


Spring June 22nd, 2010

The Effects of Reflective Thinking on Middle School Students' Academic Achievement and Perceptions of Related Instructional Practices: A Mixed Methods Study

David Denton
Seattle Pacific University

Follow this and additional works at: <https://digitalcommons.spu.edu/etd>

 Part of the [Curriculum and Instruction Commons](#), [Educational Assessment, Evaluation, and Research Commons](#), and the [Junior High, Intermediate, Middle School Education and Teaching Commons](#)

Recommended Citation

Denton, David, "The Effects of Reflective Thinking on Middle School Students' Academic Achievement and Perceptions of Related Instructional Practices: A Mixed Methods Study" (2010). *Theses and Dissertations*. 13.
<https://digitalcommons.spu.edu/etd/13>

This Dissertation is brought to you for free and open access by Digital Commons @ SPU. It has been accepted for inclusion in Theses and Dissertations by an authorized administrator of Digital Commons @ SPU.

The Effects of Reflective Thinking on Middle School Students' Academic Achievement
and Perceptions of Related Instructional Practices: A Mixed Methods Study

by

David W. Denton

Seattle Pacific University

2010

The Effects of Reflective Thinking on Middle School Students' Academic Achievement
and Perceptions of Related Instructional Practices: A Mixed Methods Study

by

David W. Denton

A dissertation submitted in partial fulfillment of the requirements for the degree of

Doctor of Education

Seattle Pacific University

2010

Approved by _____

Arthur K. Ellis, Ed.D., Professor, Chairperson

Andrew T. Lumpe, Ph.D., Professor, Committee Member

James M. Devlin, Ph.D., Professor, Committee Member

Program authorized to offer degree

Date

Rick Eigenbrood, Dean, School of Education

Copyright

by

David W. Denton

2010

All Rights Reserved

For Kari, Gretta, James, and Ruby

With thanks, and mindful that we are His workmanship, created in Christ Jesus for good works, which God prepared beforehand, that we should walk in them - Ephesians 2:10

Acknowledgements

I would like to thank my wife, Kari, for her patience and support throughout the last four years while completing the doctoral program at Seattle Pacific University. I would not have finished without her help. Thanks to my children, Gretta, James, and Ruby, for their steadfast love. I have fond memories of studying and writing papers between bath time, playing tag, and reading books together.

I am also grateful to the professors at Seattle Pacific University. I learned much more than I thought possible because of their expertise. I am especially indebted to Dr. Andrew T. Lumpe and Dr. James M. Devlin, readers of this dissertation. Their knowledge and insight significantly improved numerous parts of my study.

My gratitude is also extended to Dr. Arthur K. Ellis, the chairperson for this dissertation. Dr. Ellis inspired and challenged me throughout my time at Seattle Pacific. He generously shared his wealth of experience and wisdom so that I could finish this work. Moreover, Dr. Ellis provided a number of additional opportunities to write and collaborate with him. The time and energy that he exerted on my behalf are an indication of the positive impact he has on the field of education. Truly, classrooms are better places in which to teach and learn because of his efforts.

Finally, I thank Lauryann Evans, Scott Mauk, and Karen Soine. I cannot imagine completing a doctorate without good friends like these. Their kindness and generosity humbles me. I look forward to meeting again soon to reflect together upon the technical and practical aspects of our academic pursuits.

Table of Contents

List of Figures	iii
List of Tables	iv
List of Appendices	v
Abstract	2
Chapter One: Introduction	3
Purpose of the Study	3
Research Problem.....	5
Summary	12
Research Questions and Hypotheses.....	13
Content of the Following Chapters	15
Chapter Two: Literature Review	16
Purpose.....	16
Criteria for Literature Inclusion	17
Characteristics of Reflective Thinking.....	20
Theoretical Basis of Reflective Thinking	29
Empirical Basis of Instructional Practices Associated with Reflection.....	38
Further Research Involving Reflective Thinking.....	50
Chapter Three: Methodology.....	52
Chapter Overview	52
Research Design.....	53
Quantitative Phase.....	63

Qualitative Phase.....	76
Additional Description of Study Conditions.....	82
Summary	85
Chapter Four: Results	87
Chapter Overview	87
Results for the First Research Hypothesis	87
Results for the Second Research Hypothesis	93
Results for the Third Research Hypothesis	97
Summary	103
Chapter Five: Discussion	105
Chapter Overview	105
First Research Question	105
Second Research Question.....	108
Third Research Question.....	113
Summary	118
Limitations	120
Additional Questions.....	123
References.....	125
Appendices.....	140

List of Figures

Figure 1: Alignment of Curriculum, Instruction, and Assessment	55
Figure 2: Sequential Explanatory Mixed Methods Design.....	57
Figure 3: Summary of Differences Between Planned Contrasts	92

List of Tables

Table 1: Matched Pairs of Classrooms Forming Planned Contrasts.....	65
Table 2: Teacher Assignments for Each Designation and Class Start Times.....	66
Table 3: Testing Sequence and Descriptive Data	67
Table 4: Descriptive Statistics for the Content-Specific Test and Gain Scores.....	70
Table 5: Follow-Up Questions and Student Perspective	82
Table 6: Contrasts of Content-Specific Test Means	91

List of Appendices

Appendix A.....	140
Appendix B.....	151

Abstract

The Effects of Reflective Thinking on Middle School Students' Academic Achievement
and Perceptions of Related Instructional Practices: A Mixed Methods Study

by

David W. Denton

Seattle Pacific University

Chairperson: Arthur K. Ellis

The problem investigated in this study was whether instructional practices, which characterize reflective thinking, aided in the alignment of curriculum, instruction, and assessment at the classroom level. Participants included 259 seventh grade students, 126 female and 133 male, from 10 intact social studies classrooms. The investigator used an explanatory mixed methods design, which included a quantitative and qualitative phase.

The first phase involved testing the effects of an intervention, administered for 23 days, on students' academic achievement according to a content-specific test. In addition, this phase involved a questionnaire used to assess student perceptions of the intervention. Results of the content-specific test and student questionnaire were contrasted between four treatment, four comparison, and two control classrooms. In the second phase, 12 students participated in interviews using standardized open-ended interview procedures.

Overall results showed nonsignificant differences between treatment and comparison classrooms. However, analysis of the content-specific posttest and students' perceptions of teacher feedback showed a significant correlation at the .008 level, $r(254) = .29, p < .001$. Likewise, interview participants emphasized teacher feedback as an aid to learning and deemphasized lesson objectives as reflective writing prompts.

Chapter One: Introduction

Purpose of the Study

Searching the Education Resources Information Center database using the keyword *reflection* produces more than 750 results. Likewise, searching the same database for the phrase *reflective thinking* produces 89 results. Additionally, professional organizations such as the National Board for Professional Teaching Standards (2002), the National Council for the Social Studies (2008), and the National Foundation for the Improvement of Education (1996) have suggested that reflection is an activity that promotes learning for teachers and students alike. For instance, the National Council for the Social Studies has identified reflective writing as a method for assessing student thinking.

Theorists and researchers have defined reflection and reflective thinking in different ways (Grossman, 2009; Kompf & Bond, 1995; Rodgers, 2002). In addition, much of the theory and research involving reflective thinking has focused on teacher education. For example, Spalding and Wilson (2002) conducted case studies on changes in preservice teachers' attitudes through reflective writing. Likewise, Rich and Hannafin (2009) discussed connections between teacher reflection and video recording lessons.

Furthermore, theorists and researchers have focused less attention on reflective thinking in the context of primary and secondary education. Nevertheless, some literature has described the use of reflection in these environments. For instance, Trudeau and Harle (2006) described reflective questioning as an instructional practice for teaching kindergarten students. Likewise, Barell (1984) described characteristics of reflective

thinking and then suggested ways for integrating it as an instructional practice with gifted students.

An obstacle for theorists and researchers has been establishing clear definitions of reflective thinking and reflection in the context of primary and secondary education. This has been less of a problem with regard to teacher education and reflective practice.

Researchers dealing with this topic have had ample theoretical constructs from which to choose. For instance, Schön (1983, 1987), Brookfield (1995), and Shulman (1987) each proposed models of teacher-reflection that researchers have often referenced.

John Dewey (1997a) wrote an important definition of reflective thought. Dewey defined reflective thinking as “active, persistent, and careful consideration of any belief or supposed form of knowledge in light of the grounds that support it, and the further conclusions to which it tends” (p. 6). Likewise, Dewey proposed steps to reflection, which suggests further description of the reflective process. Dewey described these steps from the perspective of the learner and they include (a) perceiving a problem or difficulty, (b) developing a solution through rational thinking, and (c) monitoring the solution for its suitability.

A number of researchers have utilized Dewey’s (1997a) definition of reflective thinking as the theoretical basis for their investigations (King & Kitchener, 2004; Rodgers, 2002; Spalding & Wilson, 2002). However, one purpose of this study was to identify characteristics of reflective thinking appropriate for secondary education according to current research on the topic. The reason for doing this was to connect these characteristics to teaching practices for use as an intervention. The empirical portion of

this study tested the effects of this intervention on seventh grade social studies students' academic achievement and perceptions of classroom instruction.

Research Problem

The central problem of this study was whether reflective thinking promotes alignment between curriculum, instruction, and assessment in the context of seventh grade social studies classroom learning. A related question in light of these purposes and problem was whether instructional practices characteristic of reflective thinking produce alignment between these three elements resulting in an increase in student achievement according to a content-specific test and changes in students' perceptions of instruction according to a questionnaire.

Resnick, Rothman, Slattery, and Vranek (2004) defined alignment as the accuracy with which educational elements in a system work together to guide instruction and produce student learning. Generally, researchers have suggested that these three components are misaligned at various educational levels, such as state, district, and classroom (Browder, Spooner, Wakeman, Trela, & Baker, 2006; Parke & Lane, 2008; Pellegrino, 2006; Porter & Smithson, 2001; Resnick et al.; Tindal & Nolet, 1996). In this case, misalignment refers to the gap between the intended curriculum and the enacted curriculum (Porter & Smithson). The enacted curriculum consists of the actual curriculum that students encounter and learn in the classroom. Alternatively, the intended curriculum consists of standards and objectives, as defined by state, district, and professional organizations, that teachers are responsible for teaching (Porter & Smithson).

Recently, the problem of alignment has received more attention because of the standards movement and No Child Left Behind legislation (Browder et al., 2006; Resnick

et al., 2004; Roach, Niebling, & Kurz, 2008). Consequently, most of the theory and research in this area has involved aligning teacher activity to state and district standards and objectives. However, some studies have researched the problem of alignment at the classroom level. For example, studies by Tindal and Nolet (1996) and Porter, Kirst, Osthoff, Smithson, and Schneider (1993) examined the alignment of curriculum, instruction, and assessment in secondary math and science classrooms.

Researchers have proposed various reasons for the misalignment of curriculum, instruction, and assessment. For instance, Harniss (2006) suggested that misalignment stems from inappropriate or unclear learning objectives, ineffective instructional methods, and invalid assessments. Additionally, the National Research Council (Pellegrino, Chudowsky, & Glaser, 2001) and Pellegrino (2006) stated that misalignment occurs because of fragmented decision making, which dilutes efforts at improving instruction in classrooms. These sources also suggested that misalignment is a consequence of inadequate teaching and learning theories that link curriculum, instruction, and assessment together into a cohesive whole. Along these same lines, Roach et al. (2008) stated that the problem of alignment stems from curriculums that tend toward excessive content coverage, thereby limiting students' depth of understanding and causing poor performance on assessments. Additional research has supported the assertion; namely, that excessive content coverage and misalignment coincide (Goodlad, 1984; Porter et al., 1993; Schmidt, McKnight, Cogan, Jakwerth, & Houang, 1999).

Although analyzing the alignment problem in relation to the standards movement is not a central focus of this study, research in this area has identified important obstacles that inform an understanding of misalignment within classrooms. For example, in an

analysis of states' efforts to align curriculum standards with their respective tests, Resnick et al. (2004) identified two problems. First, state tests do not assess the full range of standards laid out for students and second, tests often show a higher frequency of low-level questions, such as noninference reading questions or performing simple calculations in mathematics. Because of these flaws, Resnick et al. said that teachers will begin adjusting their instruction to meet the low-demand characteristics of state tests, instead of the high-demand standards that exemplify rigorous curriculum. Research relating to social studies teaching and learning has shown some signs of this outcome already (Bolinger & Warren, 2007; Leming, Ellington, & Schug, 2006).

At the classroom level, Tindal and Nolet (1996) described misalignment in two seventh grade science classrooms and found numerous inconsistencies between instruction, curriculum, and students' achievement. For example, while the curriculum involved in the study emphasized most science concepts equally, teachers emphasized one concept more often in comparison to another. Along with this, students ranked ideas and vocabulary that they were learning differently in comparison to the curriculum that teachers were using (Tindal & Nolet). Moreover, students demonstrated inconsistencies in how they performed on content assessments. For instance, assessment results showed large achievement gaps between general education students and low performing students on a content-specific test. However, Tindal and Nolet did not find the same gaps between different achieving students on skill-specific measures.

Porter et al. (1993) conducted another study, which analyzed the instruction of 62 secondary math and science teachers across 18 schools in six states. Similar to the results of Tindal and Nolet (1996), Porter et al. observed gaps between the intended and enacted

curriculums. However, Porter et al. described another set of characteristics that accompanied their observations relating to misalignment. First, math and science instruction at the secondary level were primarily expository in nature, relying heavily on listening to lectures and reading textbooks. Second, instruction lacked content depth, which Porter et al. defined as “the number of ways a topic was taught (modes of instruction) in combination with the number of different intended student outcomes” (p. 682).

Content Coverage and Instructional Practices

Issues related to the alignment problem such as excessive content coverage and use of expository instructional methods are not unique to math and science classrooms. Research by Faulkner and Cook (2006) and Leming et al. (2006) has shown that these problems occur in middle school environments and social studies classrooms as well. For example, Faulkner and Cook surveyed 146 middle school teachers with regard to their use of instructional strategies. Their results showed that teachers acknowledged the importance of using a wide variety of instructional strategies; however, most teachers indicated that their primary teaching methods were discussion, lecture, and worksheets. Moreover, instructional strategies used the least often by middle school teachers included hands-on experimentation, reflective writing, inquiry, and integrated units (Faulkner & Cook).

In another study, Leming et al. (2006) interviewed 1,051 elementary and middle school social studies teachers and found that they rated student-centered instruction as their preferred style when teaching. Nevertheless, when asked about their use of

instructional methods, 90% of teachers indicated that they used teacher presentation and discussion more often than other types of instruction.

Likewise, Bolinger and Warren (2007) surveyed 140 elementary and secondary social studies teachers and found that lecture was the most common method of instruction at all levels, especially in secondary grades. Other frequently employed practices included text readings and worksheets (Bolinger & Warren). Similar to other research, teachers in this study reported valuing student-centered instructional methods but used teacher-centered methods most often.

Last, Raphael, Pressley, and Mohan (2008) conducted a qualitative study of 10 sixth grade classrooms in two different middle schools to identify factors that promote student engagement. Their results showed that highly engaging teachers used a wide variety of motivational and instructional practices to support student learning and engage student interests (Raphael et al.). For instance, engaging teachers worked from organized lessons and incorporated modeling, explicit directions, and thoughtful questioning. Alternatively, less engaging teachers usually did not work from an organized plan, implemented a limited range of instructional strategies, and often used worksheets or other low-level activities (Raphael et al.).

The central problem of this study was whether reflective thinking promotes alignment between curriculum, instruction, and assessment at the classroom level for seventh grade social studies students. However, researchers have suggested that the problem of alignment is associated with multiple factors, many of which relate to classroom teaching. Some of these include unclear learning objectives (Harniss, 2006), an absence of teaching and learning theories that unify educational elements (Pellegrino,

2006), excessive content coverage (Porter et al., 1993), and use of a restricted range of instructional practices (Porter et al.). Moreover, additional research has shown that some of these factors are common to middle school learning environments and social studies classrooms (Bolinger & Warren, 2007; Faulkner & Cook, 2006; Leming et al., 2006; Raphael et al., 2008).

Marginalization of Social Studies

One of the objectives of those that support the standards movement has been to reduce problems associated with alignment, especially problems that stem from employing ineffective instructional practices (Nash, 2000). However, Vogler and Virtue (2007) said that the standards movement has done little to influence positive change in the instructional practices of secondary social studies teachers. Rather, the numerous standards produced by states may be exacerbating factors associated with the alignment problem. For example, Vogler and Virtue argued that social studies teachers are increasingly concerned with covering standards and less concerned with content depth. Moreover, the pressure that teachers feel to cover numerous standards has led to increased use of teacher-centered instruction because of its perceived efficiency (Vogler & Virtue).

There is another consequence of the standards movement on social studies teaching and learning, which is perhaps more problematic. According to the National Council for the Social Studies (2008), states and schools across the country have marginalized social studies curriculum, instruction, and assessment in favor of basic education disciplines such as literacy and mathematics. Pederson (2007) reported similar findings, saying that 25 states are giving less attention and resources to nontested subjects

like social studies. Likewise, McGuire (2007) suggested that teachers and administrators are emphasizing reading and math skills at the expense of teaching students about citizenship and democratic ideals.

Before the beginning of the standards movement, researchers such as Goodlad (1984) and Jackson (1990) argued that students spent most of their time in school developing basic skills, to the exclusion of learning important ideas and using higher-level thinking. Nevertheless, the marginalization of nontested disciplines has been especially troubling since social studies classrooms are often assigned the task of cultivating students' understanding of civic responsibility and democratic systems (Dyngneson & Gross, 1999; Jenness, 1990; National Council for the Social Studies, 2008).

The primary purpose of this study was to explore reflective thinking and its related instructional practices as a means for closing the gap between curriculum, instruction, and assessment at the classroom level. However, another related goal was to focus educators' attention back to the importance of social studies teaching and learning. This was especially important in light of research which has shown that the standards movement has caused some educators to marginalize nontested disciplines such as social studies (McGuire, 2007; National Council for the Social Studies, 2008; Pederson, 2007). Likewise, various researchers have suggested that social studies teachers use a limited range of instructional practices and are perhaps overly concerned with content coverage (Bolinger & Warren, 2007; Faulkner & Cook, 2006; Leming et al., 2006). Consequently, this study sought to examine instructional practices that have been shown to have a positive effect on student achievement and thereby increase the range of techniques available to social studies teachers.

Summary

In summary, reflection and reflective thinking are relatively popular topics. However, theorists and researchers have defined these ideas differently (Grossman, 2009; Kompf & Bond, 1995; Rodgers, 2002). One purpose of this study was to establish a description of reflective thinking and then identify its characteristics according to current research. Another purpose of this study was to identify instructional practices that replicate these characteristics and then test their effects on student achievement and perceptions of classroom instruction through an intervention.

The primary problem under investigation in this study was the alignment of curriculum, instruction, and assessment at the classroom level. Some researchers have suggested that excessive content coverage and applying a restricted range of instructional practices are factors that underlie this problem (Porter et al., 1993; Tindal & Nolet, 1996). Similarly, there is evidence which has shown that these factors are part of middle school learning environments and social studies classrooms in general (Bolinger & Warren, 2007; Faulkner & Cook, 2006; Leming et al., 2006; Raphael et al., 2008). In addition, some researchers have suggested that the standards movement has added to the obstacles faced by educators as they pursue effective instructional methods in social studies (National Council for the Social Studies, 2008; Vogler & Virtue, 2007). Moreover, an overemphasis on basic skills education, perhaps because of the standards movement, has diverted resources and attention away from social studies teaching and learning (McGuire, 2007; National Council for the Social Studies; Pederson, 2007).

Research Questions and Hypotheses

The questions for this study focused on instructional practices in the context of middle school social studies given the purposes and background previously discussed. Moreover, the investigator chose reflective thinking in light of research that has shown that social studies teachers mostly use teacher-centered instructional practices (Bolinger & Warren, 2007; Faulkner & Cook, 2006; Leming et al., 2006). Likewise, the investigator chose reflection as the central element of the intervention to focus instruction on content depth instead of content coverage.

Accordingly, the research questions for this study follow:

1. Do seventh grade students show increased achievement on a content-specific test when they engage in instructional practices characteristic of reflective thinking?
2. Do seventh grade students perceive classroom instruction differently according to a student questionnaire when they engage in learning practices characteristic of reflective thinking?
3. Do characteristics of reflective thinking influence low, medium, and high achieving seventh grade students differently according to student interviews?

The reason for conducting the empirical portion of this study was to examine the effects of instructional practices characteristic of reflective thinking on students' academic achievement according to a content-specific test. Along with this, the investigator used a student questionnaire to examine students' perceptions of classroom instruction that characterizes reflective thinking. The investigator chose an explanatory mixed methods design since this study examined two measures. This method involved a quantitative and qualitative phase (Tashakkori & Teddlie, 2003). The investigator

assessed student achievement and perceptions of classroom instruction during the quantitative phase of this study. These measures were represented by the first two research hypotheses for this study.

The third hypothesis connects to the qualitative phase. Lichtman (2006) wrote that research involving qualitative methods do not generally involve hypothesis testing. Likewise, Creswell and Clark (2007) stated that qualitative researchers only research questions and not hypotheses. The reason for including a hypothesis connected to the qualitative phase was to improve the integration of data and reporting of results from each of the two phases of the study. Another reason for the third hypothesis was to show that the qualitative phase was not intended as an “add-on” to the quantitative phase, but as an aid for interpreting quantitative results (Creswell & Clark, p. 108). In addition, Tashakkori and Teddlie (2003) suggested that investigators routinely use qualitative methods to explain quantitative results with mixed methods design.

Procedures relating to the third hypothesis involved assessing students’ perceptions of instructional practices characteristic of reflective thinking. Moreover, this hypothesis also incorporated interview data collected from a purposeful sample of high, medium, and low achieving students (Patton, 1990).

The research hypotheses for this study have been written in null form in comparison to directional form (Gall, Gall, & Borg, 2003). The reason for doing this was to facilitate the investigator’s use of these hypotheses for conducting statistical tests and reporting results (Gall et al.). The research hypotheses for this study follow:

1. Instructional practices characteristic of reflective thinking do not positively influence seventh grade students’ academic achievement in social studies.

2. Instructional practices characteristic of reflective thinking do not have an effect on seventh grade students' perceptions of classroom instruction in social studies.

3. Instructional practices characteristic of reflective thinking do not have a differential effect on low, medium, and high achieving seventh grade students in social studies.

Content of the Following Chapters

The remaining chapters of this study examine these questions and hypotheses. Chapter two provides a review of literature relating to reflective thinking and then suggests four characteristics that researchers associate with it. Also in this chapter is an examination of instructional practices that replicate these characteristics. The subject of chapter three is the method of the study and this section includes other topics such as population and sample, data collection procedures, and pilot study results. Chapter four describes the results of the study, organized according to each research hypothesis. Last, chapter five presents the study's conclusions organized by each research question.

Chapter Two: Literature Review

Purpose

Researchers have proposed various descriptions of reflection and reflective thinking (Grossman, 2009; McAlpine, Weston, Beauchamp, Wiseman, & Beauchamp, 1999; Rodgers, 2002). However, describing abstract human thought processes often leads to different explanations. For instance, ancient philosophers have suggested a number of characteristics useful for describing reflection and reflective thinking. In the Old Testament, the psalmist reported meditating on the law of the Lord by talking to himself day and night (Psalm 1:2, *The New King James Bible*). Elsewhere, the Greek sage Aesop told of an old woman who, chancing upon an empty wine bottle, recollected the once fragrant contents of the remaining dregs (Aesop, trans., 1992). In the *Tao Teh Ching*, the wise master Lao Tzu reminded the disciple that in order to cultivate the mind, one must “know how to dive in the hidden deeps” (trans., 1989, p. 17). Finally, in the *Bhagavad Gita* (2:41), the hero Arjuna was advised to contemplate one action at a time in order to avoid straying onto irresolute paths and innumerable distractions.

These sources serve as a reminder of how people from different cultures have taken an interest in the human capacity for higher-level thinking, such as reflection. This interest has continued as exemplified by some sources found in educational literature, each of which has described some element of reflection (Choi, Land, & Turgeon, 2005; Kaplan, Rupley, Sparks, & Holcomb, 2007; Li & Lal, 2005; Meyers, 2006).

Nevertheless, increasing the body of research pertaining to reflective thinking has added to the number of definitions that researchers have relied upon to describe it, such

as those proposed by Grossman (2009), McAlpine et al. (1999), and Rodgers (2002). One purpose of this review was to describe characteristics of reflective thinking according to current literature on the topic. Another purpose was to investigate theoretical sources that lend support to these characteristics as instructional practices. An additional purpose of this review was to identify instructional practices that replicate these characteristics and examine their effectiveness according to a representative sample (Randolph, 2009) of empirical studies relating to them.

The first section of this review describes criteria used for including literature. This is followed by a description of included literature, which the investigator used to identify characteristics of reflective thinking. The next section explores theoretical sources that underlie these characteristics and connects them to instructional practices. The final section examines studies relating to the instructional practices in question to evaluate their potential effectiveness as an intervention.

A final note regarding vocabulary used in this review is that the author interchanges the terms reflection and reflective thinking. This was done depending on the topic at hand; however, in this review reflection is thought of as the primary activity of reflective thinking.

Criteria for Literature Inclusion

In order to identify characteristics of reflective thinking the investigator searched four databases including Education Resources Information Center (ERIC), Education Full Text, PsycINFO, and PsycARTICLES. Moreover, each search involved specific criteria to locate relevant literature, such as including peer-reviewed articles and searching for keyword phrases like reflective thinking and reflection. In addition, the investigator

excluded articles older than 1989 since this review focuses on current interpretations and descriptions of reflective thinking.

Furthermore, the investigator applied other criteria because each database has its own set of search filters. For example, searches in ERIC were limited to elementary and secondary education since seventh grade is the subject group of interest in this study. Likewise, searches in PsycINFO and PsycARTICLES included school age subjects. Because of the large number of results in PsycARTICLES and PsycINFO, searches were also limited to educational classifications such as learning and memory, learning and motivation, and classroom dynamics, among others.

Initial search results produced 868 articles. From these, the investigator chose 67 and evaluated them according to additional criteria. These criteria included frequency of citation, theoretical support, and clarity of definitions. From this process, the investigator focused on six articles, including Brown (1997), Grossman (2009), King and Kitchener (2004), McAlpine et al. (1999), Rodgers (2002), and Spalding and Wilson (2002). Furthermore, the investigator used the reference lists of these six to locate additional articles, such as those by Hatcher and Bringle (1997), Mezirow (1997), Bruner (1996a, 1996b), and Kompf and Bond (1995).

