading outcomes. Gender was also included as a covariate due to the disproportionate number of males and females in the NT and EW conditions (see Method section), coupled with the tendency of girls to be better writers than boys (Berninger & Fuller, 1992; Graham, 2006). To examine the effects of treatment, contrast coding was used to make orthogonal comparisons for Research Question 1 (the comparison of CW to RS) and Research Question 2 (the comparison of NT to EW). Two interaction terms were included in the model to examine potential heterogeneity of the effects of treatment across different levels of student writing ability (Research Question 3) for each of the comparisons: 1) Contrast 1 – [(CW versus RS) X WIAT-III], and 2) Contrast 2 – [(NT versus EW) X WIAT_III].

Multiple Choice Outcome. Results of the regression model for the multiple choice outcome can be found in Table 5, columns 2-4. All variables were entered simultaneously. The model results revealed that the variables explained 9% of the variance in the multiple choice outcome, $F(6, 185) = 2.94, p < .001$. Of the control variables, gender was not a statistically significant predictor in the model ($t = -0.79, p = .429$). However, student writing ability was a statistically significant predictor of scores on the multiple choice measure ($t = 3.36, p < .001$). The coefficient was 0.04, indicating a 10 point standard score increase on the writing pretest was associates with an increase of 0.4 questions answered correctly when controlling for gender, treatment group, and treatment by writing skill interactions.

As predicted, students in the two writing conditions outperformed students in the RS condition when controlling for initial writing ability and gender, resulting in a statistically significant main effect for treatment. The coefficient for the CW to RS contrast was significant ($t = 2.20, p = .029$) and positive ($B = 0.57$), indicating that students in CW scored an average of .57 points higher on the 15 question measure, or had 3.8% more correct answers, than students in the RS condition. This represents an effect size of 0.34 favoring the writing treatments. However, this finding is tempered by the low reliability score for the multiple choice measure. The coefficient for the interaction of writing ability with Contrast 1 was not statistically significant ($t = -1.14, p = .254$). Therefore, I cannot reject the null hypothesis that the slopes were homogeneous for the writing treatment groups when compared to the RS treatment. In other words, the positive effect of the writing was not significantly different across different levels of student ability.

There were no main effects for the second contrast included in the model. The coefficient for Contrast 2 (NT vs. EW) was not statistically significant ($t = 0.47, p = .636$). Therefore, I cannot reject the null hypothesis that there is no difference between the scores for these treatment groups. There were also no statistically significant interactions between the treatment comparisons and writing ability. The coefficient for the interaction of writing ability with Contrast 2 was not statistically significant ($t = 0.60, p = .548$), indicating I cannot reject the null hypothesis that the slopes are homogeneous for the NT and EW groups. In this case, there was not a statistically significant difference between the two writing treatments. Therefore, the lack of difference between the treatments would be expected across all levels of student ability.

Application Essay. Scores for the application essay reflect the sum of the scored on the application and elaboration rubrics, without the coherence score included. Results of the regression model for knowledge application can be found in columns 5-7 of Table 3. The variables included in the model explained 9% of the variance in the multiple

choice outcome, $F(6, 185) = 3.19, p = .005$. Gender was not a statistically significant predictor of the essay outcome ($t = 0.53, p = .594$). As expected, however, student writing ability was a statistically significant predictor of the essay scores ($t = 3.24, p = .001$). The coefficient was 0.04, indicating a 10 point standard score increase on the writing pretest was associated with an increase of 0.4 increase in the essay score when controlling for gender, treatment group, and treatment by writing skill interactions.

Contrary to predictions there were no statistically significant main effects for treatment on this outcome. We were unable to reject the null hypotheses for Question 1 that there were no differences in the scores between CW and the RS condition ($t = -0.56, p = .578$), and Question 2 that there were no differences between the scores of students in the NT and EW conditions ($t = 0.08, p = .993$). Additionally, there were no statistically significant differences for the variables included to examine whether there were interactions between the writing ability and Contrast 1 ($t = -0.27, p = 0.788$) or writing ability and Contrast 2 ($t = 1.63, p = 0.106$). In other words, the null hypothesis of similar slopes for each treatment across levels of writing ability cannot be rejected, meaning that any differences between the treatments, or lack thereof, are expected across all levels of student ability.

Topic Knowledge-Total Correct. As previously described, it was necessary to use the square-root transformed scores for this outcome measure to correct for heteroskedasticity. Results of the regression model for topic knowledge outcome can be found in columns 8-10 of Table 3. The model explained 19% of the variance in the topic knowledge outcome, $F(6, 185) = 7.38, p < .001$. Gender was not a statistically significant predictor of this outcome measure ($t = 1.14, p = .255$). Student writing ability was a statistically significant predictor of the essay scores ($t = 5.28, p < .001$). The coefficient for writing ability was 0.02, indicating a 10 point standard score increase on the writing pretest was associated with an increase of 0.2 increase in the number of

correct propositions included in the student responses to this measure when controlling for gender, treatment group, and treatment by writing skill interactions.

There were no statistically significant main effects for treatment on this outcome. Therefore, I was unable to reject the null hypotheses for Question 1 that there were no differences in the scores between CW and the RS condition ($t = -0.97$, $p = 0.335$), and Question 2 that there were no differences between the scores of students in the NT and EW conditions ($t = 1.15$, $p = .252$). The coefficients were also not statistically significant for interactions between writing ability and the treatments included in the first contrast ($t = 1.06$, $p = 0.290$) and writing ability and the treatment comparison included in the second contrast ($t = 0.86$, $p = 0.388$). That is, the lack of differences found in the contrast coded treatment comparisons would be expected across all levels of student writing ability.

Topic Knowledge-Proportion Correct. The regression model results for the proportion score of the topic knowledge outcome can be found in Table 3, columns 11-13. Contrary to predictions, the model did not explain a statistically significant amount of variance for this outcome, $F(6, 185) = 1.30$, $p = .260$. There were also no statistically significant predictors of the outcome variable included in the model. Therefore, the null hypotheses for the three research questions were not rejected for this measure.

CHAPTER V

DISCUSSION

Writing has been shown to improve learning (Bangert-Drowns et al., 2004; Graham & Perin, 2007) and reading comprehension outcomes for students (Graham & Hebert, 2010, 2011; Hebert, Graham, & Gillespie, 2012). Consequently, more attention is being paid to writing as an essential element of reading instruction (Duke, Pearson, Strachan, & Billman, 2011). However, not enough research has been conducted to tease out the nuances of how factors such as grade level, students' writing ability, and the measurement of reading comprehension may impact the effects of writing on reading for different writing tasks. In the present study, I examined whether note-taking and extended writing tasks were effective for improving the expository text comprehension of fourth grade students on three reading comprehension outcomes. I also examined whether the two writing tasks were differentially effective across measures, and whether the effectiveness of these writing tasks were moderated by student writing ability. The discussion centers on the research questions.