In this review, the investigator applied two methods for selecting and including sources regarding the underlying theory researchers have used to explain reflective thinking. First, sources were included when researchers repeatedly referenced the same author. For instance, articles pertaining to reflective thinking often cited Dewey (1997a), Vygotsky (1978), Schön (1983, 1987), Brookfield (1995), and Shulman (1987). However, these last three sources focused on reflection in the context of teacher education and the

purpose of this study was to examine reflective thinking as an instructional practice for secondary students. As such, this review attended less to these authors and more to the first two. In addition, the investigator included sources when researchers repeatedly referenced specific constructs. For example, some researchers have connected reflective thinking to metacognition or have discussed it as a component of higher-level thinking (Brown, 1997; Choi et al., 2005; Grossman, 2009; Joseph, 2006).

Next, the investigator synthesized these results into a list of classroom-based instructional practices and then searched for empirical evidence of their effectiveness. Because these practices focus on instruction in educational contexts, the investigator limited searches of empirical studies to two databases, which included ERIC and Dissertation Abstracts. This portion of the literature review searched multiple keywords and keyword combinations, such as journaling, reflective prompts, metacognition, feedback, and middle school.

Overall, the investigator included a representative sample (Randolph, 2009) of empirical studies for each instructional practice. One outcome of this approach was the inclusion of three meta-analyses. While it is the case that meta-analyses summarize multiple studies in a parsimonious manner (Gall et al., 2003), some authorities have criticized the methods researchers use to reach conclusions using meta-analytic procedures (Fink, 1998). One specific criticism is that meta-analyses do not include studies that produced nonsignificant differences between control and treatment groups. Sheskin (2007) identified this phenomenon as the “file drawer problem” (p. 1307). This means that researchers leave studies with nonsignificant results unpublished.

Nevertheless, these studies were included because of their ability to summarize multiple research findings and represent the instructional practice of interest overall.

Additional criteria applied in section three include using middle school subjects and social studies content. Nevertheless, the investigator found few studies that integrated these criteria along with characteristics of reflective thinking. However, given the marginalization of social studies in favor of basic skills learning, the dearth of empirical research relating to social studies is unsurprising (McGuire, 2007; National Council for the Social Studies, 2008; Pederson, 2007). As an alternative to examining studies involving middle school students, some studies in this review included fifth grade and 10th-grade populations. Instances of these types of studies are those by Ruiz-Primo, Li, Ayala, and Shavelson (2004), King (1991), and Martin (2005).

Characteristics of Reflective Thinking

To identify characteristics of reflective thinking the investigator used six articles from the initial search and an additional article located from a reference list. The investigator chose these articles because of their clarity of definitions, explanation of theory, and potential for informing secondary classroom instruction. While the content of these articles is different, close examination of each shows important similarities. These similarities formed the basis for identifying characteristics of reflective thinking in this review.

First, Grossman (2009) suggested that reflective activity differentiates into four levels along a continuum. The first level is content-based reflection, defined as the development of one's reflection in light of a particular learning experience and guiding objective. Hatcher and Bringle (1997) developed this idea and suggested three strategies

for promoting content-based reflection. These strategies included connecting learning objectives to learning experiences, giving opportunities for regular reflection, and providing feedback on reflections.

The second level of reflection is metacognitive, defined as an awareness and knowledge of one's own thinking (Grossman, 2009). To promote metacognitive reflection, Grossman proposed a four-step model. These steps include describing the perceived thoughts of self and others, adopting the viewpoint of others through a case study, relating the content or topic at hand to a personal experience, and reflecting on one's thoughts at a highly descriptive level. Similar to Hatcher and Bringle (1997), Grossman recommended making students explicitly aware of the various steps needed to understand both the process and outcome of reflection.

Grossman (2009) identified the third and fourth levels of reflection as self-authorship and transformative. Self-authorship is reflection that occurs in two phases. These phases include assessing one's thoughts from a metacognitive state and understanding the effects that these thoughts have on one's behavior (Kegan, 1994). Transformative reflection is the process of changing personal perspectives through the adoption of alternative viewpoints (Mezirow, 1997). Grossman stated that both types of reflection involve an emotionally engaging experience such as service learning.

In comparison to Grossman's (2009) levels of reflection, Rodgers (2002) suggested that reflective thinking lacks a clear and cohesive definition. As a result, Rodgers argued that researchers have difficulty differentiating between reflective thinking and other types of cognition. Likewise, the lack of definition means that researchers have had trouble assessing it as a skill, communicating in a common language

about it, and researching its effects on learning. To remediate these difficulties, Rodgers proposed a four-part definition derived from Dewey's (1997a) theoretical writings on the topic.

First, reflection involves continuity, which Rodgers (2002) defined as making connections between parts and weaving experiences into a whole. For instance, someone writing a journal would likely make connections from one entry to the next, showing progress over time regarding some issue or question. Second, reflective thinking is rigorous, systematic, and disciplined. These characteristics take shape as the learner engages in an experience, describes that experience, and decides upon further action in an organized manner. Third, reflective thinking is interactional, it occurs among learners in a community through communication. Last, reflective thinking requires the adoption of certain attitudes, such as directness, which Rodgers defined as a kind of honesty about one's own knowledge and abilities.

Another perspective, from Brown (1997), stated that reflective thinking emerges from multiple factors, which influence classroom culture overall. To support this assertion, Brown cited an educational program called Fostering Learning Communities (FCL). The focus of FCL has been to design learning environments where grade school students think deeply about serious matters and cultivate metacognitive skills. For example, students in FCL are encouraged to employ self-questioning strategies and spend time researching in collaborative teams. Moreover, teachers model techniques of reflective thinking to students, such as thinking aloud, as part of the FCL instructional philosophy.

Brown (1997) described three principles, which guide FCL programs. These principles include agency, reflection, and collaboration. Agency involves designing learning environments that emphasize active and strategic learning, for example, engaging students in research projects to help them connect multiple learning experiences. Brown defined reflection as guiding students toward awareness of their own strengths and weaknesses as learners and assisting them in accessing their repertoire of learning strategies through metacognition. According to Brown, activities that promote reflection include thinking aloud, peer teaching, thoughtful questioning, and opportunities to revise answers.

Brown (1997) suggested that many questions still exist regarding the environmental factors that influence reflective thinking, such as metacognition. For instance, one question involves the role of the teacher and the instructional steps that are required to promote reflection. Another question involves how teachers organize learning environments overall to maximize students' reflective capacity.

Spalding and Wilson (2002) addressed parts of these questions, albeit, in the context of preservice teacher training. Nevertheless, Spalding and Wilson proposed that students improve their capacity for reflective thinking through explicit instruction on the process and product of reflection. By explicit instruction, Spalding and Wilson meant such activities as defining terms, discussing examples, and applying an organizational model. Moreover, these researchers defined growth in reflective capacity as students' ability to (a) distinguish between narration and reflection, (b) write different kinds of reflections according to a proposed model, and (c) connect learning activities together through reflective writing.

Like other researchers (Dunlap, 2006; McMillan & Wilhelm, 2007; Pavlovich, 2007), Spalding and Wilson (2002) have assumed that reflective writing promotes reflective thinking. Moreover, these researchers have suggested that the formats in which students write reflections, such as on paper or in an email, matter less to students than the feedback they receive. Feedback methods include positive comments, questions to stimulate elaboration, personal connections regarding shared experiences, and designation of reflection type. For the last method of feedback, Spalding and Wilson utilized a model developed by Valli (1997), which identifies four types of teacher reflection, including reflection-in-action, personalistic, deliberative, and critical. As students model characteristics of these categories in their writing, Spalding and Wilson gave feedback using a letter such as D for deliberative and so on. Additionally, after receiving feedback, Spalding and Wilson gave students an opportunity to revise their reflections and improve their initial response.

Similar to the research by Spalding and Wilson (2002), McAlpine et al. (1999) proposed a model to promote reflective thinking. However, unlike Spalding and Wilson, the model proposed by McAlpine et al. analyzed the reflective thinking of university professors instead of preservice teachers.

The model proposed by McAlpine et al. (1999) had six components including goals, knowledge, action, monitoring, decision-making, and corridor of tolerance. Knowledge consists of those domains of information familiar to teachers such as content knowledge and pedagogical knowledge, while action represents the decisions teachers make to achieve learning goals. According to McAlpine et al., monitoring represents feedback processes through which teachers guide their action. Moreover, McAlpine et al.

described the corridor of tolerance as the degree to which teachers accept dissonance between student achievement and intended learning outcomes.

From these six components, McAlpine et al. (1999) defined reflection as thinking about instruction through monitoring cues so that student learning coincides with learning goals and approximates intended outcomes. Overall, the model and definition proposed by McAlpine et al. emphasized processes of metacognition, monitoring, and feedback.

These techniques are not unique to the model of reflective thinking proposed by McAlpine et al. (1999). For instance, King and Kitchener (2004) described a reflective judgment model (RJM) and suggested that monitoring is a characteristic of it.

RJM is a theoretical framework for understanding how people deal with controversial issues, such as affirmative action or the origin of humans. As part of RJM, King and Kitchener (2004) proposed three levels of reflection. These levels include prereflective thinking, quasi-reflective thinking, and reflective thinking.

The primary characteristic of prereflective thinkers is their propensity for believing that knowledge is certain and that questions have definite answers (King & Kitchener, 2004). Alternatively, quasi-reflective thinkers understand that knowledge involves uncertainty and that individuals contribute to the construction of knowledge, although with only moderate influence. In addition, quasi-reflective thinkers exhibit less dependence on external authorities and instead rely more on evidence to shape their views. Last, reflective thinkers deliberately and consistently evaluate evidence to make judgments. Moreover, reflective thinkers generate and argue knowledge claims according to available information while remaining open to new data and perspectives.

Furthermore, King and Kitchener (2004) suggested that learners' capacity for reflective thinking has stage-like properties. However, they also argued that these stages do not necessarily evolve in a mechanistic, stepwise manner. Instead, these stages intertwine and overlap. Movement between stages depends on age and developmental level, but also on environmental factors that influence reflective-skill development.

King and Kitchener (2004) divided environmental factors that influence reflective-skill development into two categories, which they labeled as functional and optimal. Students in functional environments demonstrate minimal skill proficiency because they lack support. For instance, having students self-assess their performance on a quiz without establishing assessment criteria or providing feedback. Alternatively, students in optimal environments learn skills and use them to the best of their ability with contextual support. Contextual support includes those practices that simultaneously sustain and challenge learners, such as memory aids, instruction with examples, and teacher feedback.

An important conclusion reached by King and Kitchener (2004) is that learners' capacity for reflective thinking is limited according to age and developmental level. Nevertheless, King and Kitchener also proposed that teachers can improve students' capacity for reflective thinking, just as they can improve students' skills in other areas. However, research on the instructional practices necessary for this to occur has remained largely unexplored (Brown, 1997; King & Kitchener). In addition, the research that does exist has mostly investigated functional performance rather than optimal performance. For instance, King and Kitchener have claimed that RJM measures the middle range of

students' capacity for reflective thinking rather than the upper reaches of their reflective abilities.

Instead of describing a model of reflective thinking, Kompf and Bond (1995) associate reflection with multiple characteristics. To guide this description, they used cognitive theories proposed by John Locke (1974) and John Dewey (1997a). From these two sources, Kompf and Bond argued that reflective thinking is a preferred form of thought, which is purposeful, rigorous, and organized. Alternatively, unreflective thinking is random and chaotic.

Still other characteristics dealt with the timing and substance of reflection. For instance, Kompf and Bond (1995) proposed that reflection can occur during or after a learning event. Likewise, it can focus on a specific problem or question or it may resemble meditation, in which case there is no explicit catalyst upon which to focus, but rather freedom of thought. Moreover, reflection may involve social components or it may exist as an internal manifestation of one's own mind.

An additional characteristic of reflective thinking is its potential as a summarizing activity. For example, according to Kompf and Bond (1995), reflective thinking can assist learners in forming connections between past and present events. Along with this, it can serve as an opportunity for speculating about the future in light of what has already occurred.

Generally, Kompf and Bond (1995) suggested that reflection exists as a subskill of critical thinking, wherein the learner actively analyzes, synthesizes, and evaluates information gathered through observation and experience. While both critical thinking and reflective thinking involve deliberation, critical thinking also requires the application

of a set of standards through which learners evaluate information. Other sources on the topic of critical thinking, like Ennis (1987), McPeck (1981), Paul (1982), and Lipman (1988), have associated it with notions of skepticism, self-direction, and good judgment. In this way, critical thinking is more external and informative with regard to one's future action in comparison to reflective thinking.

Summary of Characteristics of Reflective Thinking

Literature used to describe reflective thinking has shown that it consists of multiple characteristics. According to the research cited in this review, these characteristics include continuity, explicitness, interaction, and metacognition (Brown, 1997; Grossman, 2009; Hatcher & Bringle, 1997; McAlpine et al., 1999; Rodgers, 2002; Spalding & Wilson, 2002). It is the case that researchers have described additional characteristics of reflective thinking and that this review focused on a limited number of sources. Likewise, differentiating reflective thinking from other forms, such as critical thinking, is a difficult undertaking. Nevertheless, these four characteristics have appeared across sources and they provide a structure for describing reflective thinking as a learning activity.

An additional observation derived from the literature examined in this review is that students' ability to engage in reflective thinking is proportional to their age and stage of development (Grossman, 2009; King & Kitchener, 2004; McAlpine et al., 1999). Nevertheless, more than one researcher claimed that growth in students' capacity for reflective thinking is possible given the right contextual factors. Some general examples of contextual factors include instruction and curriculum (Brown, 1997; King & Kitchener; Spalding & Wilson, 2002).

Theoretical Basis of Reflective Thinking

The next section of this review examines theoretical sources of reflective thinking according to the four characteristics described previously.

First, some researchers have proposed that continuity is a characteristic of reflective thinking and words that describe continuity include connection, succession, and union (Mish, 1987). Moreover, continuity, as a characteristic of thought, is an old idea. For example, in Plato's *Meno* (trans. 2006a), Socrates associated continuity of thought with recollection. The exact word for this was anamnesis, αναμνησις (Plato, p. 302), or calling to mind (Liddell & Scott, 1996). According to Socrates, learning was a form of recollection, which meant that the knower gathered memories and ideas together, reconnecting them to form continuous thought. Mostly, recollection was the result of discussion, specifically, the asking and answering of many questions. An example of this process was Socrates guiding a young boy to understand the qualities of irrational numbers (Bruner 1996a; Plato). Socrates taught the boy by asking one question after another, until the boy identified the qualities of irrational numbers for himself.

Likewise, Rousseau (trans., 2004) described the notion of continuity as constructing a chain of thoughts. This meant that learners advance from one idea to another, taking time to become familiar with each one before moving to the next. For example, Rousseau said that teachers must take care to connect lessons together so that learners advance their understanding in an orderly sequence, from one stage to the next, each experience building on the one previous.

Dewey (1997a) also likened continuous thought to a chain. According to Dewey, a chain of thought grows from one link to another, each supporting the next in line.

Development of thought depends on the integration of previous experiences with new experiences. One result of integration is that learners come to understand how their thoughts connect and relate. According to Dewey, when learners misunderstand the consecutive relationships between their thoughts, their understanding resembles a “miscellaneous scrap-bag” of facts (p. 97). Consequently, thoughts lack cohesion and meaning so that domains of knowledge remain unconnected.

Some researchers have proposed journaling as an instructional method for promoting continuity of thought (Dunlap, 2006; Hubbs & Brand, 2005; McMillian & Wilhelm, 2007). Some attributes of student journals which support this idea include revealing student thinking to a teacher’s informal assessment, providing a vehicle for students’ inner dialogue, and affording students an opportunity to connect their thinking from one instructional event to the next (Dunlap; Hubbs & Brand; McMillian & Wilhelm).

Another characteristic of reflective thinking is explicitness, which means fully revealed, expressed without vagueness, and leaving no question as to the meaning or intent (Mish, 1987). According to the literature examined in this review, explicitness is a component of reflective thinking, which teachers make use of when they communicate the learning objective to students, connect the objective to reflection, and provide instruction on the process of reflection (Brown, 1997; Grossman, 2009; Hatcher & Bringle, 1997; McAlpine et al., 1999; Spalding & Wilson, 2002).

In this review, learning objective is defined as the general goal or purpose of each lesson that guides and informs classroom instruction, activities, and assessment. Literature on reflective thinking suggested that learning objectives serve to stimulate and

structure student reflection (Hatcher & Bringle, 1997; McAlpine et al., 1999). In addition, Gustafson and Bennett (2002) proposed that stimuli, such as questions, directions, or probes, influence the quality of student reflection. Likewise, Griffith and Frieden (2000) specifically recommended Socratic questioning, integrated with dialogue, as an instructional method for prompting reflection.

Nevertheless, King and Kitchener (2004) and Song, Grabowski, Koszalka, and Harkness (2006) stated that quasi-reflective thinkers, such as seventh grade students, do not compare their knowledge claims to evidence, but rely more on concrete explanations or external sources of information to assess their learning. In another study, Brookhart (2001) found similar results among high school English and science students when discussing their perceptions of academic performance. For example, students in Brookhart's study evaluated their success according to multiple learning events, such as teacher feedback, preparing papers, and studying for tests, among others.

Furthermore, some researchers have advised that teachers take an explicit approach to promote reflective thinking among students (Brown, 1997; Grossman, 2009; Hatcher & Bringle, 1997). Generally, this means informing students about the process and product of reflection. Although researchers have proposed different methods for accomplishing this (Gustafson & Bennett, 2002), one approach involves guiding student reflection with a specific prompt (Griffith & Frieden, 2000; Song et al., 2006).

Turning again to *Meno* (Plato, 2006a), Socrates illustrated explicitness as a characteristic of learning by focusing investigations on a specific objective. In the *Meno*, the objective under consideration was whether people can teach and learn virtue. In order to address this question, Socrates deployed different kinds of instructional strategies such

as examples, summaries, and illustrations. Moreover, Socrates continuously revisited the original objective, constructing and establishing new propositions while keeping the essential question at the center of the dialogue. At the conclusion of the *Meno*, Socrates returned to the objective by proposing an answer based on established arguments.

Likewise, Dewey (1997a) suggested that learning events are anchored by a goal or objective, called the main topic of thought. According to Dewey, the main topic of thought is an organizing structure and it serves as the basis upon which one's ideas move toward a unified conclusion. Dewey argued that guiding one's thoughts according to the main topic does not consist of fixed and mechanical action. Rather, following the main idea to its conclusion involves variety and change, but variety and change that converges to a meaningful conclusion. One result of this convergence is that students revisit the main topic, or objective, and assess their own progress according to their experience.

Researchers have proposed that teachers promote reflection by instructing students on the process of reflection (Grossman, 2009; Spalding & Wilson, 2002). For instance, Grossman outlined a series of steps for modeling metacognitive reflection. Alternatively, Spalding and Wilson described a typology of reflection and shared its categories, definitions, and related examples with students. Although numerous instructional practices exist which teachers use to develop students' cognitive processes (Gredler, 2005; Joyce & Weil, 2004), one frequently cited practice is strategy instruction (Pressley & Harris, 1990).

According to Harris and Pressley (1991), one theoretical basis of strategy instruction has been Vygotsky's (1978) zone of proximal development (ZPD). Vygotsky defined ZPD as the awakening of internal learning processes through interaction with

adults, peers, and the environment. An important characteristic of ZPD is that teachers challenge students just enough to guide them toward new levels of mastery. Furthermore, Rousseau (2004) alluded to the same theory, stating that teachers should increase the difficulty of a learning task in proportion to the child's ability. Nevertheless, with respect to strategy instruction, Harris and Pressley defined ZPD as the distance between what a learner can do independently and what a learner can do with the assistance of competent peers or adults. Moreover, Harris and Pressley argued that strategy instruction works best when teachers guide students to perform at the upper boundary of their current abilities.

Pressley and Harris (1990) recommended the following teacher behaviors for promoting strategy instruction: (a) demonstrating the strategy in the context of a meaningful academic task, (b) introducing strategies one at a time, (c) providing feedback and opportunities for practice, and (d) assisting students that struggle with the strategy on an individual basis. Additional recommendations for teachers regarding strategy instruction have included explaining the steps to direct task performance, verbal modeling, systematic prompts, and teacher to student dialogue and questioning (Reid & Lienemann, 2006).

In summary, characteristics of explicitness include the structure that teachers use to guide students' reflective thinking. One method for integrating explicit qualities into the reflective process is with a prompt (Griffith & Frieden, 2000; Song et al., 2006). Another method is strategy instruction, specifically, modeling the process of reflection for students and setting standards with regard to product, such as delineating the length of a written reflection (Choi et al., 2005; Hatcher & Bringle, 1997; Spalding & Wilson, 2002). However, along with these structural characteristics, researchers have also

emphasized interaction as a characteristic of reflection, primarily as a means for producing instructive feedback (Choi et al.; Song et al.; Winne 2005).

Interaction has been defined as a mutual or reciprocal action or a relationship that involves influence (Mish, 1987). The literature examined in this review suggested that reflective thinking includes interaction at two levels. These levels include student to student and teacher to student.

Similar to the other characteristics examined in this review, interaction is also an element of the teaching method of Socrates (Plato, 2006a, 2006b). For instance, in *Meno*, Socrates stated that teaching is not didacticism, διδάσκοντα (2006a, p. 306), but a process of joint inquiry, or more literally, seeking answers with another person (2006a, p. 314). This process of joint inquiry begins with questioning, followed by some kind of assessment, for instance, an evaluation of the validity of arguments. In *Meno*, Socrates often assessed discussion points and then followed with summary. These assessments and summaries served as starting points for additional inquiry (2006a).

Furthermore, Socrates used the same interactional methods with younger students and intellectual peers alike (Plato, 2006a, 2006b). This suggests that feedback processes, whether student to student or teacher to student, share similar qualities, such as reciprocal action and collaborative inquiry.

Likewise, Dewey (1997b) stated that interaction is a principle of learning because it involves stimulating interpersonal contact and communication. Moreover, interaction is not limited as an occurrence between people, but it can happen between people and objects in their environment. Nevertheless, social interaction, according to Dewey, is the most significant because it fluctuates and adjusts to the needs of the learner. Likewise,

social interaction means that the learner has some influence over the direction of the interaction. For instance, peer-questioning involves at least two students interacting together in a cooperative manner.

Vygotsky (1978) also emphasized social interactions as a basis for learning. According to Vygotsky, cognitive functions develop on two levels, first through interpersonal processes and then through intrapersonal processes. For example, a young child may point at an object just out of reach. However, the purpose of pointing is not to make the object move, but to secure help from nearby adults. As adults come to the child's aid, the child comes to understand the meaning of pointing, primarily through social interaction (Vygotsky).

Two practical manifestations of interaction, in the context of classroom learning, include peer-questioning (Brown, 1997; Choi et al., 2005; Spalding & Wilson, 2002) and formative assessment (Hatcher & Bringle, 1997; McAlpine et al., 1999).

King (1991) defined peer-questioning as a social-cognitive activity through which students clarify, elaborate, and reconceptualize content for one another. Alternatively, Bell and Cowie (2001) defined formative assessment as an interactional process between teacher and student for informally determining the content of student thinking and progress toward learning goals.

Furthermore, some researchers have associated instructional practices such as peer-questioning and formative assessment with metacognition (Brown 1997; Choi et al., 2005; Griffith & Frieden, 2000; Winne, 2005). Metacognition is concerned with how learners think about their own thinking and account for their own mental procedures (Bruner, 1996a; Flavell, 1979).

Researchers such as Schunk (2008) and Zimmerman (2002) situated metacognition within the larger construct of self-regulated learning. Examples of metacognition include assessing one's progress toward completing a learning objective or understanding one's personal strengths and weaknesses as a student (Flavell, 1979). Bruner (1996a) suggested that modern pedagogical practices should focus more on teaching students metacognitive strategies, since these assist learners in transcending their present cognitive capabilities.

Metacognition has been a topic of interest among researchers for more than 40 years (Brown 1978; Flavell, 1979). However, the ideas underlying metacognitive thinking have existed since ancient times and like the other characteristics of reflective thinking discussed in this review, Socrates provided examples of its qualities.

In *Theaetetus* (Plato, 2006b), Socrates defined thought as the soul conducting a detailed conversation with itself. As part of this conversation, the soul asks questions and forms arguments, which it then evaluates. Socrates identified this process as an inner-cognitive operation. This is not to suggest that thinking is the result of an individual effort. For instance, in other sections of *Theaetetus*, Socrates noted that the same process of evaluation occurs through conversation. Nevertheless, an important feature of Socrates' method in relation to metacognition is that the mind is capable of assessing its own thoughts.

In comparison, John Locke (1974) proposed that the mind obtains knowledge through sensory stimuli and reflection, as opposed to recollection or anamnesis (Plato, 2006a). According to Locke, reflection is the mind noticing its own operations and contemplating its own ideas. For instance, young children learn about the characteristics

of hot and cold through sense experience and then, through reflection, they assign these qualities to objects with words. At an advanced level of reflection, children categorize objects together depending on a particular quality, for example, milk, snow, and chalk, which all have the quality of being white (Locke). The result of reflection is understanding, which Locke defined as the mind turning in upon itself to observe its own ideas.

Just as Locke (1974) and Socrates (Plato, 2006b) described the mind turning inwards upon itself, so Dewey (2004) observed that reflection is an understanding of the details that connect an action to a consequence. Thinking without reflection is similar to trial and error since the one doing the thinking tries different solutions without attempting to understand why one method works and another fails. Moreover, thinking without reflection is random and discontinuous. However, reflective thinking, according to Dewey, involves an understanding of the connections between action and result.

More recently, Zimmerman (2002) described metacognition as a process of self-regulated learning, which involves a number of cognitive skills including (a) setting proximal goals, (b) adopting learning strategies, (c) evaluating the efficacy of one's methods, and (d) adapting future methods, among others. More practically, Dignath and Büttner (2008) suggested that planning the completion of a task, monitoring one's comprehension through self-testing, and evaluating one's learning products in comparison to a goal are indicative of metacognition. Moreover, Dignath and Büttner also emphasized the importance of communicating to students how, when, and where to apply various metacognitive strategies while also illustrating the benefits of their use.

In summary, reflective thinking consists of multiple characteristics. According to current literature relating to reflective thinking, some of its characteristics include continuity, explicitness, interaction, and metacognition. Moreover, some theoretical support exists which identifies these characteristics as elements of teaching and learning.

Although other instructional practices exhibit these traits (Gredler, 2005; Joyce & Weil, 2004), the literature examined in this review has shown that (a) student journals, (b) reflective prompts with strategy instruction, (c) feedback processes, and (d) metacognition may promote reflective thinking in classrooms. The next section of this review analyzes a representative sample (Randolph, 2009) of empirical studies that examined these practices to illustrate their potential for promoting reflection in the context of seventh grade social studies.