Research Question 1: Is writing more effective than reading and studying for improving the expository text comprehension of fourth graders?

Based on the comparison of the two writing treatments to the read and study treatment across three outcome measures in this study, the results were inconclusive. That is, questions about the multiple choice measure introduced doubts about the statistically significant finding for this measure, and some of the non-significant findings may also be a reflection of poor measurement.

First, the writing groups outperformed the read and study group on the multiple choice outcome measure, identifying more correct inferences. Inference is an especially important reading comprehension skill, and it could be argued that increasing students'

ability to think beyond the text is more important than recalling information. However, this finding was tempered by a low reliability score for the measure, which calls into question whether it is truly an assessment of students' ability to infer, or whether it might measure multiple constructs. This weakens the findings for this measure, as it is not clear what the measure captured.

Additionally, there were no significant differences between groups on the application essay. However, this measure may have suffered from a potential lack of sensitivity. One indication of this was that the scale for this test was 0-10 (based on the sum of the two essay dimensions scored), yet no students in the sample obtained a score of 9 or 10, essentially restricting the range of outcomes to an 8 point scale. This may reflect poor differentiation between essays at the top of the range. A second indication of the potential lack of sensitivity was that the measure suffered from potential floor effects. While the mean score for this outcome ($M = 2.97$) was slightly more than one standard deviation ($SD = 2.24$) higher than the lowest score of zero, 20 percent ($n = 38$) of students received a score of zero. Thus, the scale also appears to have poor differentiation at the bottom range of the measure. This lack of sensitivity leads to three potentially competing interpretations, obscuring the findings for the essay measure: 1) essay scoring was too stringent, lacking precision to discriminate true knowledge differences, 2) the essay prompt was not constructed so that it elicited responses that were representative of students' knowledge, or 3) the scores were a true representation of students ability to apply the knowledge they gained from this assessment, and students simply were not able to gain such knowledge. In other words, the lack of significant differences found on the essay outcome may indicate a true lack of differences between the treatment groups, or could simply reflect poor measurement.

Based on the results of the multiple choice and essay outcomes, there was no evidence to support that writing tasks improve the expository text comprehension of fourth grade students above and beyond reading and studying. This was further

supported by the results of the topic knowledge measure; a measure where validity issues were not readily apparent. Despite this, questions about the reliability and validity of two of the comprehension measures leave open the possibility that true differences may be found if students' comprehension is tested properly. Future research should be planned with an emphasis on improved comprehension measurement.

Is note-taking more effective than extended writing for improving the expository text comprehension of fourth grade students, after controlling for initial writing ability?

Based on the findings of this study, I cannot reject the null hypothesis that there was no difference between the two writing treatment groups for improving reading comprehension outcomes. There were four potential reasons for this: 1) lack of power, 2) questionable measurement, 3) students did not complete the writing tasks as anticipated, and/or 4) treatment contamination.

Lack of Power. First, the study may not have been sufficiently powered. With power set to .80, an alpha level of .05, and 128 participants, this comparison had the power to detect an effect size of $d = 0.36$ or greater. This difference was reasonable to expect based on effect sizes found for these two writing tasks in prior research (Graham & Hebert, 2010, 2011). However, the differences between the groups in this study were smaller, resulting in statistically non-significant effect sizes of 0.16, 0.24, 0.08 for the multiple choice, topic knowledge, and essay outcomes, respectively. This indicates that there may be potentially meaningful differences between these writing treatments, especially for topic knowledge, that could be explored with larger samples in future studies.

Questionable Measurement. Two of the measures may not be valid and, if this was the case, it may have limited the measures usefulness to capture true differences between the groups. Previously noted problems with the outcome assessments may have led to poor measurement of potential differences between the groups (see Discussion for

Question 1). Specifically, the low internal reliability coefficient for the multiple choice measure, as well as the restricted range and possible floor effects of the essay measure, challenge the validity of the findings for two of the outcomes. It is important to note, however, that there were no obvious concerns about the reliability and validity of the topic knowledge measure, and no statistically significant differences were found between the treatments on that measure, as well.

Students did complete the writing tasks as anticipated. Some students in the two writing groups did not complete the writing tasks as anticipated, which may have led to a less than perfect comparison of the treatments. Examination of the notes and essays written by the students during the experiment reveal that some deviations from writing tasks commonly recurred throughout the treatment groups. This may indicate that fourth grade students have not yet developed sufficient writing and/or reading skills to take advantage of these tasks to augment their reading comprehension without additional instruction.

Examination of students' writing artifacts completed during the experiment revealed characteristics of the students' representation of the tasks that may have compromised the effectiveness of the treatments. Not all students exhibited all of the characteristics, and some students exhibited none of them, but some of the characteristics were notable across the sample (examples of students' writing containing these characteristics are provided in Appendix H).

Common problematic characteristics of the note-taking responses included:

1. Notes were sparse, including little to none of the information from the passage
2. Notes represented only one aspect of the passage, ignoring complete sections entirely
3. Notes included superfluous information not included in the passage
4. Notes were random and unorganized, and sometimes resembled connected text

5. Notes included all or most of the information in the passage, but failed to distinguish important information
6. Students underlined or starred almost everything they wrote, failing to utilize this tool properly (when you emphasize everything, you emphasize nothing).

The first three characteristics on this list revealed that some students did not complete this writing task as anticipated. Although this does not necessarily mean these students did not engage in the thinking required for identifying and remembering the important information through writing (i.e., the writing may reflect fluency or mechanics issues rather than problems with ideation), it is possible that at least some of these students did not fully engage the text as planned. On the other hand, the last three characteristics on the list revealed that some of these fourth grade students had a difficult time distinguishing the important information from the unimportant information. This may have made note taking a problem, as students' unsystematic approach to taking notes may have lead them to remember information arbitrarily.

Problematic characteristics of students' compare and contrast responses included:

1. Inclusion of comparisons or contrasts beyond the scope of the prompt (e.g., including comparisons about where the bathroom was located)
2. Listing ways colonists stayed warm, making no comparisons or contrasts
3. Limited writing, with few to no comparisons or contrasts
4. Focus on only one aspect of the passage, ignoring other information that was relevant to the topic.
5. Improperly characterizing similarities as differences, and vice versa, illustrating a lack of understanding of how to carry out this writing task (e.g., stating that one difference is that we use blankets today, but colonists didn't).

Examples of The first four problematic characteristics noted above revealed that some students completing the extended writing tasks did not make comparisons or contrasts that allowed them to attend to or reorganize the ideas as anticipated. Characteristic number five, in contrast, revealed that some students did not have sufficient background knowledge to complete the assignment. This likely led to misunderstandings about the content and, in turn, poor comprehension of the text.

It may also be that the topic the students were asked to compare and contrast was too abstract for students at this grade level. That is, in many comparisons the students seemed to have difficulty articulating how people keep warm today, as evidence by statements such as “we just turn it to hot,” “we just put the heat on in our car,” and “their heat runs out, but ours lasts.” While these statements may be true, the information in them lacks specificity about how or why modern heating inventions work, illustrating that some of the concepts emphasized in the assessment task may have been too abstract for the students to make strong comparisons to the inventions and strategies of the colonists.

Treatment contamination. A fourth potential problem was that treatment contamination may have occurred in some instances. For example, some students in the note-taking group included comparisons and contrasts in their notes, much like the EW group. On the other hand, some students in the EW group simply listed facts about the passage, without making comparisons or contrasts, which more closely resembling the writing in the note-taking tasks. These observations revealed that the two writing treatments were not always executed as intended by all students, with some students applying procedures intended for use in the competing writing treatment.

Did writing ability moderate the effects of writing the treatment conditions for questions one and two?

The third purpose of this experiment was to examine potential interactions between the writing treatments and students’ writing ability. I assumed that students of

different writing abilities may be able to take more or less advantage of one type of writing over another, or that stronger writers may not benefit as much from writing as weaker writers due to their ability to comprehend higher level text (based on the correlation between reading and writing skill). These interactions were not realized in the current experiment, as all of the interactions between writing ability and the treatment comparisons were statistically non-significant in every model. There are two potential interpretations of these findings, both of which need to be approached judiciously.

One potential and obvious interpretation is that treatment effects of note-taking and extended writing tasks on reading outcome measures, or lack of effects for the majority of comparisons in this study, were not moderated by student writing ability. However, this interpretation is tenuous at best, due to previous concerns raised about the measures. It may also be that power to detect interaction effects was lacking in this study. Cohen, Cohen, West and Aiken (2003) indicated “the power to detect an interaction is reduced, relative to first order effects” (pp. 297). Because the study had the power to detect main effect sizes equal to or larger than $d = 0.29$ and $d = 0.36$ for questions 1 and 2, respectively, it may very well be that the study simply did not have the power to detect potential interaction effects.

A second possibility was that no moderator effects were found because the fourth grade students in this study, regardless of initial writing ability, did not write well enough to sufficiently differentiate the writing tasks. If the writing tasks were simply too difficult for the fourth grade students (of all ability levels) to complete effectively, then it does not stand to reason that there would be differential effects for tasks by students by measures. Although this point admittedly requires considerable supposition, it is important to accentuate that these interactions could potentially emerge as students become more skilled, or if they were provided more instruction in how to employ the writing tasks.

CHAPTER VI

CONCLUSION

The findings of this study were quite limited. While a significant difference was found between the two combined writing treatments and the control condition on the multiple choice assessment, this measure was found to have low internal consistency, limiting the value of this finding. Moreover, there were no other statistically significant findings in the study regarding treatment effects.

However, there are many aspects of this study which can provide valuable insight for designing future research in this area. These are best examined through the limitations of the study. These include problems with the outcome measures, students' difficulty completing the writing tasks as intended, and lack of power.

First, two of the outcome measures had questionable validity, albeit for different reasons (described in detail earlier). Improvement in the measure of reading comprehension is paramount for future research studies in this area. Lessons learned from this study should help to improve the design and scoring of such measures.

Second, the study lacked power to detect differences in treatments for effect sizes smaller than 0.36 and 0.29 for the two primary research questions. Although larger effect sizes have previously been found for these treatment comparisons, this study included students in an earlier grade level than in past research. The data from this study showed that smaller effects may indeed be evident for fourth grade, but there was not sufficient power to obtain statistical significance. This lack of power may also have led to difficulty identifying potential interaction effects in the study. Although smaller, these effects may still be practically significant and important to identify. Future research studies should be designed with smaller effects in mind.

Third, some students had difficulty completing the writing tasks as intended, which likely influenced the effectiveness of the writing approaches studied here. This is an especially important finding, as one of the goals of this study was to determine whether fourth grade students could take advantage of these tasks with minimal instruction. It appears that this was not the case. However, the lack of findings does not suggest that students would not benefit from more intensive instruction in this area, or that minimal instruction (as was applied here) would not be appropriate for students in higher grade levels. Future research on this topic conducted with students in fourth grade and earlier should almost certainly include an instructional component, while studies examining minimal instruction should be conducted with students in later grades and/or designed to look for smaller effect sizes.

A final limitation of this study was that it did not include student reading ability as a potential covariate. Reading ability would almost certainly have accounted for variability in reading outcomes, and it may potentially moderate the effects of the treatment tasks. Although problems with the pretest reading measure were only briefly touched on in the Method section, the intent was to include reading ability as a factor in the current study. Future research should control for reading ability and examine potential interactions between reading ability and the writing task comparisons.

APPENDIX A

TABLES 1 - 5

Table 1:

Demographic Information of Students by Treatment Condition

	Read and Study (n = 64)	Note Taking (n = 61)	Extended Writing (n = 67)	Total (n = 192)
Age				
Mean	10.23	10.29	10.25	10.26
SD	(0.36)	(0.38)	(0.40)	(0.38)
Gender				
Males	28 (43.8%)	36 (59.0%)	24 (35.8%)	88 (45.8%)
Females	36 (56.2%)	25 (41.0%)	43 (64.2%)	104 (54.2%)
Race				
White	53 (82.8%)	52 (85.2%)	52 (77.6%)	158 (81.9%)
Black	6 (9.4%)	5 (8.2%)	7 (10.4%)	18 (9.3%)
Asian	1 (1.6%)	0	2 (3.0%)	3 (1.6%)
Hispanic	3 (4.7%)	4 (6.6%)	5 (7.5%)	12 (6.2%)
Other	1 (1.6%)	0	0	1 (0.5%)
Unknown	0	0	1 (1.5%)	1 (0.5%)
Primary Language				
English	62 (96.9%)	59 (96.7%)	65 (97.0%)	186 (96.9%)
Spanish	1 (1.6%)	1 (1.6%)	2 (3.0%)	4 (2.1%)
Amharic	1 (1.6%)	0	0	1 (0.5%)
Unknown	0	1 (1.6%)	0	1 (0.5%)
Students with Disabilities				
Yes	9 (14.1%)	8 (13.1%)	9 (13.4%)	26 (13.5%)
No	55 (85.9%)	53 (86.9%)	58 (86.6%)	166 (86.5%)
Writing Pretest (WIAT-III)				
Mean	109.00	106.64	107.60	107.76
SD	(14.53)	(14.30)	(13.17)	(13.96)

Note. WIAT-III = Wechsler Individual Achievement Test, 3rd Edition, paragraph writing subtest.