Empirical Basis of Instructional Practices Associated with Reflection

First, Mezirow (1997) proposed that using journals as an instructional method encourages students to think about lesson content. In addition, student journals provide a practical format through which teachers can regularly assess student thinking on a topic or question.

In a related study, Ruiz-Primo et al. (2004) analyzed the effects of student journals on the communication skills of fifth grade science students. The investigators selected six intact classrooms using stratified sampling measures according to pretest and posttest scores on a content-specific science test. Then, Ruiz-Primo et al. designed and implemented a journal rating system and used it to assess student performance over the course of the school year. Independent raters scored 24 science journals with inter-rater reliability ranging from .82 to .99. Next, Ruiz-Primo et al. assessed correlations between

journal scores and science achievement according to two different performance assessments. Results showed that journal scores in areas such as communication, conceptual understanding, and teacher feedback correlated to other achievement measures, such as content-specific science tests.

In addition, Ruiz-Primo et al. (2004) observed changes in students' quality of communication between science units at different points in the school year. According to a dependent *t* test, the quality of students' journal writing decreased over time. This decrease coincided with a reduction in the number of times teachers wrote feedback in students' science journals.

The study by Ruiz-Primo et al. (2004) showed that student journals serve as an assessment tool. Likewise, with an appropriate rating system, the work in these journals correlates to other achievement measures. In addition, Ruiz-Primo et al. found that teacher feedback and monitoring affected the quality of students' written communication in journals. Other research in the area of teacher feedback supports this last finding (Black & Wiliam, 1998a, 1998b; Gallagher & Worth, 2008).

In comparison to the study by Ruiz-Primo et al. (2004), Werderich (2006) conducted a qualitative study that analyzed teacher feedback and student journals with sixth grade and seventh grade students. For this study, Werderich analyzed 600 pages of dialogue journals containing student reflections about literature.

According to Werderich (2006), the teachers in this study responded to student reflections according to four response categories. These categories included visual aid, modeling, questioning, and feedback. However, analysis showed that teachers applied these categories loosely, depending on students' writing. For example, teachers

remediated grammatical errors, reinforced ideas, and identified with students' feelings on a topic. Mostly, Werderich noted that teachers were interested in guiding student thinking through specific and individualized feedback responses.

In summary, the research by Ruiz-Primo et al. (2004) and Werderich (2006) showed that student journals serve as a repository for students' reflective writing, correlate with other content measures of achievement, and require teacher feedback and monitoring to maintain their effectiveness. Nevertheless, one topic not thoroughly discussed by Ruiz-Primo et al. and Werderich were methods for structuring students' reflective writing. Although literature in this area has proposed various structuring methods (Griffith & Frieden, 2000; Grossman, 2009; Gustafson & Bennett, 2002), two specific techniques include reflective writing prompts and providing steps or a strategy for students to follow as they reflect (Dunlap, 2006; Hatcher & Bringle, 1997; Spalding & Wilson, 2002).

Song et al. (2006) conducted a study with middle school students related to methods for structuring written reflection through problem solving. In this study, Song et al. surveyed 122 students from six intact classrooms to determine which instructional design elements were most helpful in promoting reflective thinking. For six months, students in this study engaged in problem based learning, facilitated by instructional practices indicative of reflective thinking such as reflective judgment model, collaborative learning, and scaffolding techniques. Song et al. defined scaffolding techniques as instructional practices used to promote reflective thinking such as reflective writing, teacher explanations, and reflective writing prompts.

In this study, Song et al. (2006) used a 10-item survey, which the investigators designed. According to their analysis, the survey showed sufficient reliability, $\alpha = .88$, and demonstrated adequacy with regard to additional statistical tests. However, this study involved factor analysis with 122 subjects, which Pett, Lackey, and Sullivan (2003) said was a small number of participants for conducting factor analysis procedures. Nevertheless, after administering the survey, Song et al. analyzed the results and identified two factors titled Scaffolding Methods and Constructivist Learning Environments. Scaffolding Methods included instruction involving drawing, writing, answering specific questions, and teacher explanations. Alternatively, elements of the second factor, Constructivist Learning Environments, included working on real world problems, student collaboration, and having time to think about questions before answering.

In addition, students in this study said that instruction characteristic of Constructivist Learning Environments, especially those involving social interaction, were more helpful for their learning than instruction associated with Scaffolding Methods (Song et al., 2006). Nevertheless, more than one researcher has proposed providing students with a set of steps or strategy to promote reflection (Hatcher & Bringle, 1997; Spalding & Wilson, 2002). One way to examine the potential effects of organizing students' reflective process according to a set of steps is through strategy instruction.

Beckman (2002) defined strategy as a tool or method for accomplishing a task. More broadly, strategy instruction involves teaching students how and when to use a specific learning technique. In addition, Beckman stated that strategy instruction involves

metacognition and self-regulation since one of the goals of this approach is to encourage students to oversee and monitor their use of learning strategies.

Studies by De La Paz (1999) and Graham and Perin (2007) identified effective characteristics of strategy instruction for improving middle school students' writing performance. Although improving students' writing is not the focus of this review, researchers have frequently connected writing with reflection (Dunlap, 2006; McMillan & Wilhelm, 2007; Pavlovich, 2007; Spalding & Wilson, 2002). Furthermore, the studies by De La Paz and Graham and Perin outlined general approaches regarding effective strategy instruction, applicable to various content fields, such as social studies.

De La Paz (1999) studied the effects of strategy instruction on middle school students' essay writing performance. Participants for this study included 22 students equally divided between genders; six were learning disabled. De La Paz used a baseline method to establish pretreatment measurements of essay writing ability, followed by posttest and retention test measures. The independent variable for this study was a writing strategy that included the following steps: teacher modeling, activation of prior knowledge, collaborative practice, mnemonic devices, and making a prewriting plan (De La Paz). Dependent measures included the number of words students used in an essay, the quality of students' prewriting plan, and functional essay elements. De La Paz defined functional essay elements as those writing components that directly supported the writer's propositions, such as explanations or conclusions. According to De La Paz, interrater reliability for students' prewriting plan and functional essay elements was between .78 and .90.

Although this study represented a relatively homogeneous group of students, 90 percent were Caucasian, and a small sample size, overall results showed that general education and learning disabled students improved their essay writing performance (De La Paz, 1999). Specifically, in comparison to baseline data, students doubled the length of their essays and included two to three times as many functional essay elements.

In another study, Graham and Perin (2007) conducted a meta-analysis summarizing effective practices for teaching writing to secondary students. In this study, the investigators applied multiple criteria for selecting research to include in their analysis. For instance, studies were included if they reported interrater reliability measures of writing quality at .60 or higher, applied experimental or quasi-experimental designs, and reported sufficient information for calculating effect sizes such as the number of subjects. Studies that were not experimental or quasi-experimental and that did not include measures of writing quality were the two most frequent reasons for exclusion. In addition, the meta-analysis conducted by Graham and Perin excluded studies with small sample sizes, such as the one conducted by De La Paz (1999).

Graham and Perin (2007) searched multiple databases including ERIC, Dissertation Abstracts, and PsycINFO. These searches involved keywords such as writing and composition, along with other descriptors such as strategy instruction, peer revising, and summary instruction. As studies were located, Graham and Perin organized and coded them according to grade, number of subjects, writing genre, and publication type. Overall, 123 studies were included in the meta-analysis, which were then analyzed using Cohen's *d* to calculate an effect size for each category of treatment.

According to Graham and Perin (2007), the following strategies showed effect sizes between .70 and .83: (a) teaching strategies for planning, revising, and editing, (b) teaching summarization strategies, (c) integrating collaborative writing activities, and (d) setting product goals for students' writing. Last, Graham and Perin said that strategy instruction was a particularly effective teaching practice to use with struggling students.

In summary, the research by Song et al. (2006), De La Paz (1999), and Graham and Perin (2007) described a number of approaches for structuring reflective thinking. These methods included scaffolding support such as drawing, writing, and peer collaboration (Song et al.). Additional methods, relating specifically to writing reflections, included teacher modeling, collaborative practice, and writing product requirements (De La Paz; Graham & Perin). Furthermore, the studies examined in this review involving strategy instruction also emphasized interaction or feedback processes.

However, the importance of interaction and feedback is not limited to studies involving strategy instruction. Other researchers connect these characteristics to reflective thinking (Brown, 1997; Choi et al., 2005; Hatcher & Bringle, 1997; McAlpine et al. 1999; Spalding & Wilson, 2002). Although there are various methods for facilitating classroom interaction and feedback processes (Nelson & Schunn, 2009; Wiggins, 1993), two methods connected to reflective thinking include peer-questioning and formative assessment (Brown; Choi et al.; Hatcher & Bringle; King, 2002; McAlpine et al.; Spalding & Wilson).

In a study related to these topics, King (1991) examined the effect of peer-questioning on fifth grade students' problem-solving performance. In this study, King randomly assigned students from intact classrooms into three treatment conditions. These

conditions included guided questioning, unguided questioning, and a control group. King randomly assigned each student a partner within each of these conditions. However, students received the treatment at different times during the school day. Nevertheless, students spent three weeks in each of the three conditions and at the end of this time, King tested students on several dimensions of problem solving. Measurements in this study included partner problem solving using computers, a content-based problem-solving test, and observational measures. King did not report the reliability of test measures, but did state that interrater reliability of observational measures ranged from .90 to .98.

Overall, King (1991) found that guided peer-questioning enhanced problem-solving performance and increased students' use of strategic talk, such as giving an explanation. Likewise, additional studies by Choi et al. (2005) and McDuffie, Mastropieri, and Scruggs (2009) showed that structured peer interaction has a positive effect on learning.

Another element of interaction is the feedback that teachers provide students. One way to conceptualize this interaction is through formative assessment, which multiple studies showed produces a positive effect on academic performance (Black & Wiliam, 1998a, 1998b; Fuchs & Fuchs, 1985, 1986). For instance, Black and Wiliam examined 43 studies, including subjects from kindergarten to undergraduate, on the effects of formative assessment on student learning. From their analysis, Black and Wiliam found that typical effect sizes associated with formative assessment ranged between .40 and .70 (Black & Wiliam, 1998b). One of the notable studies included in Black and Wiliam's investigation was a meta-analysis by Fuchs and Fuchs (1985).

Fuchs and Fuchs (1985) assembled studies for their meta-analysis by generating a list of keyword phrases such as student achievement, student progress, and educational effects. Then, Fuchs and Fuchs searched multiple databases including ERIC, Dissertation Abstracts, and Psychological Abstracts. The researchers followed this with a manual search of five educational journals for the same keyword phrases. Last, Fuchs and Fuchs used the reference sections of located studies to find additional titles.

Furthermore, Fuchs and Fuchs (1985) included studies that employed a control group and that evaluated the effects of systematic formative evaluation on the academic achievement of elementary and secondary students. The evaluation excluded studies that monitored nonacademic behavior, focused on behavior modification, or only provided test feedback. Next, Fuchs and Fuchs (1985) coded and organized the studies according to methodological and substantive features. The interrater reliability for the coding of these categories ranged between .86 and .92.

At the conclusion of coding and organizing procedures, Fuchs and Fuchs (1985) analyzed 21 studies, a relatively small number, according to their own assessment. Nevertheless, the investigators used Hedges's g to calculate an effect size of .72 and stated that systematic formative assessment reliably increased academic achievement.

In summary, research has indicated that feedback originating from students and teachers has a positive effect on students' academic performance. The study by King (1991) showed that peer-questioning is one method for promoting peer feedback. In addition, research by Black and Wiliam (1998a, 1998b) and Fuchs and Fuchs (1985, 1986) has shown that teacher feedback through formative assessment is an effective instructional practice. However, Black and Wiliam (1998b) also proposed that formative

assessment is insufficient by itself for sustaining long-term academic gains. Rather, they recommended accompanying formative assessment with additional instructional practices such as teaching students how to self-assess and communicating the desired learning goal to students (Black & Wiliam, 1998b).

Each of these recommendations relates to metacognition, which is associated with self-regulation and self-regulated learning (Schunk, 2008; Zimmerman, 2002). A meta-analysis by Dignath and Büttner (2008), which analyzed 74 studies, showed that interventions involving metacognitive and self-regulated learning interventions have a positive effect on student learning.

The objective of the meta-analysis conducted by Dignath and Büttner (2008) was to evaluate the impact of self-regulation strategy training on secondary students' academic achievement, strategy use, and motivation. An additional research question addressed in this study was whether there were specific training characteristics that made interventions particularly effective. According to these objectives, Dignath and Büttner searched three databases including ERIC, PsycINFO, and Psynindex, the German online database. Similar to other meta-analyses included in this review (Fuchs & Fuchs, 1985; Graham & Perin, 2007), Dignath and Büttner began by searching for keywords such as metacognition, metacognitive skills, and self-regulatory strategies.

Dignath and Büttner (2008) included studies in their analysis if they (a) represented school age children in normal school settings, (b) used pretest-posttest control-group designs or controlled for group differences before the intervention was applied, and (c) focused on self-regulated learning with one or more academic components as part of the intervention. In addition, only studies published after 1996

were included in order to compare the results with research conducted by Hattie, Biggs, and Purdie (1996), who also conducted a meta-analysis on self-regulated learning.

Next, after assembling a pool of studies, Dignath and Büttner (2008) grouped them according to three outcome categories including academic performance, use of cognitive and metacognitive strategies, and motivation. Interrater agreement for the assignment of these categories ranged from .80 to 1.00.

Dignath and Büttner (2008) analyzed 25 studies at the secondary level and extracted 94 effect sizes. However, these studies only represented published results with statistically significant findings. Nevertheless, the weighted overall mean effect size for secondary school studies was .71, while the mean effect size calculated from studies incorporating strategy use as part of the intervention was .88.

In addition, Dignath and Büttner (2008) reported statistically significant correlations between metacognitive reflection and training characteristics that incorporated strategy use. The investigators defined metacognitive reflection as training students on how, when, and where to use metacognitive strategies and communicating the benefits of their application. Furthermore, this study also found that effect sizes were greater when researchers administered interventions instead of regular classroom teachers (Dignath & Büttner). Hattie et al. (1996) also found this phenomenon in their meta-analysis. Last, Dignath and Büttner reported that treatments involving group interaction positively influenced effect sizes in studies with secondary students.

Although the findings by Dignath and Büttner (2008) suggested that metacognitive reflection coupled with strategy use is an effective instructional practice, additional research by Martin (2005) demonstrated that 10th-grade students in social

studies classrooms adopt metacognitive thinking and strategy learning in a haphazard manner.

Martin (2005) conducted a qualitative study of 40 students from two intact classrooms to explore the cognitive strategies that students used to learn information in social studies. Participants in this study consisted of an equal number of male and female students, selected according to their gender, academic level, and academic performance according to class grade. With regard to academic level, Martin chose students from average and advanced classes; however, all participants had class grades above 80 percent. Martin collected data by interviewing students and two participating teachers. Interviews consisted of open-ended questioning, which lasted 10 to 20 minutes.

Results of the interviews indicated that students in this study acquired learning strategies through trial and error, by asking a family member or friend, or imitating the strategies used by their teachers (Martin, 2005). Furthermore, Martin stated that students deployed sophisticated strategies when they saw a need, such as preparing for a test. Only two of the 40 students said that they had received explicit strategy instruction on how to learn lesson content.

However, Martin (2005) did not describe the process by which data was organized, coded, or analyzed. Nevertheless, the results reported in this study are consistent with other research. For instance, Brown (1997) stated that students do not demonstrate continued use of learning strategies when left unsupported. In addition, Bolinger and Warren (2007) and Leming et al. (2006) found that social studies teachers consistently utilize instructional practices that do not require higher-level thinking.

In summary, research by Dignath and Büttner (2008) suggested that metacognitive training has a positive effect on student achievement. These effects are enhanced when coupled with elements of strategy use and peer interaction (Dignath & Büttner). Moreover, research by Martin (2005) showed that students develop and use metacognitive learning strategies mostly through trial and error if left to their own efforts.

Further Research Involving Reflective Thinking

According to the literature examined in this review, reflective thinking consists of multiple characteristics. These characteristics include continuity, explicitness, interaction, and metacognition. Instructional practices indicative of these characteristics include student journaling, reflective prompts and strategy instruction, feedback processes, and metacognition. Other researchers and educational sources have sorted these characteristics and their corresponding practices into two constructs, including scaffolding techniques and reflective assessment (Choi et al., 2005; Ellis, 2001; Song et al., 2006; Wang & Lin, 2008; Winne, 2005).

Winne (2005) defined scaffolding techniques as instructional aides that assist learners in completing tasks or achieving goals that are beyond their unassisted efforts (Bruner, 1975). One theoretical basis of scaffolding is Vygotsky's (1978) zone of proximal development, a theory mentioned earlier in this chapter. The scaffolding techniques examined in this review include student journals, reflective prompts, strategy instruction, and peer-questioning (Choi et al., 2005; Combs, 2004; Puntambekar & Hubscher, 2005; Song et al., 2006.). Alternatively, reflective assessment combines elements of metacognition and formative assessment (Ellis, 2001; Wang & Lin, 2008).

Generally, researchers have conducted multiple studies on the topic of reflective thinking. For instance, previous research has investigated reflective thinking and its relationship to students' beliefs (Searles, 1980), use of student journals (Seifert, 2008), continuous feedback (Jimarez, 2005), and problem solving (Bayless, 1965).

Nevertheless, Winne (2005) proposed that additional studies are needed to measure the effects of various scaffolding techniques on different types of students. In addition, researchers have shown that reflective assessment is an effective instructional practice with 10th-grade English students (Evans, 2009), ninth grade science students (Bianchi, 2007), and fifth and sixth grade math students (Bond, 2003). However, these studies did not examine the effects of an intervention that combined scaffolding techniques and reflective assessment in the context of seventh grade social studies.

Chapter Three: Methodology

Chapter Overview

The subject of this chapter is the methodology of the study, described in four sections. The first section begins with restatement of the research questions and a general description of the research design, population and sample, protection of participants, and intervention. The second section describes the quantitative phase of the study and its instruments. This section also contains an account of pilot test results and analysis of instrument reliability and validity. The subject of the third section is the qualitative phase and its method. In addition, this section examines the issue of interview data credibility. Last, this chapter describes the study's internal and external threats to validity and then concludes with a summary.

Research Questions

The research questions follow:

1. Do seventh grade students show increased achievement on a content-specific test when they engage in instructional practices characteristic of reflective thinking?
2. Do seventh grade students perceive classroom instruction differently according to a student questionnaire when they engage in learning practices characteristic of reflective thinking?
3. Do characteristics of reflective thinking influence low, medium, and high achieving seventh grade students differently according to student interviews?

Research Design

The purpose of this study was to investigate the effects of reflective thinking on student achievement and student perceptions of classroom instruction. An additional purpose of this study was to examine the perceptions of low, medium, and high achieving students with regard to instructional practices that characterize reflective thinking.

Overall, the investigator chose a mixed methods design to investigate questions related to these purposes.

Tashakkori and Teddlie (2003) defined mixed methods as the integration of qualitative and quantitative approaches at each phase of the research process including design of research questions, methods, data collection, and analysis procedures. Creswell and Clark (2007) recommended the use of mixed methods research when one approach is inadequate by itself to address the research problem and its underlying questions. For instance, mixed methods is appropriate when the research question has multiple parts such as measuring the effects of a treatment, understanding complex phenomena, or testing new ideas in the form of relationships (Tashakkori & Teddlie).

In addition, Creswell and Clark (2007) suggested that using multiple data sources assists investigators in revealing mechanisms of causality. Likewise, Tashakkori and Teddlie (1998) stated that mixed methods research provides researchers with more opportunities to construct causal inferences through such processes as triangulation. Tashakkori and Teddlie defined triangulation as the combination of multiple data sources and analyses to form conclusions. Richards (2005) wrote that triangulation is the interpretation of a problem or situation through the application of different research

methods and analysis of different data types. In this study, the investigator compared quantitative and qualitative data to improve the interpretability of results overall.

Last, mixed methods reduces some of the problems associated with purely quantitative or qualitative methods such as (a) an over emphasis on statistical significance, (b) failure to incorporate a broad range of information, (c) selectivity in the reporting of results, and (d) failure to examine alternative perspectives (Tashakkori & Teddlie, 2003).

The literature reviewed in the previous chapter suggests that reflective thinking involves multiple characteristics (Grossman, 2009; Kompf & Bond, 1995; Rodgers, 2002). This study examined these characteristics and their influence on the alignment of curriculum, instruction, and assessment, as a way to improve students' achievement on a content-specific test. However, curriculum, instruction, and assessment are only three variables that influence classroom environments. One goal of this study was to unite and emphasize these three elements through the intervention. Research examined in the previous chapters suggests that uniting and emphasizing curriculum, instruction, and assessment at the classroom level has a positive effect on student achievement and perceptions of instruction (Porter et al., 1993; Roach et al., 2008; Tindal & Nolet, 1996).

The intervention in this study incorporated characteristics of reflective thinking, including continuity, explicitness, interaction, and metacognition (Brown, 1997; Grossman, 2009; Hatcher & Bringle, 1997; McAlpine et al., 1999; Rodgers, 2002; Spalding & Wilson, 2002). Moreover, the investigator for this study organized these characteristics through scaffolding techniques and reflective assessment (Choi et al., 2005; Ellis, 2001; Song et al., 2006; Wang & Lin, 2008; Winne, 2005).

Figure 1 shows an illustration of these relationships, according to the literature review. First, various factors influence student thinking in classroom environments (Jackson, 1990). Figure 1 shows some of these factors including interest, goal orientation, and prior knowledge. Also shown is the intervention, organized within the constructs of scaffolding techniques and reflective assessment. The rectangular box above the intervention describes characteristics of reflective thinking. Moreover, this part of Figure 1 shows instructional practices the investigator used to replicate these characteristics, such as student journals. The dashed lines represent the investigator's efforts at unifying curriculum, instruction, and assessment through the intervention. Last, the solid lines stand for the positive influence that the investigator predicted for bringing curriculum, instruction, and assessment together to influence student thinking.

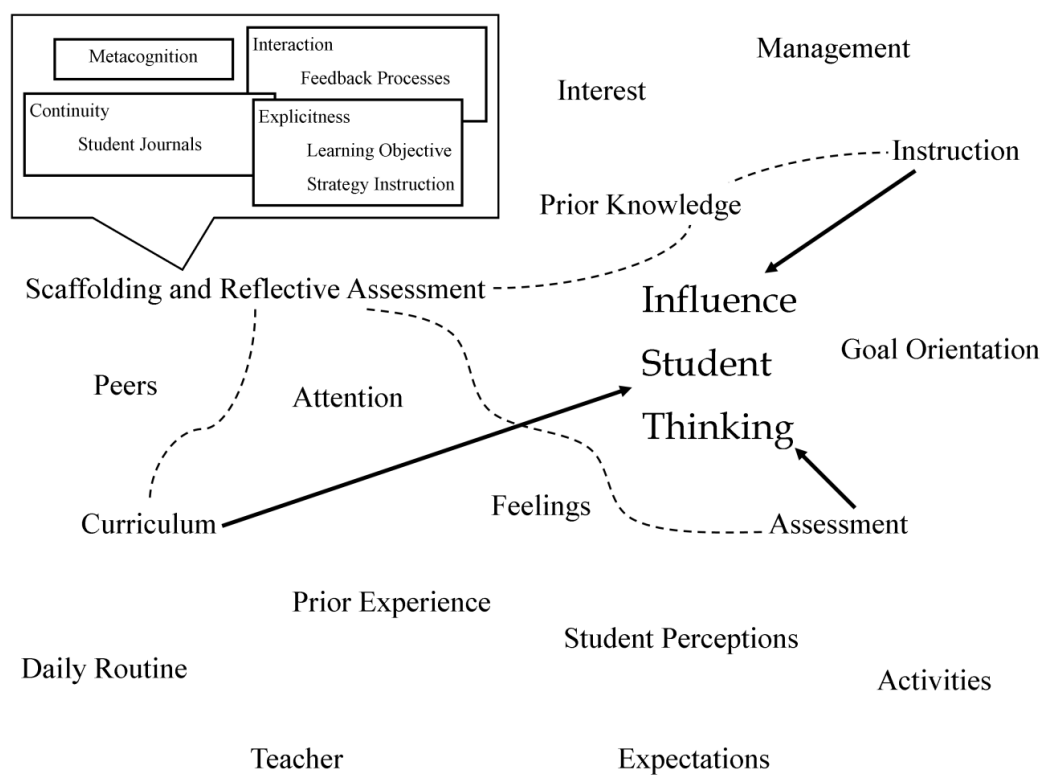


Figure 1. Alignment of curriculum, instruction, and assessment.

Similar to this study, Papanastasiou (2000) and Tapola and Niemivirta (2008) applied mixed methods approaches to their studies in order to explore multiple facets of complex phenomenon. Papanastasiou investigated teachers' job satisfaction in Cyprus while Tapola and Niemivirta investigated sixth grade students' achievement motivation. Likewise, this study examined an educational problem, which has multiple parts. Specifically, the problem investigated was the alignment of curriculum, instruction, and assessment through instructional practices, which characterize elements of reflective thinking, namely, scaffolding techniques and reflective assessment.

Besides the practical difficulties of examining complex research questions, mixed methods research has a particular set of obstacles. For instance, Onwuegbuzie and Collins (2007) identified integration as a barrier to mixed methods research. Integration is the process of combining and interpreting quantitative and qualitative data together to synthesize conclusions (Onwuegbuzie & Collins). Moreover, one facet underlying the problem of integration is prioritization. Researchers encounter this problem as they make decisions about which data source, quantitative or qualitative, to emphasize (Onwuegbuzie & Collins). Nevertheless, researchers reduce these difficulties by selecting appropriate mixed methods designs (Onwuegbuzie & Collins). In this study, the investigator chose a sequential explanatory design to reduce problems associated with integration and prioritization.

A sequential explanatory mixed methods study is a two-phase process with the purpose of explaining or building upon initial quantitative results (Creswell & Clark, 2007). Explanatory design is the least complicated variant of mixed methods (Tashakkori & Teddlie, 2003). The first phase begins with the collection and analysis of quantitative

data (Creswell & Clark). In this study, the investigator assigned priority to the quantitative phase. The investigator did this because the problem under examination involved improving student achievement according to a content-specific test. Moreover, the investigator analyzed the results of this test using quantitative procedures.