Table 2:

Key Activities Completed in Instructional Steps for Each Treatment Condition

	Read and Study*	Note-Taking*	Compare and Contrast*
Step 1:	Introduction to Strategy	Introduction to Strategy	Introduction to Strategy
Step 2:	Read the passage	Read the passage	Read the passage
Step 3:	Introduce concept: Study in sections	Introduce concept: Take notes by section	Introduce concept: Read the prompt & look for relevant information by section
Step 4:	Model: Studying important information, ignoring unimportant	Model: Studying important information, ignoring unimportant	Model: Underlining relevant information, ignoring irrelevant
Step 5:	Demonstrate: Repeating information to emphasize	Demonstrate: Underlining or starring notes to emphasize	Demonstrate: Underlining relevant information
Step 6:	Student practice: Studying and repeating important information	Student practice: Taking notes and adding emphasis	Student practice: Underlining relevant information, ignoring unimportant
Step 7:	Student practice: Ignoring unimportant information	Student practice: Ignoring unimportant information	Instructor overview of other relevant information in the passage
Step 8:	Demonstrate: Studying silently (mouthing or whispering)	Demonstrate: Organizing notes	Essay example: Read and discuss
Step 9:	Student Practice: Silent studying	Student Practice: Taking and organizing notes	Identify parts of essay: Introduction, comparing, contrasting, conclusion Introduce: compare and contrast words
Step 10:	Review and practice all study tips	Review and practice all note taking tips	Practice writing comparing sentences
Step 11:	Introduce and practice: Using all of your time	Introduce and practice: Using all of your time	Practice writing contrasting sentences
Step 12:	Conclusion and Questions	Conclusion and Questions	Conclusion and Questions

Note. *Not all activities are listed for purposes of brevity.

Table 3:

Intra-class correlations for the unconditional and fully defined multilevel models

Outcome Measure	ICC (Unconditional Model)	ICC (Full Model)
Multiple Choice	0.01	0.000
Essay (Concept Application and Elaboration)	0.01	0.000
Topic Knowledge	0.02	0.000
Topic Knowledge (adjusted)	0.00	0.000
Topic Knowledge (average proportion correct)	0.00	0.000

Note. ICC = Intraclass Correlation Coefficient

Table 4:

Means and Standard Deviations for each Treatment Condition on Three Outcome Measures

Outcome Measure	Read & Study (n = 64)	Note-taking (n = 61)	Extended Writing (n = 67)
Multiple Choice	8.00 (2.12)	8.74 (2.41)	8.37 (2.18)
Essay	5.22 (2.07)	5.00 (2.39)	4.72 (2.24)
Topic Knowledge Total Correct	5.88 (3.37)	5.83 (3.61)	5.06 (2.81)
Proportion	0.51 (0.25)	0.53 (0.26)	0.46 (0.25)

Note. Scores for the essay measure are a sum of scores on the application and elaboration rubrics.

Table 5

Summary of Regression Analyses for the Effects of Treatment and Writing Skill on Four Reading Outcomes

Variable	Multiple Choice			Application Essay			Topic Knowledge					
							(Total Correct)			(Proportion)		
	<i>B</i>	<i>SE_B</i>	<i>t</i> (185)	<i>B</i>	<i>SE_B</i>	<i>t</i> (185)	<i>B</i>	<i>SE_B</i>	<i>t</i> (185)	<i>B</i>	<i>SE_B</i>	<i>t</i> (185)
Intercept (<i>B</i> ₀)	8.19	0.24	34.65**	4.58	0.23	19.51**	2.02	0.07	27.32**	0.48	0.03	18.14**
Gender ^d	-0.28	0.35	-0.79	0.18	0.35	0.53	0.13	0.11	1.14	-0.01	0.04	-0.14
Writing Ability ^a	0.04	0.01	3.36**	0.04	0.01	3.24**	0.02	0.003	5.28**	0.002	0.001	1.58
CW vs RS ^b	0.57	0.26	2.20*	-0.14	0.26	-0.56	-0.08	0.08	-0.97	-0.03	0.03	-1.01
NT vs. EW ^c	0.10	0.22	0.47	0.02	0.22	0.08	0.08	0.07	1.15	0.03	0.02	1.13
(CW vs. RS) X Writing Ability	-0.18	0.02	-1.14	-0.01	0.02	-0.27	0.01	0.003	1.06	0.002	0.002	1.17
(NT vs. EW) X Writing Ability	0.01	0.01	0.60	0.02	0.01	1.63	0.004	0.003	0.86	0.001	0.002	0.78
Model Fit	$R^2 = .09, R^2_{adj} = .06$			$R^2 = .09, R^2_{adj} = .06$			$R^2 = .19, R^2_{adj} = .17$			$R^2 = .04, R^2_{adj} = .01$		
Omnibus Test	$F(6, 185) = 2.94^{**}$			$F(6, 185) = 3.19^*$			$F(6, 185) = 7.38^{**}$			$F(6, 185) = 1.30_{ns}$		

Note. CW = Combined Writing Treatments. RS = Read and Study. NT = Note-Taking. EW = Extended Writing.

Due to contrast coding and the inclusion of Treatment by Writing Skill interaction terms, the regression coefficients for Writing Skill represent the gain in the outcome associated with an increase in pretest writing ability when controlling for gender and treatment.

^a Centered. ^b Contrast Coded Treatments: Note-taking = .5, Extended Writing = .5, Read and Study = -1. ^c Contrast Coded Treatments: Note-Taking = 1, Extended Writing = -1, Read and Study = 0.

* $p < .05$. ** $p < .001$.

APPENDIX B

STUDY TIPS AND EXERCISE SHEETS

Study Tips

STUDY TIP 1:

- **Study in sections**
 - **Look for headings that break up sections**
 - **Identify important information after each heading**

STUDY TIP 2:

- **Study important information**
- **Ignore unimportant information**
- **Use short phrases to remember big ideas**

Thinking Example:

male penguins:
tough, don't eat in winter, take care of eggs

STUDY TIP 3:

- **Repeat information more than once to help you remember it**

Thinking Example:

ice, southernmost Earth, Antarctica:
frigid, only penguins in winter

Repeat:

ice, southernmost Earth, Antarctica:
frigid, only penguins in winter

STUDY TIP 4:

- **Study silently (Mouth the words to yourself)**

Study these ideas in your head:

Dark, males loses half his weight,
female comes back, male eats,
chick hatches, parents take turns holding chick

STUDY TIP 5:

- **Use all of your time and study the information again**

Study Exercises

STUDY EXERCISE 1: Practice studying important information in short phrases and words and repeating it

- Study the important information in “Foothold for Family”

STUDY EXERCISE 2: Practice ignoring unimportant information

- Study the important information in “Warm up for Dads”
- Remember that there is information that is not important
- Ignore the unimportant information

STUDY EXERCISE 3: Practice studying silently

- Study the important information in “Snack Time”
- Mouth or whisper the important ideas to yourself

STUDY EXERCISE 4: Review all of the study tips

Thinking Example:

Water is closer, Adults leave chicks, Learn to swim and find food

Independence:

Note-Taking Tips

NOTE-TAKING TIP 1:

- Take Notes Section by Sections
 - Look for headings that break up sections
 - Identify important information after each heading
 - Write Notes

NOTE-TAKING TIP 2:

- Take notes on important information,
- Ignore unimportant information, and
- Use words or short phrases to remember big ideas.