The second phase, which is qualitative, further supports and explains the quantitative findings. As part of sequential explanatory design, the investigator adjusts the data collection procedures of the second phase to better interpret the findings from the first phase (Creswell & Clark, 2007). In this study, this meant making changes to the interview questions as the investigator gathered and analyzed quantitative data. Aldridge, Fraser, and Huang (1999) used similar procedures to adjust the second phase of their study, which involved assessing student perceptions of school environments in America and Taiwan.

Figure 2 shows the order and components of sequential explanatory design. The first square labeled QUAN represents quantitative methods and data collection. In addition, Creswell and Clark (2007) suggested capitalizing QUAN to show priority over qual, which represents the collection and analysis of qualitative data. Furthermore, this design uses the qualitative portion to explain quantitative results, which means that the qualitative phase occurs after the investigator analyzes data from the quantitative phase (Creswell & Clark; Tashakkori & Teddlie, 2003).

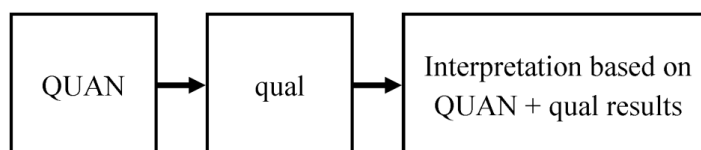


Figure 2. Sequential explanatory mixed methods design with notation according to Creswell and Clark (2007).

Creswell and Clark (2007) also stated that sequential mixed methods is less likely to bias results of the second phase in comparison to a concurrent approach. Moreover, in this study, the purpose of the qualitative phase was to explain quantitative results. The investigator reduced the potential for bias by applying each phase in sequence and establishing the purpose and priority of the second phase before collecting any data. Furthermore, since this study involved intact classrooms, the generalizeability of results was limited regardless of the methodological approach taken (Campbell & Stanley, 1963; Gall et al., 2003).

The quantitative sources of data for this study included an investigator made content-specific test (CS) and student questionnaire (SQ). Although the investigator considered alternative tests and scales for use in this study, none matched the curriculum needs of the participating teachers or the factors of interest as they related to the research questions. For instance, searching Mental Measurements Yearbook database showed two social studies tests, each pertaining to economics content. The teachers and students in this study were required to cover Washington State history according to school district expectations.

The investigator encountered similar issues when searching for preestablished scales. First, some scales measured dimensions that were too general, such as the Classroom Environment Scale (Trickett & Moos, 2002). Other scales measured dimensions that were unique to a particular learning situation, such as the one developed by Song et al. (2006), which assessed student attitudes in the context of problem based learning environments.

Slavin (1992) suggested that investigator made content-specific tests and questionnaires are suitable for measuring student achievement and attitudes in the context of educational research. However, Slavin recommended that investigators take appropriate steps to establish the reliability and validity of these measures, such as conducting pilot tests.

For the second phase of this study, the qualitative source of data included interviews of students from low, medium, and high achievement designations according to multiple measures. The investigator chose standardized open-ended interviews (Fontana & Frey, 2005; Patton, 1990) to explain the results of the CS and SQ and triangulate results overall. Moreover, the investigator used interview data to explore additional characteristics of reflective thinking that the SQ did not measure.

Population and Sample

The district where the investigator conducted the study is located in a town 12 miles from a large city in Washington State. This district has approximately 6,700 students. This study took place at one of the district's middle schools. This middle school has 788 students, 29% of whom receive free or reduced price meals, 13% are designated learning disabled, and 3% are transitional bilingual (Office of Superintendent of Public Instruction, 2008).

Participants for this study included 259 seventh grade students, 126 female and 133 male. Twenty-three of these students receive special education services and four students have learning disability 504 plans. Seven students are English language learners and two of these seven receive special education services. Twenty-nine percent or 75 students receive free or reduced price lunches.

The design of this study did not include random assignment of students to groups but consisted of 10 intact classrooms. All 10 classrooms contained seventh grade students studying social studies content. In this study, the investigator administered 23 days of intervention. This was the number of days required to complete one unit of study. The title of this unit was trees, technology, and the environment. Moreover, the design involved three participating teachers. One of these teachers, who taught five of the 10 intact classrooms, was also the investigator.

The rationale for using these participants was twofold. First, these students were part of the investigator's regular classroom load, which facilitated supervision of the project. Second, previous studies have shown a need to research instructional strategies for promoting reflective thinking among seventh grade social studies students (Bianchi, 2007; Bond, 2003; Evans, 2009; Song et al., 2006; Winne, 2005). Previous studies explored this topic with other grades or in other settings, but few studies focused on middle school students in the context of social studies learning.

For the quantitative phase of this study 10 intact classrooms, consisting of 259 seventh grade students, completed the CS pretest, posttest, and retention test. During the CS posttest, students also completed the SQ. In addition, students in treatment, comparison, and control classrooms completed a retention test 12 weeks from the conclusion of the intervention period.

Sampling procedures for the qualitative phase consisted of purposeful sampling, which involved the selection of participants based on predetermined characteristics related to the research questions (Patton, 1990; Tashakkori & Teddlie, 2003). In this study, participant criteria included high, medium, and low achieving students, ranked

according to students' sixth grade Washington Assessment of Student Learning reading and math scores, and CS posttest scores. The investigator organized interview participants into three achievement categories, including low, medium, and high. Then, the investigator and another participating teacher selected 12 students. Three students from each of the four intact treatment classrooms were selected for participation. Each group of three students consisted of one low, one medium, and one high achieving student. These three types of students participated in student interviews simultaneously, during the same session.

Protection of Participants

The quantitative phase of this study involved typical classroom instruction and assessment procedures, which did not require informed consent from participants. Nevertheless, the investigator informed students' parents or guardians about the study using an informational letter; see Appendix A. The investigator prepared this letter with the help of the school's principal. Students carried this letter home before the study began.

For the qualitative phase of this study, the investigator conducted student interviews during the school day. Moreover, the investigator interviewed students during their social studies class to reduce the amount of missed instruction. Since conducting these interviews was an unusual activity, and because students left social studies class to participate, the investigator received informed consent from students' parents or guardians to conduct interviews. The letter sent home with students granting the investigator informed consent is located in Appendix A.

Neither students nor teachers received compensation for participating in this study. However, the benefits of the study to students included an increase in reflective writing, an increase in peer discussion regarding unit content, and an increase in understanding of Washington State history. The benefits to participating teachers included an increase in collaboration and an increase in teachers' use of different instructional strategies. Participating in this study posed no risks to students or teachers.

Last, the investigator ensured the privacy of participants' data by storing it on secure computers and replacing student and teacher names with numbers.

Intervention

Participating teachers applied the intervention for 23 consecutive days during the first two months of the school year. Again, this was the number of days required to complete one unit of study. Each application of the intervention lasted for approximately 10 minutes per class period. Class periods last for 50 minutes at the school where this study took place. Teachers applied the intervention at the conclusion of each class, at which time students wrote, and sometimes illustrated, a reflection in a journal along with a prompt. In this study, students wrote the lesson objective, or a slight variation of it, as their reflective prompt. Upon writing their reflections, participating teachers instructed students to read their reflections aloud to a nearby peer. At the beginning of the next application of intervention, teachers chose three to four student reflections to read to the class. The purpose of doing this was to model characteristics of exemplary reflections.

During the first five days of the intervention, students received strategy instruction regarding the contents of a model reflection. These instructions included (a) reading the lesson objective from the dry board, (b) thinking about one's learning in

comparison to the objective, (c) writing or drawing a description of one's learning, and (d) comparing one's reflection to a peer's reflection by reading it aloud.

Last, participating teachers collected students' reflection journals each day and assessed their content. To assess student journals, teachers wrote different types of comments depending on the student's reflective response. Comment types included questioning, correcting, and approving of students' thinking. The investigator did not analyze the reliability of these comments.

Students who did not receive the treatment engaged in additional practice of lesson content. Marzano, Pickering, and Pollock (2001), Joyce and Weil (2004), and Mager (1984) argued that engaging students in practice activities is one method for improving student learning. For example, comparison classrooms read textbook passages as a class and summarized the contents. At another time, comparison classrooms practiced the steps for analyzing artifacts by examining Native American woodworking tools.

Quantitative Phase

In the quantitative phase of this study, the investigator applied a within-teacher random assignment design (Slavin, 1992) to match four pairs of classrooms according to multiple achievement measures. The investigator used CS pretest scores and Washington Assessment of Student Learning reading and math scores to match eight intact classrooms into four pairs of planned contrasts. After organizing classrooms into four pairs, the investigator assigned comparison or treatment status randomly to each class. Overall, the investigator analyzed four pairs of planned contrasts, each consisting of a

treatment and comparison classroom, and two control classrooms for a total to 10 intact classrooms.

Analysis of variance results on CS pretest scores showed nonsignificant results between all 10 classrooms at a significance level of .05, $F(9, 239) = .44, p = .91$.

Likewise, contrasts of Washington Assessment of Student Learning reading and math scores produced nonsignificant results between all 10 intact classrooms at a significance level of .05, $F(9, 212) = 1.07, p = .39$ and $F(9, 212) = 1.20, p = .30$, respectively.

Moreover, the investigator conducted the same analyses after removing special education students, $n = 23$, and English language learners, $n = 7$. Analyses conducted without these students produced nonsignificant results between planned contrasts across CS pretest, reading, and math scores.

Control classrooms were not randomly assigned their designation. This was the case because there were three teachers involved in the study and one of these teachers joined the project after initial planning had taken place. The late arrival of one of the participating teachers was due to school scheduling requirements. The investigator assigned control classrooms to the teacher who joined the study late for convenience.

Table 1 shows each planned contrast and the different measures the investigator used to match pairs of classrooms. In this study, the investigator defined planned contrasts as two classrooms paired together according to similar performance on multiple achievement measures. Again, the investigator randomly assigned treatment and comparison designations for each pair. In addition, Table 1 shows the number of participants in each class along with control group data. The investigator did not randomly assign control groups their designations.

Table 1

Matched Pairs of Classrooms Forming Planned Contrasts and Control Classroom Data

Designation	Comparison 1	Treatment 1
<i>n</i>	24	31
CS Pretest Mean	16.39	16.31
Reading WASL Mean	415.90	416.89
Math WASL Mean	411.29	411.78
Designation	Comparison 2	Treatment 2
<i>n</i>	23	29
CS Pretest Mean	16.43	16.71
Reading WASL Mean	413.67	411.56
Math WASL Mean	405.17	405.85
Designation	Comparison 3	Treatment 3
<i>n</i>	28	29
CS Pretest Mean	16.54	17.66
Reading WASL Mean	415.84	403.09
Math WASL Mean	400.76	402.57
Designation	Comparison 4	Treatment 4
<i>n</i>	23	24
CS Pretest Mean	15.96	15.65
Reading WASL Mean	406.67	408.78
Math WASL Mean	393.94	390.96
Designation	Control 1	Control 2
<i>n</i>	23	25
CS Pretest Mean	17.33	17.79
Reading WASL Mean	409.17	418.91
Math WASL Mean	396.39	417.77

Note. CS is an abbreviation for content-specific and WASL is an abbreviation for Washington Assessment of Student Learning.

Along with organizing planned contrasts between four pairs of classrooms, the investigator examined attendance records for comparison and treatment groups during the 23-day study period. Contrasts of student absences for treatment and comparison classrooms showed nonsignificant results at an alpha level of .05, $F(7, 203) = .57, p = .78$.

After organizing four pairs of planned contrasts, the investigator randomly assigned designations to the eight classrooms taught by two teachers involved in the study, teacher A and teacher B. Teacher A received four comparison designations and one treatment designation. Teacher B received three treatment designations. Teacher C taught two social studies classes, which the investigator nonrandomly designated as control classrooms. In summary, the investigator arranged four pairs of planned contrasts; three of these occurred between teacher A and B. The fourth planned contrast occurred between two classrooms taught by teacher A.

Table 2 summarizes the designations of treatment, comparison, and control classrooms for each participating teacher. In addition, Table 2 shows that two of the planned contrasts occurred at the same time of day, 12:45 p.m. and 1:40 p.m.

Table 2

Teacher Assignments for Each Designation and Class Start Times

Class Time	Teacher A	Teacher B	Teacher C
8:00 a.m.	Comparison 1		Control 1
8:56 a.m.	Comparison 2		Control 2
9:46 a.m.	Treatment 1	Treatment 2	
12:45 p.m.	Comparison 3	Treatment 3	
1:40 p.m.	Comparison 4	Treatment 4	

Students in each of the 10 intact classrooms completed the CS pretest, posttest and retention test. The investigator administered the retention test 12 weeks from the conclusion of the intervention period. Students in control classrooms experienced the same lessons used by teacher A and B during the 23-day study period only after completing the CS posttest. Students in control classrooms did not engage in the intervention or additional practice before taking the retention test. However, students in control classrooms did experience the same content as students in treatment and comparison classrooms, according to the design of scripted lessons. Table 3 shows the testing sequence applied by the investigator. In addition, in Table 3, X₁ represents the intervention and X₂ represents additional practice.

Table 3

Testing Sequence and Descriptive Data for Treatment, Comparison, and Control Designations

Designation	No. Classrooms	<i>n</i>	CS Pretest	Intervention	CS Posttest	CS Retention Test
Treatment	4	98	O	X ₁	O	O
Comparison	4	113	O	X ₂	O	O
Control	2	48	O		O	O

Note. CS is an abbreviation for content-specific.

Content-Specific Test

For this study, the investigator used a content-specific test as the dependent variable (Slavin, 1992). This test measured students' knowledge of Washington State history and related topics such as Native American logging techniques, influence of geography on human activity, and economic benefits of the lumber industry. Students in

comparison and treatment classrooms received 23 days of instruction relating to the content of this test. Students in the control classrooms received 23 days of instruction in an alternate unit, which did not include content relating to the content-specific (CS) pretest-posttest.

Pilot Testing the Content-Specific Test

The investigator pilot tested the CS test with 93 seventh grade students during the previous school year. This test consisted of 34 multiple-choice items. Students completed the CS pilot test online. According to Cronbach's alpha, the reliability of this test was .75 while the Spearman-Brown coefficient for split halves was .71.

Administration

Participating teachers administered the CS test online in two computer labs at the school where the investigator was conducting the study. See Appendix A for a sample of CS test questions in their online format.

Before beginning the test, each teacher read a set of instructions to students. These instructions included information regarding participation, grading, and the importance of the test, among other topics. See Appendix A for CS pretest and posttest administration instructions.

Because students took the CS test online, they were seated in the computer lab next to each other. To increase test security, the investigator created two equivalent forms for each CS test administration. For example, the investigator made CS posttest form A and CS posttest form B. These forms were identical except for the order of the questions. Teachers assigned form A or form B to students in an alternating pattern, such as A B A B. In addition, the investigator used a computerized test-writing program that randomized

the answers for each multiple-choice question. In summary, students sitting next to one another during CS testing took equivalent but alternate forms of the test and saw answers to each question in random order.

Reliability of the Content-Specific Test

The CS pretest, posttest, and retention test were identical and consisted of 38 items, four more items than the pilot test. The investigator and participating teachers added four items to adjust for changes made to lessons prior to the beginning of the study. According to Cronbach's alpha, the reliability of the CS pretest was .71, $n = 237$. Likewise, the Spearman-Brown coefficient for split halves was .71. The reliability of the CS posttest, according to Cronbach's alpha, was .83 while the Spearman-Brown coefficient for split halves was .80, $n = 250$. The CS retention test showed similar results with regard to reliability, $\alpha = .80$, $n = 242$.

The investigator calculated gain scores as the difference between CS posttest and CS pretest results for each student. Furthermore, Table 4 shows skewness and kurtosis values for pretest, posttest, retention test, and gain scores. These values ranged between .14 and .60. These values showed sufficiently normal distributions so that the investigator tested differences using parametric statistics (Green & Salkind, 2005). The minimum and maximum scores on the pretest were 5 and 29. For the posttest, the minimum was 6 and the maximum was 37. In addition, 251 out of 257 students scored less than 35 out of 38 points on the posttest. These values suggested that the CS pretest and CS posttest did not demonstrate ceiling and floor effects with regard to students' test performance. Results for the retention test showed similar results.

Table 4

Descriptive Statistics for the Content-Specific Test and Gain Scores

	CS Pretest	CS Posttest	CS Retention Test	Gain Scores
<i>n</i>	249	257	241	247
Missing	10	2	18	12
<i>M</i>	16.69	23.64	23.91	6.91
<i>SD</i>	5.24	6.50	6.03	5.65
Skewness	0.22	-0.29	-0.49	-0.15
<i>SE</i> of Skewness	0.15	0.15	0.16	0.16
Kurtosis	-0.52	-0.60	-0.47	-0.14
<i>SE</i> of Kurtosis	0.31	0.30	0.31	0.31
Minimum	5	6	8	-8
Maximum	29	37	36	21

Note. CS is an abbreviation for content-specific.

Validity of the Content-Specific Test

This study consisted of 23 days of intervention for one unit of study. This unit of study consisted of 23 lessons pertaining to seventh grade social studies. For each lesson covered in the unit, there were one or two test items. For instance, the objective of lesson number 12 was to have students define the word mechanization; consequently, the test item corresponding to this objective asked students to identify the definition of mechanization. Each lesson objective and its corresponding test item were refined during the pilot study of the CS test and subsequent planning for the unit. Results from the retention test showed that students in control classrooms produced similar scores in

contrast to treatment and comparison classrooms. Students in control classrooms experienced the same set of scripted lessons used during the 23-day study before taking the retention test. This outcome suggests that the lessons used in this study corresponded to CS test items.

Netemeyer, Bearden, and Sharma (2003) wrote that researchers typically assess the face validity of a test between applications. In this study, the investigator assessed face validity according to students' observed fluency of working with the CS pilot and posttest. In each application, students completed the test in the allotted time. In addition, the Flesh-Kincaid grade level readability statistic for the posttest test was 5.1. This implies that the test appeared to students in an easy to use and readable format.

Netemeyer et al. (2003) also said that content validity is based upon thoughtful item generation and judging efforts. In this study, the investigator constructed CS test items to correspond to each of the 23 lessons used by teachers during the 23-day intervention period. In this way, the test covered the target construct (Netemeyer et al., 2003).

Last, Netemeyer et al. (2003) stated that researchers prove convergent validity by showing correlations between existing measures. In this study, the investigator compared the results of the CS posttest for students in treatment and comparison designations to Washington Assessment of Student Learning reading and math scores. Results showed significant correlations between posttest scores and reading and math scores at a significance level of .01 using two-tailed tests. The correlation between CS posttest scores and reading scores was $r(179) = .68, p < .01$. For math scores the correlation was

$r(179) = .72, p < .01$. These results suggest that the CS posttest showed convergent validity when compared to assessments of general aptitude.

Student Questionnaire

Along with the CS test, the investigator created a student questionnaire (SQ) to examine differences in students' perceptions regarding classroom instruction. The investigator used the SQ to explain CS posttest results and as a source of information for creating student interview questions in the qualitative phase of the study. The initial goal of the SQ was to measure each characteristic examined in the previous chapter, including continuity, explicitness, interaction, and metacognition. However, analysis of the SQ showed only two factors. These factors represented facets of metacognition and interaction and the investigator named these Self-Assessment and Teacher Interaction, according to the content of their corresponding items.

Netemeyer et al. (2003) identified several sources for creating items. These include prior research, open-ended elicitation from samples of representative subjects, and researcher-generated statements based on the researcher's knowledge. In this study, the investigator initially designed items according to prior research and knowledge of the constructs under examination. Then, the investigator refined these items according to pilot test results. In this study, the sources of SQ items came from a combination of these sources including prior research, investigator's understanding, and pilot test results.

Pilot Testing the Student Questionnaire

The investigator pilot tested the SQ with 210 seventh grade students during the previous school year. Although the questionnaire for the pilot test began with 11 items, the final version had six. The investigator assigned a five-point Likert scale to each item;

5 = strongly agree, 4 = agree, 3 = neutral, 2 = disagree, and 1 = strongly disagree.

Cronbach's alpha for these six pilot test items was .78. Students took the SQ online at the end of the CS pilot test.

The Kaiser-Meyer-Olkin statistic for the SQ pilot was .73, while Bartlett's test of sphericity was significant at the .05 level, $df = 15$, $p < .01$. Moreover, the correlation matrix for the SQ showed correlations between .14 and .70. Pett et al. (2003) recommended that item correlations for questionnaires range between .30 and .80.

The investigator chose to extract factors using principal components analysis because it efficiently summarizes data sets (Pett et al., 2003). In addition, this method extracted 69% of the total variance for three factors during pilot testing analysis. Pett et al. wrote that factors extracted using principal components analysis account for at least 50% of shared variance. Furthermore, Pett et al. recommended this standard since error variance contributes to shared variance when using principal components extraction method.

The investigator rotated two factors, according to scree plot results and eigenvalues greater than one. The investigator used Varimax rotation to simplify factors since this method produces interpretable solutions and assumes independent subscales (Pett et al., 2003). Analysis of SQ pilot test results showed two factors with eigenvalues of 2.84 and 1.44. These factors accounted for 71% of shared variance.

Administration

For this study, the investigator placed SQ questions at the end of the CS posttest. Students completed the SQ in the same setting with the same set of instructions as those

given for the CS posttest; see Appendix A for a sample of SQ questions in their online format.

Reliability of the Student Questionnaire

Final analysis of the SQ showed Cronbach's reliability of .80. The Kaiser-Meyer-Olkin statistic for the SQ was .80 and Bartlett's test of sphericity was significant at the .05 level, $df = 15$, $p < .01$. Correlations between items ranged between .29 and .64. Additional data, including means and standard deviations, for each SQ item are located in Appendix A.

The investigator extracted two factors with eigenvalues greater than 1 and according to scree plot results. Similar to pilot trials of the SQ, the investigator used principal components analysis with Varimax rotation. Results showed two factors with eigenvalues of 3.04 and 1.03 accounting for 68% of shared variance. The rotated factor matrix for this analysis is located in Appendix A.

Validity of the Student Questionnaire

Netemeyer et al. (2003) stated that researchers prove content validity by showing that elements of a measurement represent the target constructs. The investigator used two approaches to coordinate SQ items with the constructs under examination.

First, the investigator linked items to elements of the intervention. In this study, the intervention included lesson objectives as reflective prompts and teacher feedback in the form of written comments. For each day during the 23-day intervention period, participating teachers engaged students in each of these elements. For instance, students wrote lesson objectives as part of their reflective response to guide reflection. In addition, teachers collected students' reflection journals at the end of each class period and

provided feedback in the form of written comments. Items on the SQ included words that corresponded to these elements of the intervention.

Second, results from pilot test and final administration analysis showed the same factors between two different groups of seventh grade social studies students. Each factor contained items that related to one another as well as to its associated factor overall. See Appendix A for the rotated factor matrix for the student questionnaire. The investigator named one factor Self-Assessment because items asked about student learning. For example, one item read, “I ask myself if I have met the learning target for each lesson.” The investigator named the second factor Teacher Interaction because the items in this group incorporated questions about teacher-student interaction. For example, two items reference teacher feedback and teacher comments.

With regard to face validity, the investigator gave the SQ online in the same format as CS tests. The Flesh-Kincaid grade level readability statistic for the questionnaire was 5.8. Pilot test results and final administration results showed that students found SQ items easy to answer and read.

Last, Tashakkori and Teddlie (2003) identified triangulation as the combination of the results of two or more studies to provide a comprehensive picture of results overall. In this study, the investigator triangulated results between measures to show convergent validity of the SQ.

First, the investigator compared the results of the SQ to transcript data gathered during the qualitative phase of this study. Results showed that students were able to answer questions about related questionnaire constructs, including lesson objectives as reflective prompts and teacher feedback. Categories derived from transcripts also showed

that students emphasized teacher feedback as a useful source of information that guided their learning. However, interview participants deemphasized lesson objectives as learning cues. These results, discussed in depth in the next chapter, converged with SQ results overall. Second, statistical analysis showed a significant correlation between CS posttest performance and factor scores for Teacher Interaction from the SQ. In this way, the SQ demonstrated convergent validity through the triangulation of CS posttest results, SQ results, and categories derived from student interviews.

Qualitative Phase

One rationale for applying sequential explanatory mixed methods design is to use the second phase to explain the results of the first phase (Creswell & Clark, 2007). In this study, the investigator interviewed students using standardized open-ended interviews (Fontana & Frey, 2005; Patton, 1990) to explain the results of the CS and SQ. Another purpose of the interviews was to examine students' perspectives regarding characteristics of reflective thinking according to the third research question and hypothesis of this study.

Overall, the method of the qualitative phase was phenomenological. Patton (1990) defined this approach as focused inquiry to understand the structure and essence of someone's experience. In this case, the investigator interviewed student participants to understand their perspective regarding characteristics of reflective thinking.

The investigator chose students for participation based on three criteria. First, in cooperation with another teacher involved in the study, the investigator chose participants for their ability to communicate openly with an adult and provide a rich source of information (Patton, 1990). Second, the investigator sorted students from each treatment classroom into three achievement groups, including, low, medium, and high. The

investigator established these groupings according to Washington Assessment of Student Learning reading and math scores and CS posttest scores. Third, the investigator chose participants according to additional characteristics to represent each gender, along with special and general education students. Specifically, there were five males and seven females. Furthermore, there was one male special education student and one female special education student.

Richards (2005) suggested that descriptive coding is appropriate for classifying participants according to their attributes. In this study, descriptive coding was limited to high, medium, and low achievement designations. The purpose for applying this restriction was to focus on answering the third research question, which investigates differences between students according to their achievement level.

In summary, interview participants included 12 students total, four students each from low, medium, and high achievement groupings. There were seven female and five male students. Two special education students participated, one female and one male.

Administration

Interviews began two weeks after students completed the CS posttest and SQ. The investigator interviewed 12 students in the library of the school where the study took place. Interviews took one day to complete and each interview session lasted between 7 and 10 minutes. The investigator recorded interviews using a digital audio device.

Students interviewed in groups of three to promote information-rich responses (Patton, 1990). In addition, nine of the students had never met the investigator conducting the interviews. The investigator and participating teacher concluded that interviewing students in groups of three would promote student comfort with the interview process and

encourage discussion. Each group of three consisted of one low, one medium, and one high achieving student from the same treatment classroom. Three of the nine students were part of the investigator's regular classroom assignment.

Before the beginning of each interview, the investigator read an explanation to participants. The explanation read as follows:

Mrs. M and I have asked you to be here so that we can improve our teaching and learning in social studies. We think that you have an important perspective about how we can make that happen. And that's what the interview is about: We want to talk to you about ways to make class better. Also, we thought it was a good idea to have three of you together so that you could share ideas; we are just asking that each of you listens to the other and does your best to contribute. Today, we have four questions. The interview will last about 15 minutes and we are talking with three kids from each class, about 12 total in case you were curious. There are two recorders here; one is digital and the other uses a tape, in case one doesn't work.

Any questions before we begin?

Interview Questions

The purpose of the qualitative phase of this study was to explain quantitative results and to investigate the four characteristics of reflective thinking, including interaction, continuity, explicitness, and metacognition. Moreover, the investigator wrote each question with the intent of examining characteristics of reflection from the perspective of students. For instance, question two asked about the kind of feedback that students perceive as helpful for their learning. Feedback from teachers is one facet of interaction, as discussed in chapter two. Likewise, the purpose of question three was to

determine what students thought about using reflection journals as a method for summarizing their thinking and making connections between lesson content. Similar to the relationship between feedback and interaction, summary and making connections are facets of continuity. Last, the intent of question one and four was to explore students' perceptions of explicitness and metacognition, respectively.

The interview questions read as follows:

1. How do you know what you are supposed to learn in social studies class each day?
2. What kind of feedback was most helpful for your learning?
3. How did writing reflections in your journal help you summarize or make connections in the lesson?
4. How do you know when you have successfully learned something?

The investigator edited these questions into their shown form several months before collecting any data. First, the investigator wrote nine questions during the previous school year and pilot tested these with the same group of seventh grade students who pilot tested the CS and SQ. Pilot test results showed that students fluently discussed and wrote answers to these preliminary questions. However, analysis of the SQ showed that it measured fewer factors than anticipated. As a result, the investigator reduced the number of interview questions from nine to four to coincide with factors from the SQ. In addition, the investigator refined interview questions to assist in explaining results of the CS posttest.

Creswell and Clark (2007) suggested that adjusting the second phase of an explanatory mixed methods design facilitates interpretation of initial results. In this study,

this meant reducing the number of interview questions and focusing them on results of the SQ and CS test. This also meant preserving substantive content from the preliminary list of questions to investigate the third research question and hypothesis.

Credibility of Qualitative Data

In this study, the investigator interviewed participants using standardized open-ended interview methods (Fontana & Frey, 2005; Patton, 1990). Using these methods meant that each participant heard the same set of interview questions in the same order. Patton suggested that standardized open-ended interviews are inflexible, not allowing the interviewer to pursue unanticipated topics. Alternatively, Patton also stated that the collection of interview data using this format minimizes issues of credibility since each participant answers the same question, in the same order.

The investigator transcribed interview recordings within two days of conducting interviews. Following this, the investigator read transcripts and identified topic codes (Richards, 2005). The investigator then reexamined transcripts along with accompanying topic codes and identified analytical codes (Richards, 2005). The investigator further summarized transcript data from these codes into categories. See Appendix A for a sample of transcript data, topic codes, and category. As a final step, member checking procedures were conducted to determine whether interview participants agreed or disagreed with the investigator's interpretation of categories (Richards, 2005).

Richards (2005) defined member checking as a technique for producing feedback regarding a researcher's interpretation of data. As an example, Werderich (2006) conducted follow-up interviews as a method of member checking to validate categories in

a study that examined teachers' use of feedback in middle school students' dialogue journals.

In this study, the investigator conducted member checking procedures by having interview participants answer eight follow-up questions. The investigator wrote two questions for each category identified from transcript coding. Responses from these follow-up questions were assessed using an artificial dichotomous score (Gall et al., 2003). Limiting participants' responses with artificial dichotomous scores allowed the investigator to display results as percentages. Taking this step also simplified students' range of responses and increased the interpretability of member checking results. See Appendix A for member checking follow-up questions and instructions.

An examination of responses to follow-up questions showed that students agreed with the investigator's interpretation between 67% and 100% of the time, $n = 12$. According to Table 5, students agreed more often with statements regarding feedback and knowing what to learn in class. Table 5 also shows that students were less likely to agree with the investigator's interpretation about indicators of successful learning. Overall, the investigator focused the results section of this study on conclusions relating to the first four questions from the member checking questionnaire. These questions showed that interview participants generally agreed with the investigator's interpretation. The investigator made fewer claims for categories represented by the last four follow-up questions since participants showed lower rates of agreement.

Table 5

Follow-Up Questions and Student Perspective in Comparison to Investigator's Interpretation

Question	Student Response
I know what I'm supposed to learn in social studies from lots of things, like the learning target, agenda, bell work [entry question], and the activities that the teacher has me do.	100%
I know what to learn in social studies by thinking about the whole lesson, not just one part.	100%
The most useful kind of feedback is constructive because it helps me know what I did right or wrong.	92%
Good feedback helps me to know what I need to improve on or if I'm doing something right.	92%
Writing, telling, or drawing my reflections helps me show or tell about my thinking.	75%
Making a reflection means that I can show whatever I'm thinking or learned from a lesson.	83%
I know when I have learned something because it's just right there in my mind and I don't have to think too hard to answer a question about it.	67%
I can tell when I have learned something because it will come to my mind later when I need to use it.	75%

Additional Description of Study Conditions

Lichtman (2006) recommended that qualitative researchers keep a reflection journal to examine during and after data collection to inform analyses and conclusions. Consequently, the investigator wrote daily reflections related to the study during the 23-day intervention period. The subject of reflective entries included processes, problems, and questions that were related to the study (Lichtman, 2006; Richards, 2005). In addition,

entries did not focus exclusively on the qualitative phase. For instance, one entry described an interruption to the intervention, when specialists tested students' hearing and vision. This prevented one treatment group and one comparison group from completing the planned lesson.

In another entry, the journal showed that teacher A, also the investigator, completed planned lessons in less time in comparison to teacher B, especially at the beginning of the study. As a result, the investigator adjusted lesson pacing.

In a third entry, the investigator's journal showed entries about differences between students' reflections. Even though students experienced the same lesson and wrote the same learning objective as a reflective prompt, students' journal entries showed a wide range of content. This phenomenon raised questions regarding convergent and divergent thinking (Ellis, 2001). An example of convergent knowledge includes skill development such as memorizing history facts. Alternatively, an example of divergent knowledge includes sharing one's opinion about a historical event.

Internal and External Threats

The design of this study was quasi-experimental since planned contrasts involved nonequivalent groups. Additional characteristics relating to the quantitative phase incorporated control groups with pretesting and posttesting. Campbell and Stanley (1963) suggested that the selection of participants is a threat to internal validity with nonequivalent control group design. However, by organizing four pairs of planned contrasts according to multiple achievement measures the investigator reduced this threat as a source of invalidity (Slavin, 1992). Nevertheless, only one contrasting pair, treatment 1 and comparison 1, involved the same teacher. As such, only this pairing controlled for

teacher effects. The remaining pairs contrasted treatment groups and comparison groups for different teachers according to Table 2. Likewise, experimenter effect is another threat to internal validity since teacher A was also the investigator.

Campbell and Stanley (1963) identified the interaction of testing and the intervention as a threat to external validity. However, analysis of gain scores for the control classrooms overall showed a mean change of minus .14 points. This small change indicated that students in this study did not demonstrate increased achievement from pretest sensitization.

Finally, Campbell and Stanley (1963) identified regression, interaction of selection and the intervention, and reactive arrangements as additional areas of concern regarding external validity with nonequivalent control group design. For example, the investigator accounted for absence rates between each pair of planned contrasts, but not for other variables such as frequency of discipline referrals across classrooms. Similarly, with regard to reactive arrangements, two of the planned contrasts occurred at the same time of day, while two contrasts occurred at a different time of day.

A final consideration with regard to the external validity of this study relates to the use of standardized open-ended interviews in the qualitative phase. As Patton (1990) suggested, standardized open-ended interviews are inflexible and do not allow the investigator to pursue unanticipated topics. Furthermore, the investigator wrote interview questions to examine specific characteristics of reflection. The combination of utilizing these rigid procedures may have forced interview participants to answer in ways that tended toward the investigator's bias regarding the results of the study.

Fontana and Frey (2005) summarized three sources of error stemming from this type of structured interviewing. First, participants may deliberately try to please the interviewer, also known as a reactive arrangement (Campbell & Stanley, 1963). Second, error can result from inconsistent administration of interview procedures. Finally, data from structured interviews may include error because the investigator is unable to communicate clearly with participants.

In summary, in this study the investigator controlled for some threats to internal and external validity, such as test sensitization. However, the investigator did not control for other threats, primarily those dealing with the selection of students to treatment, comparison, and control classroom designations. In addition, the qualitative phase of this study involved standardized open-ended interview procedures. Such procedures increase the credibility of interview data, but introduce the possibility of error from various sources, such as reactive arrangements (Fontana & Frey, 2005; Patton, 1990).

Summary

The investigator of this study analyzed the effects of reflective thinking on seventh grade social studies students, $N = 259$. Moreover, the investigator chose a sequential explanatory mixed methods design to answer the research questions. This type of design involves a quantitative and qualitative phase (Creswell & Clark 2007; Tashakkori & Teddlie, 2003). In the quantitative phase, the investigator organized four pairs of planned contrasts using within-teacher random assignment (Slavin, 1992). In addition, the investigator organized pairs according to multiple achievement measures. The study also incorporated two control classrooms.

Before beginning the study, the investigator administered the CS pretest to all participating students. Moreover, students in treatment classrooms experienced the intervention, which involved instructional practices characteristic of reflective thinking. Alternatively, students in the comparison classrooms engaged in additional practice. The intervention lasted for 23 days. At the conclusion of the intervention, all students completed the CS posttest and SQ. Furthermore, all students completed a retention test 12 weeks after the conclusion of the intervention period.

After analyzing quantitative results, the investigator interviewed 12 students using standardized open-ended interviews (Fontana & Frey, 2005; Patton, 1990). The investigator recorded interviews and transcribed them. Next, the investigator coded interview transcripts using topic and analytical methods (Richards, 2005). The purpose of this phase of the study was to explain quantitative results and explore characteristics of classroom instruction associated with the intervention, as well as gather data to investigate the third research question and hypothesis of this study.

Chapter Four: Results

Chapter Overview

This chapter presents the results of the study according to each research hypothesis. The first section examines results of the content-specific test (CS) and compares each statistical test that the investigator conducted. Section two presents results of the student questionnaire (SQ) and then describes the method the investigator used for calculating factor scores from the SQ. Moreover, this section describes results of the SQ for planned contrasts, pooled contrasts, and correlations between factor scores and CS posttest scores. The third section examines results of student interviews. In addition, this section organizes results according to the categories the investigator identified from interview transcripts. The chapter concludes with a summary.

Results for the First Research Hypothesis

The first research hypothesis of this study was that instructional practices characteristic of reflective thinking do not positively influence seventh grade students' academic achievement in social studies.

In this study, the researcher measured student achievement using a content-specific pretest and posttest. However, researchers have proposed different approaches when comparing nonequivalent groups using pretest-posttest measures (Dimitrov & Rumrill, 2003; Gall et al., 2003; Sheskin, 2007; Trochim, 2006). For instance, Sheskin recommended using the pretest as a covariate and the posttest as the dependent variable. Trochim recommended a similar approach except that the researcher adjusts pretest scores to account for pretest error. The formula recommended for making adjustments to

pretest scores is located in Appendix B. Similarly, Dimitrov and Rumrill recommended analyzing gain scores or analysis of covariance using the pretest as the covariate.

The investigator in this study conducted each test to explore the possibility that one statistical approach would produce different results in comparison to another. Moreover, the investigator assigned treatment, comparison, and control classrooms as the independent variable for each of these tests. For gain score analysis, the dependent variable was gain score group mean. The investigator calculated gain scores as the difference between CS pretest and CS posttest scores. For analysis of covariance, the dependent variable was CS posttest scores and the covariate was CS pretest scores or adjusted CS pretest scores (Trochim, 2006). Supplementary data tables for each of these tests are located in Appendix B.

Gain Scores

Before conducting statistical tests on gain scores, the investigator contrasted classrooms according to their homogeneity of variances using Levene's statistic. This test yielded no significant differences between gain score variances across classrooms according to ANOVA results at a significance level of .05, $F(9, 237) = .95, p = .49$.

Gain score tests began with a one-way analysis of variance to examine the relationship between planned contrasts according to gain score classroom means. The ANOVA was significant at the .05 level, $F(9, 237) = 15.86, p < .01$. The investigator conducted post hoc tests using Tukey *a* to control for Type I error (Sheskin, 2007). Post hoc tests showed that control classrooms scored significantly lower than treatment and comparison classrooms. However, all four planned contrasts between treatment and comparison classrooms showed nonsignificant results according to post hoc analyses. In

summary, gain score ANOVA revealed significantly lower scores for control classrooms, but nonsignificant differences between all planned contrasts for treatment and comparison classrooms.

Analysis of Covariance

Before conducting an analysis of covariance, the investigator contrasted pretest scores between classrooms to determine if there was an interaction effect, also known as the homogeneity-of-slopes test (Green & Salkind, 2005). Results showed that there was no significant interaction between pretest scores and classrooms at a significance level of .05, $F(9, 227) = .98, p = .46$. Likewise, Levene's statistic did not show a statistical difference between classroom error variances at the .05 level, $F(9, 237) = 1.01, p = .43$.

Analysis of covariance on pretest-posttest scores showed statistically significant results at a significance level of .05, $F(9, 236) = 15.81, p < .01$. The investigator conducted post hoc tests using Bonferroni contrasts to reduce the likelihood of Type I error (Sheskin, 2007). However, post hoc analyses applied a significance level of .10 to compensate for a reduction in statistical power (Sheskin). Nevertheless, ANCOVA results showed nonsignificant differences between planned contrasts for treatment and comparison classrooms. Alternatively, control classrooms scored significantly lower than treatment and comparison classrooms.

Last, the investigator conducted an analysis of covariance on adjusted pretest scores (Trochim, 2006). An examination of potential interaction effects for this test yielded identical results to analysis of covariance with unadjusted pretest scores, $F(9, 227) = .98, p = .46$. Likewise, Levene's test of equal error variance was identical to analysis using ANCOVA on unadjusted pretest scores, $F(9, 237) = 1.01, p = .43$.

Results from ANCOVA using adjusted pretest scores showed significant results at the .05 level, $F(9, 236) = 10.83, p = < .01$. However, post hoc tests using Bonferroni contrasts with the significance level set at .10 showed nonsignificant differences between planned contrasts for treatment and comparison classrooms. Again, control classrooms scored significantly lower than treatment and comparison classrooms.

There was one difference between ANCOVA using adjusted pretest scores in contrast to the other two statistical tests. On the posttest, the highest-mean control classroom was not statistically different in contrast to the lowest-mean treatment classroom. This outcome contradicts results using gain score ANOVA and ANCOVA with unadjusted pretest scores. Otherwise, these three tests yielded similar results with regard to tests of statistical significance for all omnibus F and post hoc contrasts.

In summary, this study found nonsignificant results between planned contrasts of treatment and comparison classrooms. However, treatment and comparison classrooms outperformed control classrooms. Generally, gain score ANOVA, ANCOVA using unadjusted pretest scores, and ANCOVA using adjusted pretest scores produced similar results regarding statistical significance. For instance, the omnibus F for gain score ANOVA and ANCOVA using unadjusted pretest scores was 15.86 and 15.81, respectively.

Table 6 summarizes the results of each statistical test conducted on CS test results by showing means and probability statistics for each planned contrast. In addition, Table 6 shows contrasts between the greatest-mean control classroom and the lowest-mean treatment and the lowest-mean comparison classroom. Additional values relating to each statistical test are located in Appendix B.

Table 6

Contrasts of CS Test Means According to Gain Score ANOVA and ANCOVA Analyses

	Comparison 1	Treatment 1	<i>p</i>
Gain Score Mean	9.39	10.83	.98
Posttest Mean, Pretest Covariate	25.78	27.14	1.00
Posttest Mean, Adjusted Pretest Covariate	25.54	26.81	1.00
	Comparison 2	Treatment 2	<i>p</i>
Gain Score Mean	8.19	7.86	1.00
Posttest Mean, Pretest Covariate	24.62	24.57	1.00
Posttest Mean, Adjusted Pretest Covariate	24.41	24.67	1.00
	Comparison 3	Treatment 3	<i>p</i>
Gain Score Mean	8.07	7.18	1.00
Posttest Mean, Pretest Covariate	24.61	25.21	1.00
Posttest Mean, Adjusted Pretest Covariate	24.52	26.02	1.00
	Comparison 4	Treatment 4	<i>p</i>
Gain Score Mean	8.48	7.39	.99
Posttest Mean, Pretest Covariate	24.43	23.04	1.00
Posttest Mean, Adjusted Pretest Covariate	23.73	22.02	1.00
	Greatest-Mean Control	Lowest-Mean Treatment	<i>p</i>
Gain Score Mean	0.50	7.18	< .01
Posttest Mean, Pretest Covariate	18.29	23.04	< .01
Posttest Mean, Adjusted Pretest Covariate	19.53	22.02	1.00
	Greatest-Mean Control	Lowest-Mean Comparison	<i>p</i>
Gain Score Mean	0.50	8.07	< .01
Posttest Mean, Pretest Covariate	18.29	24.43	< .01
Posttest Mean, Adjusted Pretest Covariate	19.53	23.73	.05

Note. CS is an abbreviation for content-specific. Gain score contrasts were conducted using Tukey *a* at the .05 level. Contrasts for ANCOVA were conducted using Bonferroni at the .10 level.

Figure 3 shows another representation of CS pretest-posttest results for each statistical test. Figure 3 also shows contrasts between the greatest-mean control classroom and the lowest-mean treatment and lowest-mean comparison classroom.

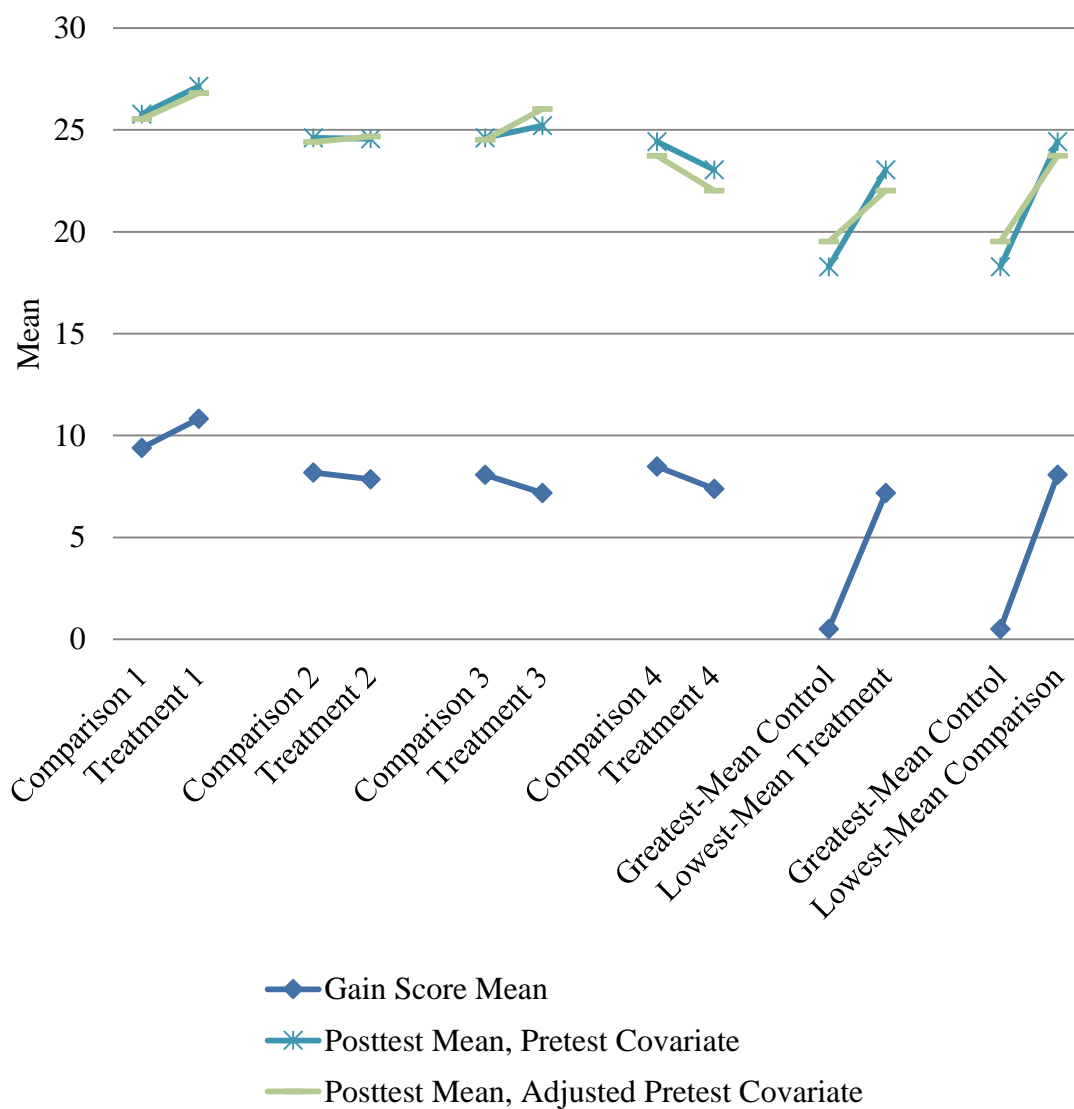


Figure 3. Summary of differences between planned contrasts and statistical tests.

Retention Test Results

Students from all 10 classrooms took a retention test 12 weeks from the conclusion of the intervention period. Before taking the retention test, students in the two control classrooms experienced the same lessons that teacher A and B used during the 23-day study. However, these students did not engage in the intervention or additional practice. Nevertheless, a one-way analysis of variance on retention test data for all 10 classrooms showed nonsignificant results at the .05 level, $F(9, 231) = 1.63, p = .11$.

The retention test counted as the third time that students had taken the CS test. Students may have sustained or improved upon their scores because of test sensitization (Campbell & Stanley, 1963). This could be the case even though students in control groups showed no gains between pretest and posttest administrations, as mentioned previously. Additional values relating to the retention test are located in Appendix B.

Results for the Second Research Hypothesis

The second research hypothesis of this study was that instructional practices characteristic of reflective thinking do not have an effect on seventh grade students' perceptions of classroom instruction in social studies.

To test this hypothesis, the investigator measured students' perceptions of classroom instruction using the SQ. According to the analysis discussed in the previous chapter, the SQ consisted of two factors, Self-Assessment and Teacher Interaction. Students who took the SQ ranked each item according to a 5-point Likert scale. Although this scale represents ordinal data, Grimm and Yarnold (1995) stated that researchers often analyze ordinal scales using parametric statistics. In addition, the investigator chose to summarize SQ results by calculating and analyzing factor scores. Pett et al. (2003)

suggested that calculating and analyzing factor scores is a parsimonious method for examining questionnaire results.

In this study, the investigator calculated two factor scores, one for Self-Assessment and one for Teacher Interaction. Additional calculations regarding factor scores include extracting factors using principal component analysis with Varimax rotation. Although there are alternative methods for calculating factor scores (Pett et al., 2003), the investigator chose regression because it is commonly used (Pett et al.). Moreover, alternative methods for calculating factor scores, such as Bartlett, produced identical results in comparison to regression.

Pett et al. (2003) recommended that a coefficient matrix for factor scores show low correlations between items that relate to different factors. Alternatively, Pett et al. also said that a coefficient matrix shows higher correlations between item scores and their related factors. The coefficient matrix for the calculated factor scores in this study indicated low correlations across factors, $r < .19$, and higher correlations between item scores and their related factors, $r > .35$. In summary, the factor score coefficient matrix replicated factor matrix correlations. Additional values related to this analysis, including the factor score coefficient matrix, are located in Appendix B.

Analysis of Factor Scores between Planned Contrasts

Before conducting statistical tests on factor scores for planned contrasts, the investigator analyzed classroom variances using Levene's statistic. Neither of the factor scores showed statistically significant differences at the .05 level; Self-Assessment, $F(9, 246) = 1.47, p = .16$; Teacher Interaction, $F(9, 246) = 1.05, p = .40$.

Analysis of variance results between planned contrasts of treatment, comparison, and control classrooms showed a significant difference for Self-Assessment at the .05 level, $F(9, 246) = 2.05, p = .03$. Again, the investigator conducted post hoc tests using Tukey *a* to control for Type I error (Sheskin, 2007). However, analysis of post hoc tests showed one statistically significant difference at the .05 level between the two control classrooms, control 1 and control 2. None of the planned contrasts between treatment and comparison classrooms for Self-Assessment produced statistically significant results. Likewise, analysis of Teacher Interaction factor scores between planned contrasts showed nonsignificant results at the .05 level, $F(9, 246) = 1.79, p = .07$.

Analysis of Factor Scores between Pooled Groups

The investigator conducted additional analysis of the SQ by pooling results from individual classrooms into three groups. These groups consisted of treatment, comparison, and control, $n = 112, n = 98, n = 46$, respectively. By pooling results together, the power of subsequent statistical tests increased, along with the probability of committing Type I error (Sheskin, 2007). To decrease the potential of Type I error, the investigator reduced the significance level to .01 for omnibus *F* and post hoc comparisons using pooled results on factor scores.

Before conducting statistical tests, the investigator examined pooled factor scores for similar variances using Levene's statistic. Analyses showed nonsignificant results at the .05 level; Self-Assessment, $F(2, 253) = .59, p = .56$, and Teacher Interaction, $F(2, 253) = .05, p = .95$.

Analysis of variance results indicated a significant difference for Teacher Interaction at a significance level of .01, $F(2, 253) = 6.53, p < .01$. Post hoc tests using

Tukey *a* with the significance level set at .01 showed a significant difference for the control group in contrast to treatment and comparison groups, $p < .01$. Analysis indicated that the control group had lower factor scores for Teacher Interaction. However, contrasts between treatment and comparison groups did not yield a significant difference at the .01 level using Tukey *a* as the post hoc contrast method, $p = .79$. In addition, analysis showed nonsignificant differences on Self-Assessment between each group. Supplementary values related to these analyses are located in Appendix B.

Factor Scores and Content-Specific Posttest Correlations

Pett et al. (2003) suggested comparing factor scores with other variables to evaluate the effects of interventions and produce information regarding the generalizability of findings. Consequently, the investigator also analyzed correlations between factor scores and CS posttest scores.