NOTE-TAKING TIP 3:

- Underline or star the really important notes to make them stand out

NOTE-TAKING TIP 4:

- Use all of your time to examine if there is any important information you missed

Note-Taking Exercises

EXERCISE 1: Practice taking notes and emphasizing important information in “Foothold for Family”

EXERCISE 2: Practice ignoring unimportant information in the section “Warm up for Dads.”

EXERCISE 3: Organize the notes for the section “All for One”:

Unorganized Notes: Female is gone, always dark, males loses half his weight, Female comes back, male eats, chick hatches, they take turns holding chick

Female is Gone:

Female Comes Back:

<hr/> <hr/> <hr/> <hr/>	<hr/> <hr/> <hr/> <hr/>
-------------------------	-------------------------

EXERCISE 4 (Putting it all together): Organize and emphasize information in your own notes for the section “Snack Time”

EXERCISE 5: Review all of the note-taking tips

Compare and Contrast Tips

Example Writing Prompt:

Write multiple paragraphs to compare and contrast how penguins take care of their offspring with what you know about how humans take care of their offspring.

COMPARE AND CONTRAST TIP 1:

- Identify relevant information in each section
- Look for headings that break up sections

COMPARE AND CONTRAST TIP 2:

- Underline information that is relevant to my topic.
- Ignore information that is not relevant

COMPARE AND CONTRAST TIP 3:

- Practice ignoring information that is not relevant

Compare and Contrast Exercises

COMPARE AND CONTRAST EXERCISE 1:

You try it: Practice underlining relevant information and ignoring information that is not relevant

- Underline the relevant information in “Foothold for Family”

Example Essay:

Penguins Parents and Human Parents

Penguins and humans are very different types of animals, but they also have some similarities. One of the things they both have to do is take care of their young. In the next two paragraphs, you’ll find out the similarities and differences about how humans and penguins take care of their young.

Penguins have their chicks and take care of them in very different ways than people take care of their babies. One big difference is that mother penguins lay eggs, but mother humans give birth to live babies. Another way they differ is that penguins hold their eggs and baby chicks on their feet, while humans usually hold their babies in their arms. Speaking of holding their babies, the male penguin will keep the egg warm against his body for two months without stopping for anything, even food. On the other hand, human parents often put their babies down in cribs or seats, and they eat every day. One final contrast is that adult penguins leave their chicks to be on their own after 6 months, but human children are not old enough to be on their own until they are 18 years old!

Although there are many differences between how penguins and humans take care of their babies, there are also some similarities. One thing they have in common is that both penguins and humans have to keep their babies warm. Another comparison is that both types of animals feed their babies until the babies are old enough to feed themselves. Lastly, penguin parents and human parents are alike because both types of parents take turns holding their babies.

There are probably many more similarities and differences between people and penguins and how they take care of their babies. Can you think of any?

COMPARE AND CONTRAST EXERCISE 2:

- Identify the paragraph that compares penguins and humans, and the paragraph that contrasts penguins and humans.
- Circle the words in each paragraph that indicate whether the sentences are comparing or contrasting the topics.

Contrasting Words
Different
Differ
But
Yet
On the other hand
Contrast
However
Differently
Dissimilar
In a different way
Another way

Comparing Words
Similarities
Both
And
Have in common
Comparison
Alike
Similar
Similarly
Same
Likewise
In the same way

COMPARE AND CONTRAST EXERCISE 3: *Write a sentence contrasting two ideas*

Below you will find one sentence about female penguins and one sentence about male penguins. Read the sentences and then write your own sentence contrasting male and female penguins. Use one of the words or phrases from the “Contrasting Words List” to help you.

Sentence 1: *The female penguin lays the egg.*

Sentence 2: *The male penguin keeps the egg warm for two months.*

Your New Sentence:

COMPARE AND CONTRAST EXERCISE 4:

The next two sentences tell how penguins hold their chicks. Read the sentences, and then write a sentence comparing male and female penguins using one or more “Comparing Words”

Sentence 1: *The male penguin takes his turn holding the chick on his feet.*

Sentence 2: *The female penguin takes her turn holding the chick on her feet.*

Your New Sentence:

APPENDIX C

TOPIC KNOWLEDGE MEASURE

Write Down Everything You Know

Directions: For each of the following topics, please write down everything you can remember or think of from the reading passage. You can write your ideas as single words, short phrases, or complete sentences.

Center of Family Life in the Colonial Home

Foot Stoves

APPENDIX D

TOPIC KNOWLEDGE ANSWER KEY

Topic Knowledge Answer Key

Center of Family Life in the Colonial Home		
Fireplace or Hearth	Candlelight or light of the fire	Fireside activities
Fireplace was wide and high	Animal skins sealed drafty windows	Bathed in toasty space by the hearth
large fire	Blocked out the Daylight	Cooks; cooking; hours at the hearth
chimney was large	Living area was gloomy	Kettle; corn pudding; baking bread
gusts of cold air; cold air blew into the house	Circle of light at the hearth	Reading or Needlework
Families huddled close to fireplace for warmth	Bathroom was in the kitchen	

Foot Stoves		
Tin boxes; metal boxes	Held burning coals	Winter rides; travelers
Tucked under their blankets	Hot smoke puffed from holes/lid	Took to Sunday services; Meeting Houses Had no heat
Soothed freezing feet & legs		

Bathing in Colonial Times		
Bathroom was the kitchen	Partially filled a tub with cold water	Blanket draped from chairs
Toasty space by the hearth	warmed it with hot water	Privacy
Did not bathe as often	water heated in the fireplace	Blankets let fire's warmth surround the bather

Staying Warm at Bedtime		
Hot bricks/soapstones	Brick kept them warm as long as the heat lasted	Bed Warmer/metal pan - long wooden handle
Tucked into their beds	Brick → turned to cold stone	Pan held embers
Cuddled with bricks	Allowed fire to die down; covered w/ashes	Warmed bedding
Bricks heated in fireplace	Early risers dressed under covers	Had to wait for sheets to cool
Wrapped bricks in cloths		Animal skins sealed windows