In order to control for Type I error the investigator divided the initial significance level by the number of correlations, $.05/6$ (Green & Salkind, 2005). This adjustment yielded a significance level of .008 using a two-tailed test. The results of the correlation analysis showed that Teacher Interaction was statistically significant in relation to CS posttest scores, $r(254) = .29, p < .001$. Analyzing the correlation between Teacher Interaction and CS posttest scores using Spearman's rho produced similar results, $r_s(254) = .29, p < .001$. According to Green and Salkind, researchers in the behavioral sciences generally view Pearson coefficients of .3 as medium. Squaring the coefficient showed an effect size of 8.41. Generally, students who did well on the CS posttest also indicated that they knew how they were doing on their social studies assignments and that they received feedback from their teacher at a significance level of .008, $r(254) = .29, p < .001$.

Alternatively, Self-Assessment did not show a statistically significant correlation with CS posttest scores at a significance level of .008, $r(254) = -.15$, $p = .02$, and $r_s(254) = -.14$, $p = .03$. Additional values, including a correlation scatter matrix of factor scores and CS posttest results, are located in Appendix B.

In summary, analysis of SQ factor scores showed nonsignificant differences between planned contrasts of treatment and comparison classrooms. However, analysis of pooled group contrasts showed significantly lower scores for the control group on Teacher Interaction at a significance level of .01 using Tukey *a* post hoc contrasts, $p < .01$. Last, analysis between Teacher Interaction factor scores and CS Posttest scores showed a statistically significant correlation at a significance level of .008, $r(254) = .29$, $p < .001$. This last result indicated that students who did well on the CS posttest also said that they knew how they were doing on their social studies assignments and that they received feedback from their teacher.

Results for the Third Research Hypothesis

The third research hypothesis of this study was that instructional practices characteristic of reflective thinking do not have a differential effect on low, medium, and high achieving seventh grade students in social studies. As mentioned, Lichtman (2006) said that research involving qualitative methods does not generally involve hypothesis testing. Similarly, Creswell and Clark (2007) stated that qualitative researchers only research questions and not hypotheses. However, the reason for including a hypothesis as part of the qualitative phase was to improve the integration of data and reporting of results from each of the two phases of the study.

To test this hypothesis, the investigator interviewed 12 students using standardized open-ended interviews (Fontana & Frey, 2005; Patton, 1990). After transcribing audio recordings of interviews, the investigator applied topic and analytical codes (Richards, 2005). Through the application of these codes, the investigator identified four categories. These categories were overall view, constructive feedback, self-expression, and fluency.

The first category, overall view, described students' understanding of what they were supposed to learn from each lesson. During interviews, students suggested that they knew what they were supposed to learn from a variety of sources. According to interview transcripts, some of these sources included the teacher telling students the learning objective, the teacher's tone of voice, and the activities that the teacher organized. Other sources included the entry question, lesson agenda, and directions on worksheets.

Analysis of transcript topic and analytical coding (Richards, 2005) did not show considerable differences between students of different achievement designations concerning this category. For example, one high achieving student said, "Mrs. M, she has like the orange folder where she puts her bell work [entry question] and it puts her agenda on it where it has like the bell work we do and then it tells us what we're going to be learning." A low achieving student, listening to this comment followed up, "yah, the learning target pretty much explains the whole lesson." Alternatively, a high achieving student from a separate interview stated, "I just don't really worry about what we're trying to learn today I just learn whatever is put in front of me."

Interview participants suggested that they knew what to learn during a lesson by gathering information from many sources. These sources included the lesson's agenda,

activities, and teacher instruction. Likewise, students mentioned that the learning objective was a useful guide for knowing what to learn. However, transcript data showed that students perceived the learning objective as a matter of procedure, rather than a monitoring cue to which they could compare their own thinking.

The second category, constructive feedback, described students' perspective about the kind of feedback that promotes learning. More than one student, from different achievement levels, suggested that constructive criticism was an important element of useful feedback. For instance, one medium achieving student stated, "like when she points out what you did wrong so you can work on it better." Again, a lower achieving student reported that helpful feedback included "good criticism that you're getting, like to help you out with what you're supposed to be doing in class."

Alternatively, students suggested that the length of written feedback was less important than the quality and thoughtfulness of it. Moreover, students said that helpful feedback gave them insight about whether what they were thinking was correct or incorrect. According to one medium achieving student, feedback should contain "a message" from the teacher. This message should confirm student understanding or correct it. As one high achieving student from another interview session stated, "[feedback] that agreed or that helped out a bit... how the feedback kind of gives more on the subject... [or] helps you know where you need to focus and improve on."

Overall, students across achievement levels stated that helpful feedback involved constructive elements. For instance, interview participants suggested that helpful feedback told them whether what they were thinking was correct or incorrect. Likewise,

the length of written feedback was less important to students in comparison to the message that it conveyed.

The third category, self-expression, described students' perspective regarding the purpose of reflection. During interviews, students reported a preference for writing, and illustrating reflections in their journals according to their own interpretation of what they learned. For example, one high achieving student stated, "I kind of liked the drawings as long as you could put your own thoughts." Likewise, a low achieving student in the same interview group reported, "[illustrating a reflection] let us be creative and draw."

Another instance of students' desire for self-expression in the context of reflecting involved an activity where students pretended to be loggers and wrote letters home. This particular reflective journal entry was popular with interview participants. When asked which journal entry participants thought was their best, four out of 12 chose this entry. In addition, one low achieving student stated, "we like writing in our journals because sometimes you have activities like writing a letter to your parents pretending you are a logger."

Alternatively, three participants reported feeling less interested in writing reflections in comparison to other forms of communication such as discussing and illustrating. For example, one low achieving student said, "sometimes it helps you to understand it [lesson content] and read it... and then write it down, but sometimes I just want to write it down real quick without thinking about it." Again, one high achieving student simply stated that, "I like more discussion."

Nevertheless, other students reported that writing reflections was helpful. For instance, one high achieving student said that, "it kind of locked the information in your

brain so you would remember it better.” In response to this comment, a low achieving student reported, “yah, cause at the end of the day when you were done learning and it reminds you of what you learned again.”

According to topic and analytical coding of transcripts, students across achievement levels reported that self-expression was an important part of reflecting. Moreover, students suggested that using various forms of communication helped them reflect in their journals. These forms of communication included writing, illustrating, and discussing. Generally, transcripts showed that students favored writing reflections the least. However, this was not the case when writing meant integrating creative elements, such as pretending to be someone else and writing a letter home to relatives.

The fourth category, fluency, described students’ response when asked how they knew they had successfully learned something. Students associated successful learning with a variety of activities, the most common of which was memory. For example, one low achieving student stated, “well, when you successfully learn something it won’t go away, you remember it and she [the teacher] mentions something and you remember it right.” Likewise, one medium achieving student in a different interview said, “like when you understand it all, what you’re learning about when you don’t really have to think, that you just have it right in mind.”

Although students frequently associated memory and successful learning together, transcripts showed other topics as well. More than one student said that being able to discuss content and answering questions about it meant that they had learned something successfully. For example, students made comments such as, “when I can explain it,” and

“telling about it,” and “I understand... when there are not that many more questions that I have.”

Two high achieving students reported additional insights about successful learning. One student said, “I think you can’t be too sure whether you have learned something like totally.... I don’t think there is a way of knowing whether or not you’ve learned something... as long as you continue to be able to explain it... that’s probably good enough.” Another student in a different interview stated, “to me, if you know something really well you can define it... then you can express it without difficulty and you can take advantage of it.” However, these types of perspectives were not common among interview participants in general.

Coding of transcripts indicated that most participants associated successful learning with their ability to remember something fluently. Likewise, students associated learning with their ability to discuss and answer questions about a topic. However, two students reported deeper insights about what it meant to know something.

In summary, results from interviews showed four categories. These categories were overall view, constructive feedback, self-expression, and fluency. Member checking procedures conducted after coding of interview transcripts showed that participants agreed more with the investigator’s understanding of overall view and constructive feedback. Interview participants reported higher frequencies of agreement, between 92% and 100%, when asked about these two categories. Alternatively, students reported less agreement regarding self-expression and fluency. Member checking results showed participants in agreement with the investigator’s interpretation of these categories between 67% and 83% of the time.

Summary

Analysis of CS test data showed nonsignificant results between planned contrasts of treatment and comparison classrooms. According to gain score ANOVA and ANCOVA results using unadjusted pretest scores, treatment and comparison classrooms significantly outperformed control classrooms. In addition, different statistical tests, such as gain score ANOVA and pretest and posttest ANCOVA produced similar results.

Analysis of SQ factor scores for Self-Assessment and Teacher Interaction showed nonsignificant differences between planned contrasts of treatment and comparison classrooms. However, pooled results of factor scores for Teacher Interaction showed significant results at the .01 level, $F(2, 253) = 6.23, p < .01$. In this analysis, the control group showed lower factor scores for Teacher Interaction in comparison to treatment and comparison groups. Analysis of correlations between Teacher Interaction factor scores and CS posttest scores indicated a significant correlation at the .008 level, $r(254) = .29, p < .001$. Generally, students who performed well on the CS posttest also indicated that they knew how they were doing on their assignments and that they received feedback from their teacher. Alternatively, analysis indicated a nonsignificant correlation between Self-Assessment and CS posttest performance at a significance level of .008, $r(254) = -.15, p = .02$.

Furthermore, coding of interview transcripts showed four categories, including overall view, constructive feedback, self-expression, and fluency. Generally, interview participants suggested they knew what to learn during a lesson from multiple sources, such as the agenda, activities, and teacher instruction. However, students did not emphasize the learning objective as a monitoring cue for their learning. In addition,

students suggested that the most useful feedback was constructive and helped them know whether what they were thinking was correct or incorrect. Moreover, transcripts showed that students favored making reflections when the process involved elements of self-expression, creativity, and multiple forms of communication such as writing, illustrating, and discussing. Last, interview participants suggested that their memory of a topic was an indicator of successful learning. However, two of the interview participants gave sophisticated explanations of successful learning, which went beyond activities associated with memory.

Chapter Five: Discussion

Chapter Overview

This chapter presents the conclusions of the study, organized in three sections. Each section addresses one of the study's research questions. The subject of the first section is whether instructional practices characteristic of reflective thinking have a positive effect on student achievement. This section also compares the three statistical tests applied to content-specific (CS) test data. The subject of the second section is whether students perceive classroom instruction differently according to a student questionnaire (SQ) when they engage in learning practices characteristic of reflective thinking. This section focuses on results of the SQ including discussion of factor scores for Self-Assessment and Teacher Interaction. The subject of the third section is whether reflective thinking influences low, medium, and high achieving seventh grade students differently according to student interviews. This section organizes the discussion according to each of the categories identified from transcript data. In addition, this section triangulates results of the CS and SQ since one of the purposes of conducting student interviews was to explain the quantitative phase of this study. Contents of the final section include overall conclusions, limitations, potential improvements, and additional research questions.

First Research Question

The first research question for this study was whether seventh grade students demonstrated increased achievement on a content-specific test as they engaged in instructional practices characteristic of reflective thinking. According to CS test results,

students in treatment and comparison classrooms performed significantly better compared to students in control classrooms. The investigator expected this outcome since students in control classrooms experienced an alternate unit of study in contrast to the other two groups.

Previous research has shown that instructional practices characteristic of reflective thinking have a positive effect on student achievement (Dignath & Büttner, 2008; Gustafson & Bennett, 2002; Ruiz-Primo et al., 2004; Song et al., 2006; Spalding & Wilson, 2002; Werderich, 2006). However, in this study, students in treatment classrooms performed equally well on the CS test in contrast to students in comparison classrooms. Nevertheless, with the exception of a meta-analysis by Dignath and Büttner, the studies cited above did not involve middle school students in the context of social studies learning.

Perhaps one reason for the discrepancy in results between previous studies and this study has to do with the brevity of the intervention period. Students in treatment classrooms received the intervention for approximately 10 minutes, for 23 days. Brown (1997) suggested that improving students' capacity for reflective thinking may require long-term changes to instructional practices as well as classroom environments overall.

Alternatively, perhaps the lessons taught by teachers during the study acted as a confounding variable for treatment and comparison groups alike. For example, the investigator designed each lesson to correspond to one or two CS test items to ensure test validity. Taking this step likely improved the alignment between curriculum and assessment for all students, regardless of their designation in treatment or comparison classrooms. This would explain why students in each classroom were able to spend 10

minutes each day engaged either in reflective thinking or in additional practice and still demonstrate similar achievement. Moreover, students in control classrooms showed similar results on the 12-week retention test after experiencing the same lessons used during the 23-day intervention period. However, students in control classrooms did not engage in the intervention or additional practice. This suggests that the lessons used in this study produced a considerable effect on students' CS test performance. This interpretation is not without precedent. For example, Raphael et al. (2008) found that effective and engaging middle school teachers worked from a well-planned lesson.

In addition, in their meta-analyses of metacognitive interventions, Dignath and Büttner (2008) and Hattie et al. (1996) found larger effect sizes in studies conducted by the principal investigator in comparison to those conducted by regular classroom teachers. Likewise, the treatment classroom taught by the investigator in this study showed higher gain scores and posttest scores in comparison to treatment classrooms taught by another teacher; see Table 6. This outcome is likely the result of the investigator being the author of the CS test as well as the principal designer of each lesson used during the 23-day study.

Another topic related to the first research question involves the statistical tests used by the investigator to contrast pretest and posttest results. As mentioned, researchers have proposed different methods for contrasting nonequivalent groups using pretest-posttest data (Dimitrov & Rumrill, 2003; Gall et al., 2003; Sheskin, 2007; Trochim, 2006). These methods include gain score ANOVA (Dimitrov & Rumrill), ANCOVA using pretest scores as the covariate (Sheskin), and ANCOVA using adjusted pretest scores as the covariate (Trochim). Results from these three statistical methods produced

similar results. For example, these three tests showed similar outcomes in 44 out of 45 post hoc contrasts. Likewise, gain score ANOVA and ANCOVA using unadjusted CS pretest scores as the covariate yielded omnibus F statistics of 15.86 and 15.81, respectively.

In summary, instructional practices characteristic of reflective thinking did not have a positive or negative effect on student achievement according to CS test results. Moreover, engaging students in reflective thinking activities or additional practice activities appeared to produce the same effect on students' CS test scores. Instead, according to retention test results, most of the achievement students demonstrated seems to be the result of the lessons used in this study. Last, despite different recommendations for comparing nonequivalent groups using pretest-posttest data, analyses in this study showed similar outcomes for each type of statistical test. This was especially the case when examining results for post hoc contrasts of statistical significance and gain score ANOVA and ANCOVA using unadjusted pretest scores as the covariate.

Second Research Question

The second research question for this study was whether seventh grade students perceived classroom instruction differently according to a student questionnaire as they engaged in learning practices characteristic of reflective thinking.

According to SQ results, analysis of Self-Assessment and Teacher Interaction factor scores between planned contrasts for treatment and comparison classrooms showed nonsignificant differences. This suggests that students in treatment classrooms and comparison classrooms perceived no differences with regard to these two constructs.

This was the case even though students in treatment classrooms wrote the learning objective as a reflective prompt for nearly each day of the 23-day study. Likewise, teachers in treatment classrooms instructed students to compare their reflection to the learning objective for each administration of the intervention. Moreover, students in treatment classrooms received feedback from teachers in their reflection journals each day during the 23-day study period. Students in comparison classrooms did not write lesson objectives or receive daily feedback about their learning. However, results showed that students in each group perceived no difference in instruction regarding lesson objectives or feedback from the teacher.

Results of factor score analysis for Self-Assessment and Teacher Interaction were similar even after pooling results of the SQ and comparing groups instead of classrooms. For instance, contrasting treatment and comparison groups produced nonsignificant results for Teacher Interaction using Tukey *a* post hoc contrasts at a significance level of .01, $p = .79$. Likewise, analysis showed nonsignificant differences between pooled results for treatment, comparison, and control groups on Self-Assessment factor scores. However, students in treatment and comparison groups showed significantly higher factor scores on Teacher Interaction in contrast to the control group using Tukey *a* post hoc contrasts at a significance level of .01, $p < .01$.

There were two differences between the control, treatment, and comparison groups, which could account for the lower rating by control group students on Teacher Interaction. First, the students in the control group experienced an alternate unit of study, with entirely different lessons. Second, a different teacher taught students in the control group. Perhaps the combination of these variables led students in control classrooms to

perceive the amount of teacher feedback they were receiving as less than students in the treatment and comparison classrooms.

The similarities that groups showed on Self-Assessment factor scores could be explained by the administrative expectations of teachers at the school where the study took place. All three participating teachers are expected to post a lesson objective for students to read for each lesson. Perhaps the posting of the learning objective caused students in control and comparison classrooms to perceive items relating to Self-Assessment in the same way as students in treatment classrooms.

Additional analysis of the SQ showed a significant correlation between Teacher Interaction and students' performance on the CS posttest at a significance level of .008, $r(254) = .29, p < .001$. This suggests that students who did well on the CS posttest also indicated that they knew how they were doing on their assignments and that they received feedback from their teacher. Alternatively, factor scores for Self-Assessment did not show a statistically significant correlation with CS posttest results. The relationship between Self-Assessment factor scores and CS posttest scores showed a negative correlation at a significance level of .008, $r(254) = -.15, p = .02$.

Furthermore, other research substantiates the significant correlation found in this study between Teacher Interaction and the posttest (Black & Wiliam, 1998a; Choi et al., 2005; Ruiz-Primo et al., 2004; Spalding & Wilson, 2002). For example, Ruiz-Primo et al. found that teacher feedback in students' science journals showed a positive correlation on related content measures. Nevertheless, as mentioned in previous chapters of this study, one element of the intervention involved peer feedback. However, the investigator was unable to measure students' perceptions regarding this form of interaction. The SQ only

assessed a limited number of characteristics involving interaction between teachers and students. Consequently, the results of this study neither substantiate nor refute research findings regarding the positive effects of peer-feedback (King, 1991) on student achievement.

Alternatively, according to factor scores for Self-Assessment, there was no relationship between students' attention to the learning objective and their performance on the CS posttest. This suggests that students in this study did not use lesson objectives as monitoring cues for their learning. Brookhart (2001) found a similar phenomenon with high school English and science students. Namely, explicitly communicating learning goals to students does not mean that they assess their own performance in comparison to those goals.

Likewise, McAlpine et al. (1999) described this idea but from the perspective of the teacher. According to McAlpine et al., teachers have a corridor of tolerance through which they accept dissonance between student achievement and intended learning outcomes. For example, students may not meet the teacher's lesson goal exactly as intended, but they may demonstrate enough learning so that the teacher moves to the next lesson, but with remediation. Perhaps students in this study followed a similar set of principles. Specifically, when confronted with an explicit learning objective, students were more or less content with the dissonance between their actual learning and the learning that the teacher intended.

Interview data from the qualitative phase of this study confirmed this interpretation. Primarily, students in this study did not perceive the learning objective as a significant monitoring cue and students seemed content with the dissonance between

their understanding of a topic and the intended learning communicated from the lesson's objective. For instance, when the investigator asked students how they knew what they were supposed to learn during a lesson, students suggested a variety of sources besides the lesson objective.

As mentioned, Dewey (1997a) suggested that learning is guided by a main topic of thought, which acts as an organizing structure and serves as the basis upon which one's ideas move toward a unified conclusion. Dewey also said that guiding one's thoughts, according to the main topic, does not consist of fixed and mechanical action. Interview data from students in this study seemed to support this theory. Students suggested that they knew what to learn from a variety of sources relating to the lesson, not necessarily from a single source.

These results bring into question the methods through which teachers challenge students' thinking in light of Vygotsky's (1978) theory of the zone of proximal development (ZPD). An important characteristic of the ZPD is that teachers assist students just enough to guide them toward new levels of mastery. However, in this study, the nonsignificant correlation between Self-Assessment and CS posttest results seems to suggest that students did not perceive lesson objectives as a significant part of this assistance. Perhaps, as Dewey (1997a) asserted, learning for the students in this study involved a larger perspective, a perspective chiefly guided by teacher feedback and classroom experiences.

In support of this conclusion, Brookhart (2001) found that successful students do not compare what they know to a given standard in a concrete manner. Rather, students evaluate what they know through general notions of success, such as "making

improvement” or “being good” at a particular discipline. Moreover, Brookhart said that students acquire these perceptions, in part, from teacher feedback.

Results from the SQ confirmed the findings of Brookhart (2001). In this study, students did not identify the learning objective as a monitoring cue. However, outcomes from this study showed that students who did well on the CS posttest also indicated that they knew how they were doing on their social studies assignments and that they received feedback from their social studies teacher. One could interpret this as meaning that students used teacher feedback as a monitoring cue, especially in comparison to learning objectives, which the investigator used as reflective writing prompts.

Third Research Question

The third research question for this study was whether characteristics of reflective thinking influenced low, medium, and high achieving seventh grade students differently according to student interviews. Interview transcript topic and analytical codes showed four categories, which the investigator titled overall view, constructive feedback, self-expression, and fluency. The investigator used these categories to triangulate results from the quantitative phase to improve the interpretability of results overall.

Interview participants suggested that they knew what to learn during each class day according to a variety of sources. These sources formed students’ overall view of intended learning outcomes for each lesson. Furthermore, participants stated that learning objectives were part of this view, but not an essential part. Likewise, interview participants seemed to imply that writing or reading the learning objective was a procedural matter, one that they did not associate with self-assessment.

Furthermore, the suggestion that students adopted an overall view in terms of learning benchmarks corroborates with research by Brookhart (2001), who found that students self-assess in terms of general notions of performance, not in terms of specific learning goals. A comment made by one high achieving interview participant in this study communicated this idea, “I just don’t really worry about what we’re trying to learn today, I just learn whatever is put in front of me.”

Another finding from interview transcripts relates to the idea of constructive feedback. Students indicated that helpful feedback involved constructive elements. Similar to the category overall view, the perception that feedback should be constructive appeared to be consistent for different achieving students. Moreover, students suggested that the length of written feedback from teachers was less important in comparison to the message that it conveyed. Students also stated that useful feedback helped them understand whether what they were thinking was correct or incorrect.

Previous research by Black and Wiliam (1998a, 1998b) and Fuchs and Fuchs (1985) has shown that teacher feedback, through formative assessment techniques, has a positive effect on student achievement. Results of student interviews support this assertion. Likewise, results of the SQ showed a significant correlation between Teacher Interaction and CS posttest scores at a significance level of .008, $r(254) = .29, p < .001$. These findings suggest that constructive teacher feedback and student achievement coincide. The triangulation between (a) the significant correlation of CS posttest scores and SQ factor scores for Teacher Interaction, along with (b) the category of constructive feedback derived from topic and analytical coding of interview transcripts support this claim.

However, these conclusions do not explain why the SQ showed nonsignificant results on Teacher Interaction between students in treatment and comparison classrooms. In other words, the increased amount of feedback that students received in treatment classrooms did not make a difference in student perceptions of instruction. One possible explanation for this discrepancy is that students in this study attended more to the quality of feedback than to its length or frequency. For instance, students in treatment groups received daily teacher feedback in their reflection journals. Comparison students did not receive daily feedback from teachers. Nevertheless, over the course of the 23-day study, students in each group did receive feedback. However, this feedback came in the form of progress reports, returned assignments, and parent conferences. Perhaps students perceived these instances of teacher feedback as meaningful and constructive. Possibly, this influenced students' perceptions enough to produce similar results between treatment and comparison groups on the SQ for Teacher Interaction. If this is the case, then one might argue that the frequency of feedback is less important than how students perceive its attributes, specifically, whether or not the feedback is constructive and meaningful.

Additional analysis of interview transcripts using topic and analytical coding indicated that students preferred self-expression as part of the reflective process. Likewise, participants stated that they favored integrating multiple forms of expression, such as writing, discussion, and illustration into their reflections. These findings are somewhat different in comparison to research examined in chapter two, which suggested that reflective thinking involves explicit characteristics (Brown, 1997; Grossman, 2009; Hatcher & Bringle, 1997; McAlpine et al., 1999; Spalding & Wilson, 2002). For instance, Brown found that teaching students specific learning strategies, such as classifying and

summarizing, has shown a positive effect on student learning. However, Brown also stated that students stop using these strategies without specific guidance.

Perhaps this means that reflective thinking involves tension between students' desire for self-expression and the structure and strategies teachers use to promote reflection. One approach for discussing this issue is through categories of knowledge, such as convergent and divergent (Ellis, 2001). Convergent knowledge means that students think the same thing about a topic, such as the definition of a word, the date of a historical event, or application of a particular skill. Alternatively, divergent knowledge means that students are free to form various perspectives on a topic, such as stating an opinion, choosing a method for solving a complex problem, or debating a controversial issue.

In *Theaetetus* (Plato, 2006b), Socrates alluded to elements of divergent thinking, specifically, the desire to think freely. For instance, Socrates suggested that the mind of the philosopher desires to explore various phenomena, completing a full investigation of each, according to one's interest. Transcript coding showed instances of students' preference for characteristics of freedom and divergence as part of the reflective process. For example, participants made statements such as "I kind of liked the drawings as long as you could put your own thoughts," and "we like writing in our journals because sometimes you have activities like writing a letter to your parents pretending you are a logger."

These examples suggest that there may have been a cognitive mismatch between students' reflections and the CS test used to measure achievement. This means that students were thinking about meaningful topics related to the content at hand. However,

the writing and illustrating students produced in their journals on these topics did not match to the cognitive demands of the CS test.

Moreover, perhaps reflective thinking is not as organized as Kompf and Bond (1995) suggested. Rather, thinking reflectively could also mean allowing students freedom to express their minds, according to their own preferred style. Indeed, interview participants in this study suggested that they preferred writing reflections that involved creativity and self-expression. However, questions remain about what this means with regard to raising student achievement on content-specific measures, which arguably assess convergent forms of knowledge.

Another category from student interviews involved students' perceptions regarding successful learning. When the investigator asked interview participants how they knew that they had successfully learned something, most mentioned some characteristic of fluent memory. However, students also said that discussing and answering questions about a topic were indicators of learning. Two higher achieving students reported deeper insights about what it meant to know something, such as one's ability to use knowledge in new situations.

Analysis of interview transcript coding in relation to students' sense of successful learning corroborated research by King and Kitchener (2004), who suggested that middle school students exhibit prereflective and quasi-reflective characteristics. According to King and Kitchener, students who exhibited prereflective thinking had a propensity for believing that knowledge was certain and that questions have definite answers. Moreover, students who exhibited quasi-reflective thinking were skeptical about knowledge claims and they tended to believe that individuals contribute to the construction of knowledge,

although with only moderate influence. According to the categories described by King and Kitchener, most interview participants exhibited prereflective qualities, while two students exhibited quasi-reflective qualities.

At the very least, interview participants demonstrated a range of prereflective and quasi-reflective characteristics, according to King and Kitchener's (2004) categories. This may suggest that students involved in this study across all 10 classrooms demonstrated a wide range of capacities for reflection, from prereflective to quasi-reflective.