APPENDIX E
APPLICATION ESSAY RUBRICS

Correct Application of Concepts (Passage Dependent)	
5	<p>Describes and applies at least 4 ideas presented in the passage, with the following information included:</p> <ul style="list-style-type: none"> • heating bricks or soapstones to heat the bed; • heating the sheets of the bed with a pan; • huddling together near the fire; • starting a fire in the fireplace or hearth; • using animal skins to block drafts; • blankets or furs to keep warm when outside or traveling; • footstoves to keep feet warm; • heating water for hot baths in the fireplace; • getting dressed under the covers <p>Refers to colonial times, history, or the reading passage.</p>
4	<p>Describes at least 4 ideas presented in the text, correctly applying at least 2 of them using the verbiage specified under a score of 6.</p> <p>Refers to colonial times, history, or the reading passage.</p>
3	<p>Describes at least 3 ideas presented in the text, correctly applying at least 1 of them using the verbiage specified under a score of 6.</p>
2	<p>States 2 or more ideas from the passage, but may use some incorrect verbiage or leave out important information about the usage.</p> <p>Incompletely or incorrectly describes the application of them.</p>
1	<p>States 1 or more ideas from the passage as a way to keep warm, or includes vocabulary from the passage related to the topic without providing complete information about usage.</p> <p>Does not apply any of the ideas. OR Mentions incorrect information about keeping warm in colonial times.</p>
0	<p>Does not mention any ideas presented in the passage ways to keep warm.</p> <p>AND</p> <p>Does not mention anything related to keeping warm.</p>

Elaboration on Ideas (Presented in the Text)	
5	<p><i>Introduces new applications</i> for elements described in the text (e.g., warming something by the fire that was not discussed in the passage).</p> <p>AND</p> <p><i>States</i> realistic or sensible reasons for wanting to use each of the particular strategies mentioned (from the reading passage) for keeping warm.</p> <p>AND</p> <p><i>Elaborates</i> appropriately on how or why the strategy is used.</p> <p>*Must both elaborate AND provide a reason on at least 1 of the concepts</p> <p>(These may or may not be indicated by the use of the word because or another indicator.)</p>
4	<p><i>Introduces new applications</i> for elements described in the text (e.g., warming something by the fire that was not discussed in the passage).</p> <p>AND</p> <p><i>States</i> realistic or sensible reasons for wanting to use each of the particular strategies mentioned (from the reading passage) for keeping warm.</p> <p>AND</p> <p><i>Elaborates</i> appropriately on how or why the strategy is used.</p> <p>*Elaborations and reasons do not have to be on the same concept</p> <p>(These may or may not be indicated by the use of the word because or another indicator.)</p>
3	<p>Completes 2 of the following (does not have to apply to the same strategy):</p> <p>Introduces new applications of elements described in the text (e.g., warming something by the fire that was not discussed in the passage).</p> <p>OR</p> <p>States realistic or sensible reasons for using <u>most</u> of the particular strategies they mention for keeping warm. <i>(Strategies must be related to the reading passage)</i></p> <p>OR</p> <p><i>Elaborates</i> appropriately on how or why the strategy is used.</p> <p>(These may or may not be indicated by the use of the word because or another indicator.)</p>
2	<p>States realistic or sensible reasons. <i>(Strategies must be related to the reading passage)</i></p> <p>OR</p> <p><i>Elaborates</i> appropriately on how or why the strategy is used.</p>
1	<p>Provides a reason or elaborates, but the reason or elaboration may be incomplete or is not reasonable or sensible.</p>
0	<p>Does not provides any reasons or elaborations for keeping warm.</p>

***Rule:**

If one item from the rubric score is not addressed, can still receive the higher score.

If more than one item is not addresses, must give lower score

Coherence of Ideas (Passage Independent)	
5	<ul style="list-style-type: none"> • Provides an introductory statement introducing what the paper is about. (e.g., thesis statement, topic sentence, previews) • Includes smooth transitions between ideas, reasons, and elaborations. • No grammatical mistakes, easy for the reader to understand • Proper use of any literary devices used (e.g. bullet points, parentheses, etc.) to convey thoughts. • Uses multiple paragraphs; the paragraphs discusses related ideas (stays on topic within the paragraph) • Stays on topic of staying warm without electricity • Includes a conclusion statement (e.g., wraps up the paper)
4	<ul style="list-style-type: none"> • Provides an introductory statement introducing what the paper is about (e.g., thesis statement, topic sentence, preview) • Includes smooth transitions most of the time to introduce ideas, reasons, and/or elaborations. • Few grammatical mistakes, does not hinder the reader from understanding the paper • Proper use of any literary devices used (e.g. bullet points, parentheses, etc.) to convey thoughts. • Stays on topic of staying warm without electricity • Uses multiple paragraphs; most of the paragraphs discusses related ideas (stays on topic within the paragraph) • Includes a concluding statement (e.g., wraps up the paper)
3	<ul style="list-style-type: none"> • Provides an introduction, although not clearly stated • Includes some transition words to introduce ideas, reasons, and/or elaborations • Some grammatical mistakes, potentially one instance the reader is confused • Mostly stays on topic of staying warm with logical transitions between ideas, but may include some related topics (e.g., discusses getting food in a snow storm, shoveling the sidewalk) • Multiple paragraphs are not used • Provides a concluding statement, although not clearly stated
2	<ul style="list-style-type: none"> • Provides an introduction OR conclusion, although not clearly stated • Addresses the topic of staying warm but discusses other ideas within the realm of cold, winter, snow storm(e.g., discusses getting food in a snow storm, shoveling the sidewalk) • Some grammatical mistakes, potentially 2 -3 instances the reader is confused, but overall, the reader comprehends the paper • Includes Few Transitions • Multiple paragraphs are not used
1	<ul style="list-style-type: none"> • Does not include introduction. • Addresses the topic of staying warm, but gets far from the topic at times (e.g., discusses playing video games in winter and how to get to the next level) • May not include transitions • Many grammatical mistakes, more than 3 instances the reader is confused; reader struggles with comprehension • Does not include Conclusion
0	<ul style="list-style-type: none"> • Ideas are disjointed • Does not include an introduction, thesis, or concluding statement • Does not transition from one idea to the next • Does not address topic of staying warm • Many grammatical mistakes, too difficult for the reader to understand

APPENDIX F

MULTIPLE CHOICE INFERENCE MEASURE

Multiple Choice Test

For "A Brick to Cuddle Up to"

Directions: Read each question and circle the BEST answer from the four choices provided.

1) Why did the colonists heat bricks to keep themselves warm instead of only heating the sheets?

- a) the small bricks were easier to fit in the bed
- b) the sheets might catch on fire
- c) the sheets were too hot when they heated them
- d) the bricks stayed warm longer than the sheets

2) What is the most likely reason that the colonists wrapped the bricks in cloths before they tucked them into their beds?

- a) it kept the brick warm
- b) the bricks were too hot to touch on their skin
- c) they did not want to get their sheets dirty
- d) it kept the bricks from breaking if they fell on the floor

3) Why did colonists have a difficult time staying warm in the winter?

- a) only the meeting house had adequate heating
- b) the cold air from the big chimney always blew the fires out
- c) their houses didn't hold the heat in very well
- d) the colonists did not have a difficult time staying warm

4) Which date could be included in the time period described in the passage?

- a) 1752
- b) 1852
- c) 1900
- d) 1952

5) Based on the passage, what is the best definition for the word *embers*?

- a) fiery pieces of coal or wood
- b) wood that the colonists used to build their fires
- c) metal pans that were used in the fire place
- d) hot smoke from the fire

6) Why did colonists take baths in the kitchen?

- a) so they could use the bath water to put out the fire afterward
- b) it was the warmest place in the house
- c) they didn't bathe as often as we do
- d) they kept their water there

7) Why was the fireplace considered the center of the colonial home?

- a) the colonists put their beds around it to keep warm
- b) it was where the family would meet before going to Sunday services
- c) it was in the middle of the house
- d) the colonists gathered around it for light and warmth

8) Why did the colonists put ashes over the fire to make it die down instead of putting it out with water?

- a) Because they didn't want to waste clean water
- b) So the light from the fire wouldn't keep them awake
- c) So they could use the embers to start the fire in the morning
- d) Because nobody was awake to take care of the fire

9) Why did early Americans bathe less than we do?

- a) setting up the bathtub was a difficult process
- b) the only bathroom was in the kitchen
- c) they didn't get as dirty as we do today
- d) they didn't want to waste their drinking water

10) The author wrote that modern ways of staying warm didn't take over until recently in some parts of the country. Which part of the country could the author be talking about?

- a) Big Cities
- b) Mountains
- c) Deserts
- d) Small Cities

11) Why did the colonists have to fill a tub partially with cold water instead of only using water heated from the fire place?

- a) the water heated in the fireplace was too hot
- b) they couldn't heat enough water in the fireplace to fill the tub
- c) if they heated too much water in the fireplace it might put the fire out
- d) it was safer to check for leaks in the tub using cold water

12) Why were animal skins used to block the drafts from the windows?

- a) they didn't have enough blankets to cover the windows
- b) the animal skins were nice for decorating the windows
- c) the animal skins blocked out the daylight
- d) animal skins were thick enough to keep out the cold air

13) If the colonists let the fire die down at night, why would an early riser hurry to the hearth?

- a) because the embers were still warm
- b) because they would get dressed under the covers
- c) because the hearth is where they would make breakfast
- d) because they needed to make sure the fire was still out

14) Why did colonists bring foot stoves with them when they traveled?

- a) they were often out in the cold open air
- b) they used the embers in the stove to rebuild their fires
- c) they needed to have something to cook with if they got stranded on long trips
- d) feet are the most important part of the body to keep warm

15) Why were the chimneys for fireplaces made to be so large in colonial homes?

- a) they only had large rocks to build with
- b) because the colonists needed to let some cold air into the house to cool it off
- c) so families would gather closer together around the fire
- d) so they could let smoke from the fire out of the house

APPENDIX G

NOTE-TAKING INSTRUCTIONS

Note-Taking Instructions:

- 1) Write your name at the top of the blank lined paper.
- 2) Read the following passage: “A Brick to Cuddle Up to”
- 3) After reading the passage, take notes about all of the important information in the passage using the lined paper.

Remember the note-taking tips:

- Take notes on important information
- Ignore unimportant information
- Write short phrases and words to remember big ideas
- Underline or star especially important notes to remember
- Organize your notes

- 4) Use all of your time

APPENDIX H
STUDENT WRITING EXAMPLES

Example of sparse notes

ID # 0298

EMA RKA

Nice warm bed

*heated up stone

keeping warm in e/sai here

*footstovis

Example of notes representing only one aspect of the passage

ID # 0342

MH RKH

A Brick to Cuddle Up To.

* Own hot brick.

* If you lived in colonial times.

* Hard winter in New World.

* At bedtime, they heated soapstones, (Bricks)
in the fireplace.

wrapped it in cloths, tucked them into thier
bed.

* Kept them warm at night for as long as
it lasts.

Example of notes including superfluous information not included in the passage

They had to light a fire to keep from freezing to death. The chimney was big but the problem was that cold air kept flowing in the chimney so it put out the fire alot. they had to huddle by the fire in order to survive a frost bite. or if they felt cold. It was very open when they had to huddle because it was like below 0° out there!

Thats why If you went out there you would have to bring... A heavy coat, water, scarf, fluffy boots, mittens, a snow hat, food or anything else you think you might need.

Because IF you DONT! you will... starve to death, die of hypothermia, be dehydrated + die, or freeze to death.

And there is almost no sound in antarctica with is why if you want to talk some people even bring heatsets!

SO If you want to go to ANTARCTICA you better bring: food, water, a heavy fluffy coat, earmuffs, a winter hat, a scarf, a flashlight, matches, tang other stuff you could think of! !!

Example of notes resembling connected text

(mH)

*They heated soapstones or bricks in the fireplace. They wrapped them in cloths and tucked them into their beds. It warmed them until the heat cools down. They rubbed their feet in a bed warmer. In the winter they bring blankets and animal skins. Near their feet were small tin boxes called foot stoves. The meeting houses had on heat of their own until the eighteenth hundreds. At home they had to huddle close to enough to the fireplace. Later early Americans didn't bathe as often as we do. When they did their bathroom was the kitchen. The household cooks spent hours at the hearth.

Example of notes failing to distinguish important information

The cold: hot brick, nose tingles.

Warm Beds: Winters were hard. clever to keep warm. Heated soapstones, or bricks. Wrapped bricks. Didn't last that long. Before going to bed they rubbed icy sheets with a bed warmer (Metal pan) held hot embers from fireplace warmed the bedding, covered with skins and blankets. under feet (foot stove) held burning coals. Hot smoke puffed thru small holes in the small lid. Smoothing feet and legs. They brought foot stoves, skins, and blankets to Sunday service. Sunday service didn't have any heat until the 1800's. At home colonial family huddled close together. The fireplace was wide and high enough to hold a huge fire. But its chimney was large so if the fire was out all the cold air would come inside the house. Reading and other stuff was done by a candle. household spent hours on the hearth. They cooked and did other activities with the hearth

Example of notes in which student excessively underlined or starred information

A Nice Warm Bed.

- *Heated soapstone or bricks for beds.
- *Metal pan with a large wooden handle held members.
- So hot they had to wait to get in.

Keeping Warm Elsewhere

- *Covering themselves with animal furs or blankets.
- *Houses had no heat until 1800s.
- They had foot stoves at the end of their beds to keep their feet warm.

The center of Family life

- *had no lights except candlelight or fire.
- hot water cools cold water when they bathe.
- *Fire place was the center of the colonial home.

A Sleeping Fire

- *The fire could snuff out at bedtime.
- In the morning brick became cold stove.

Catching up to modern times

- *Most parts of countries modern ways didn't take over.
- *you having a brick to cuddle up to.

Example of compare and contrast writing that included comparisons or contrasts
beyond the scope of the prompt

People in Colonial America used hot bricks to be warm, but we don't use hot bricks, we use blankets. We use heating, they used a bed warmer. Another way they are different is that their bathroom was in the kitchen, but ours isn't. There are many more differences, but I can't think of any more.