Summary

The results of this investigation suggest a number of conclusions. First, engaging students in reflective thinking activities in comparison to additional practice activities produced the same results on a content-specific test. Moreover, in this study teachers engaged students in reflective thinking or additional practice for 20% of students' classroom instruction time over the course of 23 days, but with the same effect. In addition, during interviews, students commented that they preferred to engage in various forms of communication, including writing, illustrating, and discussion as part of the reflective process. Perhaps this means that social studies teachers can engage similar aged students in reflective thinking activities, using different forms of communication, without reducing students' understanding of lesson content.

Second, different statistical tests of pretest and posttest data showed more similarities than differences. This was the case despite alternative methods proposed by researchers for contrasting nonequivalent groups (Dimitrov & Rumrill, 2003; Gall et al., 2003; Sheskin, 2007; Trochim, 2006). Conceivably, this means that the choice between

analyzing gain scores using ANOVA and analyzing pretest-posttest scores using ANCOVA is more a matter of preference, at least for researchers conducting studies similar to this one. Another approach, the one taken by the investigator of this study, was to apply each test as a way to cross-validate results. Researchers may find this technique especially warranted given the number of special problems associated with quasi-experimental research (Gall et al., 2003).

Third, students in this study attended more to teacher feedback as a monitoring cue in comparison to learning objectives. In addition, students stated that constructive feedback was more important than feedback length or frequency. Perhaps this means that teachers should focus on constructive feedback in comparison to frequent feedback. Alternatively, students in this study did not perceive the learning objective as a monitoring cue. Rather, interview participants suggested that they gathered information about what to learn during a lesson from a variety of sources. Perhaps this means that teachers should present lesson objectives to students in some other format. For instance, it could be beneficial to rephrase learning objectives in the form of questions for use as a reflective prompt.

Alternatively, it could be beneficial to engage students with a prompt that simulates specific skills, such as those recommended by the National Council for the Social Studies (2008). For instance, National Council for the Social Studies recommended that students develop their skills of comparing, describing, generalizing, differentiating fact from opinion, and evaluating the validity of sources, among others. An example reflection following this course might present students with a set of lessons

on a particular topic for multiple days and then have them write a reflection, forming generalizations about related historical events.

Limitations

Although other research and the results of the interview phase corroborate some of the conclusions reached in this analysis, there are a number of limitations to this study. The most significant of these include (a) teacher effects, (b) experimenter effects, (c) student selection, and (d) intervention fidelity.

First, three different teachers were involved in this study. It is likely that each of these teachers exerted a unique effect on students' achievement and perceptions of instruction. However, analysis of the 12-week retention test showed similar results between treatment, comparison, and control groups after experiencing the same lessons. This outcome suggests that the effects of individual teachers were limited. Moreover, one of the teachers involved in this study was also the investigator. This means that the investigator had supplementary knowledge pertaining to the purposes and hypotheses of the study, which could have influenced results. For instance, the treatment classroom taught by the investigator showed higher gain scores and posttest scores in comparison to treatment classrooms taught by another teacher. However, if an experimenter effect was the cause for student gains in treatment 1, then the effect was not limited to this classroom but showed in comparison classrooms as well. For example, comparison classrooms, all taught by the experimenter, showed substantial gain scores and posttest means in contrast to treatment classrooms; see Table 6.

Second, the investigator did not randomly assign students to treatment and comparison classrooms; instead, this study involved contrasts of intact classrooms.

Despite taking steps to determine the equivalency of contrast pairs, such as matching classrooms according to multiple achievement measures and accounting for student absences, the underlying design of this study did not control for interaction effects between the selection of students and the intervention (Campbell & Stanley, 1963).

Last, the investigator did not observe the administration of the intervention to determine whether it was being applied according to its design. One reason for this was that the investigator was teaching comparison classrooms while the other participating teacher was administering the intervention.

Potential Improvements

Research has suggested that reflective thinking involves multiple characteristics, such as continuity, explicitness, interaction, and metacognition (Brown, 1997; Grossman, 2009; Hatcher & Bringle, 1997; McAlpine et al., 1999; Rodgers, 2002; Spalding & Wilson, 2002). One purpose of the SQ was to assess students' perceptions of these characteristics. However, pilot study results of the SQ did not produce reliable data concerning additional factors besides Self-Assessment and Teacher Interaction. Moreover, the factors that the SQ did measure were rather limited in their scope, since the investigator wrote corresponding items to investigate specific research questions. Furthermore, results could have been more informative if the SQ had assessed additional factors and included additional items. Likewise, it is likely that the results of this study could have been more valid if the investigator had located a preexistent instrument for assessing the dimensions of interest. Perhaps researchers will design these kinds of instruments in the future.

Additionally, the investigator applied a within-teacher random assignment design to the classrooms involved in this study (Slavin, 1992). This means that the investigator matched pairs of classrooms according to multiple achievement measures. Then, the investigator randomly assigned treatment and comparison designations to four pairs of classrooms. By matching classrooms and then randomly assigning designations, one teacher taught three treatment classrooms and no comparison classrooms. As a result, planned contrasts were less likely to control for teacher effects. However, results of the 12-week retention test showed that the lessons, which the investigator specifically designed for use with the CS tests, accounted for most of the gains made by students. Nevertheless, an improvement to this study could have included teachers instructing both treatment and comparison classrooms and thereby serving as their own control (Slavin, 1992).

Furthermore, Brown (1997) suggested that improving students' capacity for reflective thinking may require long-term changes to instructional practices and to the classroom environment overall. The duration of this study lasted for 23 days and accounted for 10 minutes of students' 50-minute class period. Perhaps producing changes in students' achievement and perceptions regarding classroom instruction requires an intervention period that lasts for more than 23 days and integrates characteristics of reflective thinking throughout lessons and even units of study.

Last, according to Gall et al. (2003), the statistical tests applied in this study could have only detected medium effect sizes with significance levels set at .05. In order to detect small effect sizes the investigator would have needed to add between 110 and 250 subjects. Perhaps the statistical tests applied in this study, in combination with a limited

number of subjects and chosen significance levels, were not powerful enough to detect small effect sizes, which the intervention may have produced (Gall et al., 2003).

Additional Questions

Although there were a number of limitations and steps the investigator could have taken to improve the study, the results do seem to warrant additional research. First, how do different kinds of reflective prompts affect student thinking? For instance, do different kinds of prompts promote convergent thinking or divergent thinking? Along with this, would a prompt in the form of a question, aligned with a specific lesson objective, be more effective in promoting convergent thinking among students and therefore improve achievement on a content-specific test?

Second, does increasing the amount of time that students receive an intervention related to reflective thinking change their perceptions regarding classroom instruction? For example, can different forms of instruction cause students to think differently about how they know they have learned something successfully? More practically, how often do students need teacher feedback to produce an effect in student perceptions regarding their learning progress? Likewise, can teachers improve students' sensitivity to teacher feedback and thereby increase student achievement? Finally, does writing a reflection produce different effects in comparison to other forms of communication such as discussing and illustrating?

In summary, this study sought to evaluate the effects of reflective thinking on the alignment of curriculum, instruction, and assessment. Although analyses for this study showed mostly nonsignificant outcomes, some results do contribute to solutions for the alignment problem. For instance, students who engaged in reflective thinking performed

as well as those who engaged in additional practice on a content-specific test.

Furthermore, students in this study interpreted teacher feedback as a monitoring cue but not learning objectives, which the intervention included as a reflective prompt. Last, students in this study provided a number of insights regarding useful teacher feedback and their preference for using different forms of communication while reflecting. Perhaps future studies will examine some of these topics to determine whether other forms of reflective thinking assist in aligning curriculum, instruction, and assessment within middle school social studies classrooms.

References

- Aesop. (1992). *Aesop's fables* (J. Zipes, Trans.). New York: Penguin.
- Aldridge, J., Fraser, B., & Huang, T. (1999). Investigating classroom environments in Taiwan and Australia with multiple research methods. *Journal of Educational Research, 93*(1), 48-62.
- Barell, J. (1984). Reflective thinking and education for the gifted. *Roepers Review, 6*, 194-196.
- Bayless, O. L. (1965). A pert influenced reflective thinking approach to problem-solving and discussion (Doctoral dissertation, University of Denver, 1965). *Dissertation Abstracts, 26*, 6231.
- Beckman, P. (2002). *Strategy instruction* (Report No. E638). Arlington, VA: Clearinghouse on Disabilities and Gifted Education. (ERIC Document Reproduction Service No. ED474302)
- Bell, B., & Cowie, B. (2001). The characteristics of formative assessment in science education. *Science Education, 85*, 536-553.
- Bianchi, G. (2007). Effects of metacognitive instruction on the academic achievement of students in the secondary sciences (Doctoral dissertation, Seattle Pacific University, 2007). *Dissertation Abstracts International, 68*, 2781.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education, 5*(1), 7-74.
- Black, P., & Wiliam, D. (1998b). Inside the black box: Raising standards through classroom assessment. *Phi Delta Kappan, 80*, 139-144.

- Bolinger, K., & Warren, W. (2007). Methods practiced in social studies instruction: A review of public school teachers' strategies. *International Journal of Social Education, 22*(1), 68-84.
- Bond, J. B. (2003). *The effects of reflective assessment on student achievement*. Unpublished doctoral dissertation, Seattle Pacific University, Washington.
- Brookfield, S. (1995). *Becoming a critically reflective teacher*. San Francisco: Jossey-Bass.
- Brookhart, S. (2001). Successful students' formative and summative uses of assessment information. *Assessment in Education: Principles, Policy & Practice, 8*, 153-169.
doi:10.1080/09695940120062638
- Browder, D., Spooner, F., Wakeman, S., Trela, K., & Baker, J. (2006). Aligning instruction with academic content standards: Finding the link. *Research and Practice for Persons with Severe Disabilities, 31*, 309-321.
- Brown, A. L. (1978). Knowing when, where, and how to remember: A problem of metacognition. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 77-165). Hillsdale, NJ: Lawrence Erlbaum.
- Brown, A. (1997). Transforming schools into communities of thinking and learning about serious matters. *American Psychologist, 52*, 399-413.
- Bruner, J. (1975). From communication to language: A psychological perspective. *Cognition, 3*, 255-287.
- Bruner, J. (1996a). *The culture of education*. Cambridge, MA: Harvard University Press.
- Bruner, J. (1996b). *The process of education*. Cambridge, MA: Harvard University Press.

- Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Chicago: Rand McNally.
- Choi, I., Land, S., & Turgeon, A. (2005). Scaffolding peer-questioning strategies to facilitate metacognition during online small group discussion. *Instructional Science*, 33, 483-511. doi:10.1007/s11251-005-1277-4
- Combs, D. (2004). A framework for scaffolding content area reading strategies. *Middle School Journal*, 36(2), 13-20.
- Creswell, J., & Clark, V. (2007). *Mixed methods research*. Thousand Oaks, CA: Sage.
- De La Paz, S. (1999). Self-regulated strategy instruction in regular education settings: Improving outcomes for students with and without learning disabilities. *Learning Disabilities Research and Practice*, 14, 92-106.
- Dewey, J. (1997a). *How we think*. New York: Dover.
- Dewey, J. (1997b). *Experience and education*. New York: Touchstone.
- Dewey, J. (2004). *Democracy and education*. New York: Dover.
- Dignath, C., & Büttner, G. (2008). Components of fostering self-regulated learning among students. A meta-analysis on intervention studies at primary and secondary school level. *Metacognition and Learning*, 3, 231-264. doi:10.1007/s11409-008-9029-x
- Dimitrov, D., & Rumrill, P. (2003). Pretest-posttest designs and measurement of change. *Work: Journal of Prevention, Assessment & Rehabilitation*, 20, 159-165.
- Dunlap, J. (2006). Using guided reflective journaling activities to capture students' changing perceptions. *TechTrends: Linking Research and Practice to Improve Learning*, 50(6), 20-26.

- Dynneson, T. L., & Gross, R. E. (1999). *Designing effective instruction for secondary social studies*. New York: Prentice Hall.
- Ellis, A. K. (2001). *Teaching, learning, and assessment together: The reflective classroom*. Larchmont, NY: Eye on Education.
- Ennis, R. (1987). A taxonomy of critical thinking dispositions and abilities. In J. Baron and R. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 1-10). New York: Freeman.
- Evans, L. (2009). Reflective assessment and student achievement in high school English (Doctoral dissertation, Seattle Pacific University, 2007). *Dissertation Abstracts International*, 70, 784.
- Faulkner, S., & Cook, C. (2006). Testing vs. teaching: The perceived impact of assessment demands on middle level instructional practices. *RMLE Online: Research in Middle Level Education*, 29(7), 1-13.
- Fink, A. (1998). *Conducting research literature reviews*. Thousand Oaks, CA: Sage.
- Flavell, J. H. (1979). Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry. *American Psychologist*, 34, 906-911.
- Fontana, A., & Frey, J. H. (2005). The interview: From neutral stance to political involvement. In N. K. Denzin & Y. S. Lincoln (Eds.), *The sage handbook of qualitative research* (3rd ed., pp. 695-727). Thousand Oaks, CA: Sage.
- Fuchs, L., & Fuchs, D. (1985, March). *A quantitative synthesis of effects of formative evaluation on achievement*. Paper presented at the annual meeting of the American Educational Research Association, Chicago.

- Fuchs, L., & Fuchs, D. (1986). Effects of systematic formative evaluation: A meta-analysis. *Exceptional Children*, 53, 199-208.
- Gall, M. D., Gall, J. P., & Borg, W. R. (2003). *Educational research* (7th ed.). New York: Pearson.
- Gallagher, C., & Worth, P. (2008). *Formative assessment policies, programs, and practices in the southwest region* (Report No. REL 2008-No. 041). Regional Educational Laboratory Southwest. Retrieved from <http://ies.ed.gov/ncee/edlabs>
- Goodlad, J. I. (1984). *A place called school*. New York: McGraw-Hill.
- Graham, S., & Perin, D. (2007). A meta-analysis of writing instruction for adolescent students. *Journal of Educational Psychology*, 99, 445-476. doi:10.1037/0022-0663.99.3.445
- Gredler, M. E. (2005). *Learning and instruction* (5th ed.). New York: Pearson.
- Green, S. B., & Salkind, N. J. (2005). *Using SPSS for Windows and Macintosh: Analyzing and understanding data* (4th ed.). Upper Saddle River, NJ: Prentice Hall.
- Griffith, B., & Frieden, G. (2000). Facilitating reflective thinking in counselor education. *Counselor Education and Supervision*, 40, 82-93.
- Grimm, L. G., & Yarnold, P. R. (1995). *Reading and understanding multivariate statistics*. Washington, DC: American Psychological Association.
- Grossman, R. (2009). Structures for facilitating student reflection. *College Teaching*, 57(1), 15-22.
- Gustafson, K., & Bennett, J. (2002). *Promoting learner reflection: Issues and difficulties emerging from a three-year study*. Mesa, AZ: Air Force Research Laboratory.

- Harniss, M. (2006). Instructionally useful assessment in the social studies. *Assessment for Effective Intervention, 31*(2), 23-37. doi:10.1177/073724770603100203
- Harris, K., & Pressley, M. (1991). The nature of cognitive strategy instruction: Interactive strategy construction. *Exceptional Children, 57*, 392-404.
- Hatcher, J., & Bringle, R. (1997). Reflection: Bridging the gap between service and learning. *College Teaching, 45*, 153-158.
- Hattie, J. A., Biggs, J., & Purdie, N. (1996). Effects of learning skills interventions on student learning: A meta-analysis. *Review of Educational Research, 66*, 99–136.
- Hubbs, D., & Brand, C. (2005). The paper mirror: Understanding reflective journaling. *Journal of Experiential Education, 28*(1), 60-71.
- Jackson, P. (1990). *Life in classrooms*. New York: Teachers College Press.
- Jenness, D. (1990). *Making sense of social studies*. New York: Macmillan.
- Jimarez, T. (2005). Does alignment of constructivist teaching, curriculum, and assessment strategies promote meaningful learning? (Doctoral dissertation, New Mexico State University, 2005), *Dissertation Abstracts International, 67*, 888.
- Joseph, N. (2006). Strategies for success: Teaching metacognitive skills to adolescent learners. *The New England Reading Association Journal, 42*(1), 33-39.
- Joyce, B., & Weil, M., (with Calhoun, E.). (2004). *Models of teaching* (7th ed.). New York: Pearson.
- Kaplan, D., Rupley, W., Sparks, J., & Holcomb, A. (2007). Comparing traditional journal writing with journal writing shared over e-mail list serves as tools for facilitating reflective thinking: A study of preservice teachers. *Journal of Literacy Research, 39*, 357-387. doi:10.1080/10862960701613136

- Kegan, R. (1994). *In over our heads: The mental demands of modern life*. Cambridge, MA: Harvard University Press.
- King, A. (1991). Effects of training in strategic questioning on children's problem-solving performance. *Journal of Educational Psychology*, 83, 307-317.
- King, A. (2002). Structuring peer interaction to promote high-level cognitive processing. *Theory into Practice*, 41(1), 33-39.
- King, P., & Kitchener, K. (2004). Reflective judgment: Theory and research on the development of epistemic assumptions through adulthood. *Educational Psychologist*, 39(1), 5-18.
- Kompf, M., & Bond, W. (1995, April). *Through the looking glass: Some criticisms of reflection*. Paper presented at the American Educational Research Association, San Francisco.
- Leming, J., Ellington, L., & Schug, M. (2006). The state of social studies: A national random survey of elementary and middle school social studies teachers. *Social Education*, 70, 322-327.
- Li, X., & Lal, S. (2005). Critical reflective thinking through service-learning in multicultural teacher education. *Intercultural Education*, 16, 217-234.
doi:10.1080/14675980500211832
- Lichtman, M. (2006). *Qualitative research in education*. Thousand Oaks, CA: Sage.
- Liddell, H. G., & Scott, R. (1996). *An intermediate Greek-English lexicon*. New York: Oxford.
- Lipman, M. (1988). Critical thinking: What can it be? *Educational Leadership*, 46(1), 38-43.

- Locke, J. (1974). *An essay concerning human understanding*. New York: Signet.
- Mager, R. F. (1984). *Preparing instructional objectives* (2nd ed.). Belmont, CA: David S. Lake.
- Martin, L. (2005). Use of cognitive strategies by high school social studies students. *Action in Teacher Education*, 26(4), 63-73.
- Marzano, R. J., Pickering, D. J., & Pollock, J. E. (2001). *Classroom instruction that works*. Alexandria, VA: Association for Supervision and Curriculum Development.
- McAlpine, L., Weston, C., Beauchamp, J., Wiseman, C., & Beauchamp, C. (1999). Building a metacognitive model of reflection. *Higher Education*, 37, 105-131.
- McDuffie, K., Mastropieri, M., & Scruggs, T. (2009). Differential effects of peer tutoring in co-taught and non-co-taught classes: Results for content learning and student-teacher interactions. *Exceptional Children*, 75, 493-510.
- McGuire, M. (2007). What happened to social studies?: The disappearing curriculum. *Phi Delta Kappan*, 88, 620-624.
- McMillan, S., & Wilhelm, J. (2007). Students' stories: Adolescents constructing multiple literacies through nature journaling. *Journal of Adolescent & Adult Literacy*, 50, 370-377. doi:10.1598/JAAL.50.5.4
- McPeck, J. (1981). *Promoting active learning: Strategies for the college classroom*. San Francisco: Jossey-Bass.
- Meyers, E. (2006). Using electronic journals to facilitate reflective thinking regarding instructional practices during early field experiences. *Education*, 126, 756-762.

- Mezirow, J. (1997). Transformative learning: Theory to practice. *New Directions for Adult and Continuing Education*, 74, 5-12. doi:10.1002/ace.7401
- Mish, F. C. (Ed.). (1987). *Webster's ninth new collegiate dictionary* (9th ed.). Springfield, MA: Merriam Webster.
- Nash, G. (2000). Expert opinion. In J. Cohen & J. Rogers (Eds.), *Will standards save public education?* (pp. 44-49). Boston: Beacon.
- National Board for Professional Teaching Standards. (2002). *What teachers should know and be able to do*. Retrieved from <http://www.nbpts.org>
- National Council for the Social Studies. (2008). *Curriculum standards for social studies*. Retrieved from <http://www.socialstudies.org>
- National Foundation for the Improvement of Education. (1996). *Teachers take charge of their learning*. Washington, DC: Author.
- Nelson, M., & Schunn, C. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science: An International Journal of the Learning Sciences*, 37, 375-401. doi:10.1007/s11251-008-9053-x
- Netemeyer, R. G, Bearden, W. O., Sharma, S. (2003). *Scaling procedures*. Thousand Oaks, CA: Sage.
- Office of Superintendent of Public Instruction. (2008). *Washington state report card*. Retrieved from <http://reportcard.ospi.k12.wa.us>
- Onwuegbuzie, A., & Collins, K. (2007). A typology of mixed methods sampling designs in social science research. *Qualitative Report*, 12, 281-316. Retrieved from <http://www.nova.edu/ssss/QR/QR12-2/onwuegbuzie2.pdf>

- Papanastasiou, C. (2000). Effects of attitudes and beliefs on mathematics achievement. *Studies in Educational Evaluation*, 26(1), 27-42. doi:10.1016/S0191-491X(00)00004-3
- Parke, C., & Lane, S. (2008). Examining alignment between state performance assessment and mathematics classroom activities. *Journal of Educational Research*, 101, 132-147.
- Patton, M. Q. (1990). *Qualitative evaluation and research methods* (2nd ed.). Newbury Park, CA: Sage.
- Paul, R. (1982). Teaching critical thinking in the strong sense: A focus on self-deception, world views and a dialectical mode of analysis. *Informal Logic Newsletter*, 4(2), 2-7.
- Pavlovich, K. (2007). The development of reflective practice through student journals. *Higher Education Research and Development*, 26, 281-295. doi:10.1080/07294360701494302
- Pederson, P. (2007). What is measured is treasured: The impact of the No Child Left Behind Act on nonassessed subjects. *Clearing House: A Journal of Educational Strategies, Issues and Ideas*, 80, 287-291.
- Pellegrino, J. (2006, November). *Rethinking and redesigning curriculum, instruction, and assessment: What contemporary research and theory suggests*. Paper commissioned by the National Center on Education and the Economy for the New Commission on the Skills of the American Workforce. Retrieved from <http://www.skillscommission.org>

- Pellegrino, J., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know: The science and design of educational assessment*. Washington, DC: National Research Council.
- Pett, M., Lackey, N., & Sullivan, J. (2003). *Making sense of factor analysis*. Thousand Oaks, CA: Sage.
- Plato. (2006a). *Meno* (W. R. M. Lamb, Trans.). Cambridge, MA: Harvard University Press.
- Plato. (2006b). *Theaetetus* (H. N. Fowler, Trans.). Cambridge, MA: Harvard University Press.
- Porter, A., & Smithson, J. (2001). *Defining, developing, and using curriculum indicators* (Report No. CPRE-RR-Ser-048). Philadelphia: Consortium for Policy Research in Education. Retrieved from <http://www.cpre.org/Publications/rr48.pdf>
- Porter, A., Kirst, M., Osthoff, E., Smithson, J., & Schneider, S. (1993). *Reform up close: An analysis of high school mathematics and science classrooms*. Madison: Wisconsin Center for Education Research.
- Pressley, M., & Harris, K. (1990). What we really know about strategy instruction. *Educational Leadership*, 48(1), 31-34.
- Puntambekar, S., & Hübscher, R. (2005). Tools for scaffolding students in a complex learning environment: What have we gained and what have we missed? *Educational Psychologist*, 40(1), 1-12.
- Randolph, J. J. (2009). A guide to writing the dissertation literature review. *Practical Assessment, Research & Evaluation*, 14. Retrieved from <http://pareonline.net>

- Raphael, L. M., Pressley, M., & Mohan, L. (2008). Engaging instruction in middle school classrooms: An observational study of nine teachers. *Elementary School Journal*, *109*(1), 61-81.
- Reid, R., & Lienemann, T. (2006). *Strategy instruction for students with learning disabilities*. New York: Guilford.
- Resnick, L., Rothman, R., Slattery, J., & Vranek, J. (2004). Benchmarking and alignment of standards and testing. *Educational Assessment*, *9*(1-2), 1-27.
- Rich, P., & Hannafin, M. (2009). Video annotation tools: Technologies to scaffold, structure, and transform teacher reflection. *Journal of Teacher Education*, *60*(1), 52-67. doi:10.1177/0022487108328486
- Richards, L. (2005). *Handling qualitative data*. Thousand Oaks, CA: Sage.
- Roach, A., Niebling, B., & Kurz, A. (2008). Evaluating the alignment among curriculum, instruction, and assessments: Implications and applications for research and practice. *Psychology in the Schools*, *45*, 158-176. doi:10.1002/pits.20282
- Rodgers, C. (2002). Defining reflection: Another look at John Dewey and reflective thinking. *Teachers College Record*, *104*, 842-866.
- Rousseau, J. (2004). *Émile*. (B. Foxely, Trans.). London: Everyman.
- Ruiz-Primo, M., Li, M., Ayala, C., & Shavelson, R. (2004). Evaluating students' science notebooks as an assessment tool. *International Journal of Science Education*, *26*, 1477-1506. doi:10.1080/0950069042000177299
- Schmidt, W. H., McKnight, C. S., Cogan, L. S., Jakwerth, P. M., & Houang, R. T. (1999). *Facing the consequences: Using TIMSS for a closer look at U. S. mathematics and science education*. Boston: Kluwer.