Example of compare and contrast writing with no comparisons or contrasts

You climb into bed and find the toasty treat you have been waiting for — your very own.....hot brick!!

At bedtime, they heated soapstones, or bricks, in the fireplace. They wrapped the bricks in clothes and tucked them into their beds. The brick kept them warm at night, at least for as long as its heat lasted.

Staying warm wasn't just a bedtime problem, hot smoke puffed from small holes in the stove's lid, soothing freezing feet and legs, and the meeting house had no heat until the 1800s.

Reading or needlework was done by candle work, or by the light of the fire!!!

Example of compare and contrast with limited writing

One way that ^{were} different back in 1800s they
bring a hat back to bed and we have
AC.

Example of compare and contrast with focus on only one aspect of the passage

In the story they had a bed to get in and today we do to. The thing that is different is that today we don't have to make a hot brick. But today we sleep on a bed that's not on the bed. We both have fireplaces and pillows. We don't have to sleep on the floor in a sleeping bag.

Example of compare and contrast with improperly characterized similarities and differences

In this story I am comparing and contrasting how people stay warm now, and how they stayed warm back then. It's really weird. Back then they used a brick to cuddle up to. They didn't have nice blankets like we do now.

They think a brick is comfortable. They put a brick in clothes then tucked it in their bed. However we have nice blankets that keep us warm. A brick is very hard to us. But it's nice to them. Some take animal skin and blankets to cuddle up to. But they were tucked under their feet. However we have them on our feet and pretty much everywhere else.

Their brick was put in a fire which made it warm. Both now and back then used blankets. But we use them more. There were other people back then who used blankets. But some people used bricks. It would hurt a lot of people if they put

REFERENCES

- Bangert-Drowns, R. L., Hurley, M. M., & Wilkinson, B. (2004). The effects of school-based Writing-to-Learn interventions on academic achievement: A meta-analysis. *Review of Educational Research, 74*, 29-58.
- Berninger, V., & Fuller, F. (1992). Gender differences in orthographic, verbal, and compositional fluency; Implications for assessing writing disabilities in primary grade children. *Journal of School Psychology, 30*, 363-382.
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 211-252.
- Breaux, K. C. (2010). *Wechsler Individual Achievement Test – Third Edition (WIATT-III): Technical Manual*. Bloomington, MN: Pearson.
- Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences* (Third ed.). Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Chall, J.S. (1983). *Stages of reading development*, New York: McGraw-Hill.
- Chall, J. S. (1996). *Stages of reading development* (2nd ed.). Fort Worth, Tex.: Harcourt Brace.
- Duke, N. K., Pearson, P. D., Strachan, S. L., & Billman, A. K. (2011). Essential elements of fostering and teaching reading comprehension. In A. E. Farstrup & S. J. Samuels (Eds.), *What Research Has to Say about Reading Instruction (4th ed.)*. Newark, DE: International Reading Association.
- Graham, S. (2006). Writing. In P. Alexander & P. Winne (Eds.), *Handbook of Educational Psychology*. Mahwah, NJ: Erlbaum.
- Graham, S., & Hebert, M. (2011). Writing-to-read: A meta-analysis of the impact of writing and writing instruction on reading. *Harvard Educational Review, 84*, 710-744, 784-785.
- Graham, S., & Hebert, M. (2010). *Writing to read: An evidence base for how writing can improve reading. A Time to Act Report*. Washington, DC: Alliance for Excellent Education.
- Graham, S., & Perin, D. (2007). *Writing next: Effective strategies to improve writing of adolescents in middle and high school. A report to Carnegie Corporation of New York*. Alliance for Excellent Education.

- Hebert, M., Gillespie, A., & Graham, S. (2012). Comparing effects of different writing activities on reading comprehension: A meta-analysis. *Reading and Writing: An interdisciplinary journal*. Retrieved July 12, 2012, from <http://www.springerlink.com.proxy.library.vanderbilt.edu/content/m221551722h4673r/>
- Hayes, D. A. (1987). The potential for directing study in combined reading and writing activity. *Journal of Reading Behavior*, 19, 333-352.
- Kiewra, K. A. (1989). A review of note-taking: The encoding-storage paradigm and beyond. *Educational Psychology Review*, 1, 147-172.
- Klein, P. (1999). Reopening inquiry into cognitive processes in writing-to-learn. *Educational Psychology Review*, 11, 203-270.
- Konopak, B. C., Martin, S. H., & Martin, M. A. (1990). Using a writing strategy to enhance sixth-grade students' comprehension of content material. *Journal of Reading Behavior*, 22, 19-37.
- Langer, J., & Applebee, A. (1987). *How writing shapes thinking: A study of teaching and learning*. Urbana, IL: National Council of Teachers of English.
- National Center for Educational Statistics (2009). *Reading 2009: National assessment of educational progress at grades 4 and 8*. Washington, DC: Institute of Educational Sciences.
- National Center for Educational Statistics (2010). *Grade 12 reading and math 2009 national and pilot state results*. Washington, DC: Institute of Educational Sciences.
- National Center for Education Statistics. (2011). What does the NAEP reading assessment measure? Retrieved March 4, 2012, from <http://nces.ed.gov/nationsreportcard/reading/whatmeasure.asp>
- National Center for Education Statistics. (2012). NAEP Question Tool. Retrieved February 23, 2012, from <http://nces.ed.gov/nationsreportcard/itmrlsx/search.aspx?subject=reading>
- Newell, G. E., & Winograd, P. (1989). The effects of writing on learning from expository text. *Written Communication*, 6, 196-217.
- Newell, G. E. (2007). Writing to learn: How alternative theories of school writing account for student performance. In C. MacArthur, S. Graham & J. Fitzgerald (Eds.), *Handbook of Writing Research* (pp. 235-247). New York: The Guilford Press.

- Roberts, J. K. (2007). *Group dependency in the presence of small intraclass correlation coefficients: An argument in favor of not interpreting the ICC*. Paper presented at the annual meeting of the American Educational Research Association.
- Singer, J. D., & Willet, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrence*. New York: Oxford University Press, Inc.
- Slavin, R. E. (2008a). What works? Issues in synthesizing educational program evaluations. *Educational Researcher*, 37, 5-14.
- Slavin, R. E. (2008b). Evidence-based reform in education: Which evidence counts? *Educational Researcher*, 37, 47-50.
- StataCorp. (2009). *Stata Multiple-Imputation Reference Manual: Release 11*. College Station, TX: Stata Press.
- Williams, K. T. (2001). *Technical Manual: Group Reading Assessment and Diagnostic Evaluation*. Circle Pines, MN.
- Yuan, Y. C. (1990). Multiple Imputation for Missing Data: Concepts and New Developments. Paper 267-65. Rockville, MD: SAS Institute Inc.