- Schön, D. (1983). *The reflective practitioner*. San Francisco: Jossey-Bass.
- Schön, D. (1987). *Educating the reflective practitioner*. San Francisco: Jossey-Bass.
- Schunk, D. (2008). Metacognition, self-regulation, and self-regulated learning: Research recommendations. *Educational Psychology Review*, 20, 463-467.
doi:10.1080/10573560390143094
- Searles, H. (1980). A methodological approach to the initiation of reflective thinking in the secondary social studies classroom (Doctoral dissertation, Ohio State University, 1980). *Dissertation Abstracts International*, 41, 1527.
- Seifert, K. A. (2008). The impact of a metacognitive reflection component in a problem-based learning unit (Doctoral dissertation, University of Texas A & M., 2008). *Dissertation Abstracts International*, 69, 3847.
- Sheskin, D. J. (2007). *Handbook of parametric and nonparametric statistical procedures*. New York: Chapman & Hall.
- Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1-22.
- Slavin, R. E. (1992). *Research methods in education: A practical guide* (2nd ed.). Boston: Allyn & Bacon.
- Song, H., Grabowski, B., Koszalka, T., & Harkness, W. (2006). Patterns of instructional-design factors prompting reflective thinking in middle-school and college level problem-based learning environments. *Instructional Science: An International Journal of Learning and Cognition*, 34(1), 63-87. doi:10.1007/s11251-005-6922-

- Spalding, E., & Wilson, A. (2002). Demystifying reflection: A study of pedagogical strategies that encourage reflective journal writing. *Teachers College Record, 104*, 1393-1421.
- Tapola, A., & Niemivirta, M. (2008). The role of achievement goal orientations in students' perceptions of and preferences for classroom environment. *British Journal of Educational Psychology, 78*, 291-312.
doi:10.1348/000709907X205272
- Tashakkori, A., & Teddlie, C. (1998). *Mixed methodology*. Thousand Oaks, CA: Sage.
- Tashakkori, A., & Teddlie, C. (Eds.). (2003). *Handbook of mixed methods in social and behavioral research*. Thousand Oaks, CA: Sage.
- Tindal, G., & Nolet, V. (1996). Serving students in middle school content classes: A heuristic study of critical variables linking instruction and assessment. *Journal of Special Education, 29*, 414-432.
- Trickett, E. J., & Moos, R. H. (2002). *Classroom environment scale manual* (3rd ed.). Menlo Park, CA: Mind Garden.
- Trochim, W. (2006). *Research methods knowledge base*. Retrieved from <http://www.socialresearchmethods.net/kb/conmap.htm>
- Trudeau, K., & Harle, A. (2006). Using reflection to increase children's learning in kindergarten. *Young Children, 61*, 101-104.
- Tzu, L. (1989). *Tao teh ching* (J. Wu Trans.). Boston: Shambhala.
- Valli, L. (1997). Listening to other voices: A description of teacher reflection in the United States. *Peabody Journal of Education, 72*(1), 67-88.

- Vogler, K., & Virtue, D. (2007). Just the facts, ma'am: Teaching social studies in the era of standards and high-stakes testing. *Social Studies, 98*(2), 54-58.
- Vygotsky, L. S. (1978). *Mind in society*. Cambridge, MA: Harvard University Press.
- Wang, J., & Lin, S. (2008). Examining reflective thinking: A study of changes in methods students' conceptions and understandings of inquiry teaching. *International Journal of Science and Mathematics Education, 6*, 459-479.
- Werderich, D. (2006). The teacher's response process in dialogue journals. *Reading Horizons, 47*(1), 47-74.
- Wiggins, G. P. (1993). *Assessing student performance*. San Francisco: Jossey-Bass.
- Winne, P. (2005). A perspective on state-of-the-art research on self-regulated learning. *Instructional Science, 33*, 559-565. doi:10.1007/s11251-005-1280-9
- Zimmerman, B. (2002). Becoming a self-regulated learner: An overview. *Theory into Practice, 41*(2), 64-72.

Appendix A

Informational Letter to Parents or Guardians Regarding Study Participation

September 14, 2009

Dear Parents/Guardians of Poulsbo Middle School Seventh Grade Social Studies Students

At Poulsbo Middle School, we are striving to improve an already excellent social studies curriculum. From time to time, in order to ensure that we are doing all that we can to help our students learn, teachers use new instructional strategies designed to further increase student achievement. This year, we are working with one of these new strategies.

The strategy focuses on student reflection and its connection to student achievement. Student reflection is a type of “formative assessment,” which is an instructional method that actively engages the student in measuring their own academic progress in consultation with their teacher as they move through a unit of study. The use of formative assessment aligns with district goals and “best practices” as defined by the Office of the Superintendent of Public Instruction. In addition, numerous studies in Washington State point to formative assessment as being an integral element of student preparation for the WASL.

The implementation of the new instructional strategy will coincide with students’ study of Washington State history. All students will receive all elements of the district curriculum as it relates to our study of Washington State. In fact, we are working diligently this year on collaboration in order to bring out the best in every class by pooling our efforts.

This experience will familiarize students with instructional methods that they will likely encounter later in their middle and high school experience. Furthermore, Mrs. M, Mr. Denton, and Ms. F will be able to provide school and district leadership with valuable insights regarding how the new strategies compare to other learning techniques. This valuable information will help our school to refine its social studies curriculum and instruction to ensure that all students meet their full academic potential.

We are always interested in sharing our activities with parents/guardians so if you have questions or would like additional information contact Mr. Denton at ddenton@schools.org or 360.598.1006.

Cordially, [Participating Teachers]

Informed Consent Permission Form for Participation in Student Interviews

November 9, 2009

Dear Parents or Guardians of [student name],

In cooperation with Seattle Pacific University (SPU), Mr. David Denton is conducting a study of the effects of reflective thinking practices on seventh grade social studies students. The purpose of the study is to find out how reflective thinking influences different kinds of students. Your student has been selected to participate in a 10 to 15 minute interview with Mr. Denton at Poulsbo Middle School on [date]. Your student was selected because of his/her insightful perspective regarding social studies class; about ten others students are being interviewed as well. The interview will take place in the library at Poulsbo Middle School. During the interview, students will be asked four or five questions about how they learn best, for example: "How do you know what the learning target is for each lesson?"

In order to participate, please sign the form on the following page and have your student return the form to [teacher name].

The answers that students give will be recorded on a recording device so that their responses can be considered carefully. Furthermore, the information that students give will be used in a study being conducted by Mr. Denton on the effects of reflective thinking on seventh grade students social studies learning. Student names will not be included in the study; the only information that will be kept is that which is given by students during the interview and according to the interview questions. This information will be stored on Mr. Denton's personal computer, to be used for completing the study. In addition, professors and other graduate students from SPU will read the results of these interviews. It is also possible that information from these interviews will be published in an academic journal for educational purposes; however, no student names will be used in this work.

Students will participate in interviews when it is convenient in their school day. For instance, students may be interviewed during homeroom time or at another arranged time during the school day. Every effort will be taken to ensure that students do not miss significant class time in order to participate in an interview. Plus, interviews will only be about 10 to 15 minutes long. Moreover, the benefits of participating are that students may come to understand how they learn better than they do now. And, the results of the study will yield information that assists us in understanding how grade 7 students learn social studies best. Participating in the interviews does not come with any compensation.

Participation in the interviews is voluntary; students are not required to participate. During the interviews, students may decide to stop participating at any time. Moreover, volunteers may choose to skip any question they do not want to answer. For additional information about the rights of student participants, please contact SPU at IRB@SPU.edu and reference study number 091006003 along with the expiration date 8/15/2010.

Once again, participation is voluntary and in order to participate please sign the form on the following page and have your student return the form to [teacher name].

Please contact Mr. David Denton with additional questions
360.598.1006
ddenton@schools.org

[The contents of page two follows]

Informed Consent Permission Form, Please Return

Parent/Guardian Permission to participate in study interviews on the effects of reflective thinking on student achievement (Seattle Pacific University, IRB number 091006003, exp. 8/15/2010). Please retain the copy of the attached letter for your records.

Return this signature page to [teacher name].

Parent/guardian signature

Date

Please return this form to [teacher name]

Contact Mr. Denton with additional questions.
360.598.1006
ddenton@schools.org

Content-Specific Pretest Administration Instructions

Teachers, please read to students:

“Today you will take the Social Studies preassessment. The purpose of this assessment is to find out what you know about the units that we will be studying. Another purpose is for us to learn about the way that you learn best. The information that you will be giving is very important.

Please read and answer each question carefully. Be sure to choose the best answer. If you do not know the answer, make an educated guess. Credit for this assessment is assigned according to whether or not you demonstrate positive best effort in answering each question. In addition, the information gathered in this assessment will help us adjust our instruction to improve your learning.”

Notes for teachers only:

Please alternate form A or B for students seated next to each other.

Please write down the names of any students that finish in an abnormal amount of time, such as 5 minutes.

Content-Specific Posttest Administration Instructions

Teachers, please read to students:

“Today you will take the unit test for Trees, Technology, and the Environment.

Please read each question carefully and choose the best answer. If you are not sure of an answer, make a well-educated guess.”

Notes for teachers only:

Please alternate form A or B for students seated next to each other.

3. LAST Name*

4. FIRST Name*

5. All of these are renewable*

- Animals, air, water
- Diamonds, trees, carrots
- Water, trees, oil
- Animals, fish, carrots

6. Evidence is*

- A video or picture that an event is right
- Something used in court by a lawyer
- A sign or proof of something that is true
- Something that makes the event true

7. An economic advantage of the lumber industry is that*

- Sawmills have wood
- Companies buy machines
- People earn money
- Washington exports to Japan

Figure A1. Sample of content-specific test questions in their online format.

47. There is feedback from the teacher on my social studies assignments *

Strongly Agree

Agree

Neutral

Disagree

Strongly Disagree

48. I know how I am doing on my work in social studies *

Strongly Agree

Agree

Neutral

Disagree

Strongly Disagree

51. I ask myself if I have met the learning target for each lesson in social studies *

Strongly Agree

Agree

Neutral

Disagree

Strongly Disagree

Figure A2. Sample of student questionnaire items in their online format.

Table A1

Descriptive Statistics for the Student Questionnaire

	<i>M</i>	<i>SD</i>	<i>n</i>
There is feedback from the teacher on my social studies assignments	3.63	0.91	256
I know how I am doing on my work in social studies	3.88	0.86	256
The teacher comments on my work in social studies	3.61	0.95	256
In social studies, I compare what I have learned to the learning target for that day	3.18	1.05	256
I know when I have met the learning target for each lesson in social studies	3.45	0.92	256
I ask myself if I have met the learning target for each lesson in social studies	2.99	0.99	256

Table A2

Correlation Matrix for the Student Questionnaire

Item	1	2	3	4	5	6
1	1.00	0.31	0.55	0.31	0.31	0.29
2	0.31	1.00	0.38	0.32	0.31	0.31
3	0.55	0.38	1.00	0.38	0.40	0.35
4	0.31	0.32	0.38	1.00	0.60	0.60
5	0.31	0.31	0.40	0.60	1.00	0.64
6	0.29	0.31	0.35	0.60	0.64	1.00

Note. Items in the table are numbered 1 to 6 according to the following designation:

1. There is feedback from the teacher on my social studies assignments
2. I know how I am doing on my work in social studies
3. The teacher comments on my work in social studies
4. In social studies, I compare what I have learned to the learning target for that day
5. I know when I have met the learning target for each lesson in social studies
6. I ask myself if I have met the learning target for each lesson in social studies

Table A3

Rotated Factor Matrix for the Student Questionnaire

	Factors	
	1	2
I ask myself if I have met the learning target for each lesson in social studies	0.85	
I know when I have met the learning target for each lesson in social studies	0.83	
In social studies, I compare what I have learned to the learning target for that day	0.81	
There is feedback from the teacher on my social studies assignments		0.83
The teacher comments on my work in social studies		0.81
I know how I am doing on my work in social studies		0.60

Note. Extraction method was principal component with Varimax rotation. Empty cells represent factor correlation values less than .40. The title for factor 1 is Self-Assessment and factor 2 is Teacher Interaction.

Question: What kind of feedback was most helpful for your learning?

Meaningful Feedback

C: yah it was like that's a personal **critic** and personal critics are supposed to make you better cause they teach you in the way where it's like your not good at, and they're telling you what you should **improve** on so with that in mind that should help you

Comment [D1]: Critic, from high designation student

Comment [D2]: Help you improve, high designation

S: it's pretty much **good criticism** that you're getting, like to help you out with what you're supposed to be doing in class

Comment [D3]: Critic to help, medium designation

C: even if it's positive or negative, it should still **help you**

Comment [D4]: Help you improve, high designation

P: like on where I was talking about Boeing instead of just dropping it that day **we went on** like I'd write you a little note and then after you'd said something and **it kept going on** instead of just having you wouldn't write to your teacher and just drop it that day I think it helped learn more about what I wanted to learn about.

Comment [D5]: We went on, continuity, low designation

Comment [D6]: It kept going, low designation

Figure A3. Sample of interview transcript along with topic codes and category.

Student Interview Follow-Up Questions

Hello [student name],

This is Mr. Denton and Mrs. M. You probably remember having an interview with Mr. Denton a few weeks ago. We wanted to tell you that your participation was very helpful and that we have learned a lot from the experience.

There is one final thing we need to ask from you. We listened to the answers that everyone gave in the interviews and we want to be sure that we heard all of you correctly.

To make sure we understand all of you correctly, please read each comment below and check **mostly yes** or **mostly no**. Keep in mind that there is no right or wrong way to respond. We will take these papers when you're done.

I know what I'm supposed to learn in social studies from lots of things, like the learning target, agenda, bell work [entry question], and the activities that the teacher has me do.

mostly yes ____ mostly no ____

I know what to learn in social studies by thinking about the whole lesson, not just one part.

mostly yes ____ mostly no ____

The most useful kind of feedback is constructive because it helps me know what I did right or wrong.

mostly yes ____ mostly no ____

Good feedback helps me to know what I need to improve on or if I'm doing something right.

mostly yes ____ mostly no ____

Writing, telling, or drawing my reflections helps me show or tell about my thinking.

mostly yes ____ mostly no ____

Making a reflection means that I can show whatever I'm thinking or learned from a lesson.

mostly yes ____ mostly no ____

I know when I have learned something because it's just right there in my mind and I don't have to think too hard to answer a question about it.

mostly yes ____ mostly no ____

I can tell when I have learned something because it will come to my mind later when I need to use it.

mostly yes ____ mostly no ____

Appendix B

Formula for Adjusting Pretest Error

The formula to adjust for pretest error according to Trochim (2006) follows:

$X_{adj} = X_1 + r * (X_2 - X_1)$. In this formula (a) X_{adj} is the adjusted pretest score for the individual, (b) X_1 is the mean of pretest scores to which the individual belongs, in this case the classroom, (c) X_2 is the original pretest score for the individual, and (d) r is the reliability coefficient or Cronbach's alpha of the pretest for all of the classrooms together. According to Trochim (2006), this formula adjusts each individual's pretest score in proportion to the amount of pretest error. As such, each individual's pretest score moves closer to the pretest mean of their group or in this case their classroom.

Table B1

Additional Statistics for Gain Scores

Designation	<i>n</i>	<i>M</i>	<i>SD</i>
Comparison 1	23	9.39	5.48
Comparison 2	21	8.19	4.88
Comparison 3	28	8.07	5.69
Comparison 4	23	8.48	4.62
Control 1	20	-.90	3.71
Control 2	24	.50	4.00
Treatment 1	29	10.83	4.60
Treatment 2	28	7.86	4.38
Treatment 3	28	7.18	4.05
Treatment 4	23	7.39	3.30

Table B2

Analysis of Variance Statistics for Gain Scores

Source of Variance	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Between Classrooms	2953.95	9.00	328.22	15.86*	< .01
Within Classrooms	4903.27	237.00	20.69		
Total	7857.22	246.00			

Note. * $p < .01$.

Table B3

Post Hoc Tests for Gain Scores

Designation		Mean Difference	SE	p
Treatment 1	Comparison 1	1.44	1.27	.98
Treatment 2	Comparison 2	-0.33	1.31	1.00
Treatment 3	Comparison 3	-0.89	1.22	1.00
Treatment 4	Comparison 4	-1.09	1.34	1.00
Highest-Mean Comparison 1	Lowest-Mean Comparison 3	1.32	1.28	.99
Highest-Mean Control 2	Lowest-Mean Treatment 3	-6.68*	1.27	< .01
Highest-Mean Treatment 1	Lowest-Mean Treatment 3	3.65	1.21	.08

Note. The significance level for these contrasts was set at the .05 level using Tukey *a*. * $p < .01$.

Table B4

Additional Statistics for the Posttest

Designation	<i>n</i>	<i>M</i>	<i>SD</i>
Comparison 1	24	25.46	7.10
Comparison 2	23	24.96	5.18
Comparison 3	28	24.61	6.43
Comparison 4	23	24.43	5.30
Control 1	22	16.68	4.25
Control 2	25	18.16	5.04
Treatment 1	31	27.03	5.34
Treatment 2	29	24.69	5.43
Treatment 3	28	25.21	6.91
Treatment 4	24	23.29	6.36

Table B5

Additional Statistics for Unadjusted Pretest Scores

Designation	<i>n</i>	<i>M</i>	<i>SD</i>
Comparison 1	23	16.39	4.38
Comparison 2	21	16.43	4.76
Comparison 3	28	16.54	5.59
Comparison 4	23	15.96	4.19
Control 1	20	17.33	4.81
Control 2	24	17.79	4.33
Treatment 1	29	16.31	6.49
Treatment 2	28	16.71	5.11
Treatment 3	28	17.66	6.60
Treatment 4	23	15.65	5.17

Table B6

Test of Homogeneity-of-Slopes for Unadjusted Pretest and Posttest Scores

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Corrected Model	6013.93a	19.00	316.52	16.58*	< .01
Classroom * Pretest Score	167.94	9.00	18.66	0.98	.46
Error	4333.44	227.00	19.09		
Total	148900.00	247.00			
Corrected Total	10347.37	246.00			

Note. a. R Squared = .58 and adjusted R Squared = .55. * $p < .01$.

Table B7

Analysis of Covariance for Unadjusted Pretest and Posttest Scores

Source	SS	df	MS	F	p
Corrected Model	5845.99a	10.00	584.60	30.65*	< .01
Intercept	2486.18	1.00	2486.18	130.35*	< .01
Pretest Score	3641.71	1.00	3641.71	190.93*	< .01
Classroom	2713.07	9.00	301.45	15.81*	< .01
Error	4501.38	236.00	19.07		
Total	148900.00	247.00			
Corrected Total	10347.37	246.00			

Note. a. R Squared = .57 and adjusted R Squared = .55. * $p < .01$.

Table B8

Post Hoc Tests for Unadjusted Pretest and Posttest Scores

Designation		Mean Difference	SE	p
Treatment 1	Comparison 1	1.42	1.22	1.00
Treatment 2	Comparison 2	-0.26	1.26	1.00
Treatment 3	Comparison 3	-0.52	1.17	1.00
Treatment 4	Comparison 4	-1.16	1.29	1.00
Highest-Mean Comparison 1	Lowest-Mean Comparison 3	1.28	1.23	1.00
Highest-Mean Control 2	Lowest-Mean Treatment 4	-6.36*	1.28	< .01
Highest-Mean Treatment 1	Lowest-Mean Treatment 4	3.60	1.22	.16

Note. The significance level for these contrasts was set at the .10 level using Bonferroni. * $p < .01$.

Table B9

Additional Statistics for Adjusted Pretest Scores

Designation	<i>n</i>	<i>M</i>	<i>SD</i>
Comparison 1	23	16.39	3.11
Comparison 2	21	16.43	3.38
Comparison 3	28	16.54	3.98
Comparison 4	23	15.96	2.98
Control 1	20	17.33	3.42
Control 2	24	17.79	3.08
Treatment 1	29	16.31	4.62
Treatment 2	28	16.71	3.63
Treatment 3	28	17.65	4.69
Treatment 4	23	15.65	3.67

Table B10

Test of Homogeneity-of-Slopes for Adjusted Pretest and Posttest Scores

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Corrected Model	6013.93a	19.00	316.52	16.58*	< .01
Classroom * Adjusted Pretest Score	167.94	9.00	18.66	0.98	.46
Error	4333.44	227.00	19.09		
Total	148900.00	247.00			
Corrected Total	10347.37	246.00			

Note. a. R Squared = .58 and adjusted R Squared = .55. * $p < .01$.

Table B11

Analysis of Covariance for Unadjusted Pretest and Posttest Scores

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Corrected Model	5845.99a	10.00	584.60	30.65*	< .01
Intercept	18979.98	1.00	18979.98	995.09*	< .01
Adjusted Pretest Score	3641.71	1.00	3641.71	190.93*	< .01
Group	1859.10	9.00	206.57	10.83*	< .01
Error	4501.38	236.00	19.07		
Total	148900.00	247.00			
Corrected Total	0347.37	246.00			

Note. a. R Squared = .57 and adjusted R Squared = .55. * $p < .01$.

Table B12

Additional Statistics for the Retention Test

Designation	<i>n</i>	<i>M</i>	<i>SD</i>
Comparison 1	23	25.70	7.39
Comparison 2	22	24.86	5.78
Comparison 3	27	25.19	5.68
Comparison 4	23	22.70	5.96
Control 1	23	22.00	5.99
Control 2	23	23.87	4.38
Treatment 1	27	26.22	4.12
Treatment 2	26	23.12	5.38
Treatment 3	28	23.00	6.87
Treatment 4	19	21.84	7.53

Table B13

Analysis of Variance Statistics for the Retention Test

Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Between Classrooms	520.31	9	57.81	1.63	.11
Within Classrooms	8198.86	231	35.49		
Total	8719.17	240			

Table B14

Factor Score Coefficient Matrix

	Self- Assessment	Teacher Interaction
I ask myself if I have met the learning target for each lesson in social studies	0.45	
I know when I have met the learning target for each lesson in social studies	0.42	
In social studies, I compare what I have learned to the learning target for that day	0.41	
The teacher comments on my work in social studies		0.49
There is feedback from the teacher on my social studies assignments		0.55
I know how I am doing on my work in social studies		0.35

Note. Extraction method was principal component with Varimax rotation. Empty cells represent factor correlation values less than .20.

Table B15

Factor Score Means

Designation	<i>n</i>	Mean Factor Score	
		Self-Assessment	Teacher Interaction
Comparison 1	24	0.20	0.23
Comparison 2	23	0.11	0.09
Comparison 3	28	-0.31	-0.12
Comparison 4	23	0.01	0.05
Control 1	21	0.55	-0.35
Control 2	25	-0.53	-0.56
Treatment 1	31	0.10	0.26
Treatment 2	29	-0.05	0.19
Treatment 3	28	-0.06	0.01
Treatment 4	23	0.12	0.09
Total	255	0.00	0.00

Table B16

Analysis of Variance for Factor Scores by Individual Classrooms

Factor	Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Self-Assessment	Between Classrooms	17.75	9	1.97	2.05*	.03
	Within Classrooms	236.38	246	0.96		
	Total	254.13	255			
Teacher Interaction	Between Classrooms	15.62	9	1.74	1.79	.07
	Within Classrooms	238.40	246	0.97		
	Total	254.02	255			

Note. * $p < .05$.

Table B17

Pooled Factor Score Statistics

Factor	Source	<i>n</i>	<i>M</i>	<i>SD</i>	<i>SE</i>
Self-Assessment	Comparison	98	-0.01	0.99	0.10
	Treatment	111	0.02	0.95	0.09
	Control	46	-0.04	1.14	0.17
	Total	255	0.00	1.00	0.06
Teacher Interaction	Comparison	98	0.05	1.00	0.10
	Treatment	111	0.14	0.97	0.09
	Control	46	-0.46	0.96	0.14
	Total	255	0.00	1.00	0.06

Table B18

Analysis of Variance for Pooled Factor Scores

Factor	Source	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>p</i>
Self-Assessment	Between Groups	0.12	2	0.06	0.06	.94
	Within Groups	254.00	253	0.96		
	Total	254.13	255			
Teacher Interaction	Between Groups	12.46	2	6.23	6.53*	< .01
	Within Groups	241.56	253	0.96		
	Total	254.02	255			

Note. * $p < .01$.

Table B19

Post Hoc Tests for Factor Scores between Pooled Groups

Dependent Variable	Group		Mean Difference	SE	p
Self-Assessment	Comparison	Treatment	-0.04	0.14	.97
		Control	0.03	0.18	.99
	Treatment	Comparison	0.04	0.14	.97
		Control	0.06	0.18	.94
	Control	Comparison	-0.03	0.18	.99
		Treatment	-0.06	0.18	.94
Teacher Interaction	Comparison	Treatment	-0.09	0.14	.79
		Control	0.52*	0.17	< .01
	Treatment	Comparison	0.09	0.14	.79
		Control	0.61*	0.17	< .01
	Control	Comparison	-0.52*	0.17	< .01
		Treatment	-0.61*	0.17	< .01

Note. The significance level for these contrasts was set at the .01 level using Tukey *a*.

* $p < .01$.

Table B20

Factor Score and Content-Specific Posttest Correlations

		Content-Specific Posttest	Teacher Interaction	Self-Assessment
Content-Specific Posttest	<i>r</i>	1.00	.29*	-.15
	<i>p</i>	–	.001	.02
Teacher Interaction	<i>r</i>	.29*	1.00	0.01
	<i>p</i>	.001	–	1.00
Self-Assessment	<i>r</i>	-.15	0.01	1.00
	<i>p</i>	.02	1.00	–

Note. The significance level for these correlations was set at the 0.008 level, $n = 256$.

* $p < .008$, two-tailed.

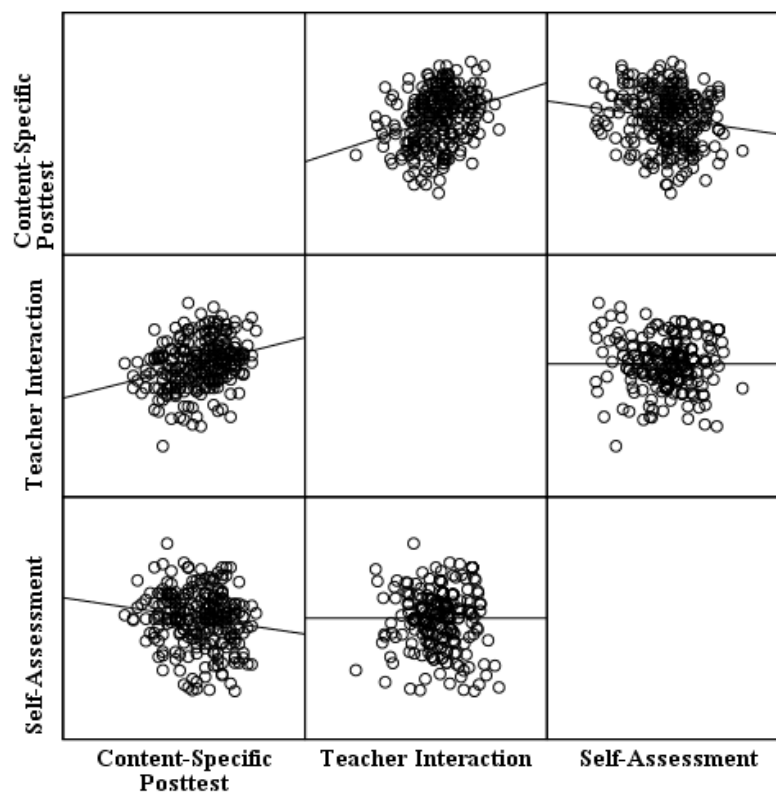


Figure B1. Factor score and content-specific posttest correlation scatter matrix.