

2017

Differentiating Literacy Instruction for Digital Learners: The Effect of Multimedia Think-Aloud Worked Examples on Adolescent Analytical Reading Comprehension

Diana Combs Neebe

University of San Francisco, dianacneebe@gmail.com

Follow this and additional works at: <https://repository.usfca.edu/diss>

 Part of the [Educational Methods Commons](#), and the [Reading and Language Commons](#)

Recommended Citation

Neebe, Diana Combs, "Differentiating Literacy Instruction for Digital Learners: The Effect of Multimedia Think-Aloud Worked Examples on Adolescent Analytical Reading Comprehension" (2017). *Doctoral Dissertations*. 335.
<https://repository.usfca.edu/diss/335>

This Dissertation is brought to you for free and open access by the Theses, Dissertations, Capstones and Projects at USF Scholarship: a digital repository @ Gleeson Library | Geschke Center. It has been accepted for inclusion in Doctoral Dissertations by an authorized administrator of USF Scholarship: a digital repository @ Gleeson Library | Geschke Center. For more information, please contact repository@usfca.edu.

The University of San Francisco

DIFFERENTIATING LITERACY INSTRUCTION FOR DIGITAL LEARNERS:
THE EFFECT OF MULTIMEDIA THINK-ALoud WORKED EXAMPLES
ON ADOLESCENT ANALYTICAL READING COMPREHENSION

A Dissertation Presented to
The Faculty of the School of Education
Department of Learning & Instruction

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Education

by
Diana Combs Neebe
San Francisco
May, 2017

THE UNIVERSITY OF SAN FRANCISCO

Dissertation Abstract

Differentiating Literacy Instruction for Digital Learners: The Effect of Multimedia Think-Aloud
Worked Examples on Adolescent Analytical Reading Comprehension

Learning by example is nothing new to the education landscape. Research into think-aloud protocols, though often used as a form of assessment rather than instruction, provided practical, content-specific literacy strategies for crafting the instructional intervention in this study. Additionally, research into worked examples—from the earliest pen-and-paper studies of algebra and statistics, to more recent multimedia studies of legal reasoning and writing—shaped the conceptual framework for the present study by detailing a series of design principles for effective multimedia worked examples. This study aimed to reimagine the face-to-face, teacher-facilitated think-aloud as a multimedia worked example, which could be leveraged for differentiated, blended instruction to support adolescent readers.

The purpose of this study was to investigate how multimedia worked examples that explicitly model the reading habits of successful readers through teacher think-alouds could facilitate effective differentiated analytical reading instruction for high school English language arts students who have access to 1:1 technology. The study tested the worked examples principle on the ill-defined problem of analytical reading comprehension in the naturalistic setting of a high school English language arts classroom using the practitioner model of the think-aloud as guidance. The study considered the effect of multimedia think-aloud worked examples on analytical reading comprehension and mental effort, as well as on the student experience of studying complex passages from literary texts.

In this experiment, an explanatory sequential mixed-methodology study, 34 sophomore English students were randomly assigned to either the worked examples treatment condition or the traditional instruction comparison condition. Using a classic treatment-comparison repeat measures pretest-posttest design, students' analytical reading comprehension and perceived mental effort was assessed. Later, in the qualitative phase of the study, the participant experience was described through interviews and annotations in order to more deeply understand the quantitative data collected.

Quantitative data were analyzed using a series of t tests between treatment and comparison groups for each phase of the study, as well as for gain scores from the baseline assessment to each of the intervention phases and to the posttest. Statistically significant differences were found between the treatment and comparison conditions for the analytical reading comprehension dependent variable at the first phase of the intervention and at the delayed posttest. No statistically significant results were found for the mental effort dependent variable. Qualitative data were coded and analyzed for emerging themes and patterns. These data revealed that students in the treatment group included higher quantity and quality annotations on their passages than did students in the comparison condition. Moreover, interviews revealed that students perceived the think-aloud process as distinct from their own analytical reading process, and they expressed that the think-aloud worked example videos increased their attention to detail, depth of analysis, ease of study, level of focus, and willingness to persist in a challenging task.

This dissertation, written under the direction of the candidate's dissertation committee and approved by the members of the committee, has been presented to and accepted by the Faculty of the School of Education in partial fulfillment of the requirements of the degree of Doctor of Education. The content and research methodologies presented in this work represent the work of the candidate alone.

Diana Combs Neebe
Candidate

May 5, 2017
Date

Dissertation Committee

Mathew Mitchell, PhD
Chairperson

May 5, 2017

Helen Maniates, PhD

May 5, 2017

Kevin Oh, PhD

May 5, 2017

Acknowledgements

A trailblazer whom I admire once famously quipped, “It takes a village.” And although writing a dissertation is a far cry from raising a child, it indeed took a village to bring this baby into existence. I’ve been blessed with the greatest village a teacher-mom-grad-student could ever dream of. To you, my village, I offer my deepest gratitude:

To my students—the 35 who willingly participated in this study, and the thousand fantastic teenagers who came before you—for inspiring me to ask better questions, collaborating with me to find answers, and sharing your experience to benefit the students who come after you.

To my Sacred Heart family, especially my English department colleagues, Dr. Jennie Whitcomb, Dr. Joy Lopez, Dr. James Everitt, Rich Dioli, Jake Moffat, Stephanie Bowe Ullman, and Barbara Intersimone, for encouraging me to pursue my passion and supporting me with your time, talent, and treasure.

To my esteemed professors, Dr. Heather Lattimer, Dr. Yvonne Bui, Dr. Nicola McClung, and Dr. Ben Baab; my dissertation committee, Dr. Xornam Apedoe, Dr. Helen Maniates, and Dr. Kevin Oh; and my chairperson, Dr. Mathew Mitchell, for challenging me to think deeply and write clearly.

To my sisters in the trenches, Melisa Kaye, Elizabeth Schroeder, and Colette Roche, for motivating me with your own fierce determination and sharp intellect.

To my dear friends, Katie Dunkelberger, Teresa Filice, Emily Liebenberg, Claire Pence, Cori and John Poffenberger, Jen Roberts, Shari Stone, and Christy Utter, for cheering me on and exhibiting tremendous patience with a long-lost friend who had too much on her plate.

To my work wife, Fehmeen Picetti, for working shoulder-to-shoulder with me as we revised rubrics, scored reading responses, conducted pilot tests, and re-arranged our classroom and calendar to make this study happen. You jumped in and ran beside me when the race was getting long and I was growing weary.

To my grandparents, Ann and Bob Bouchard, and Lee and Brownie Combs, for the sacrifices and commitments you made to ensure that my parents—and I—would have access to a fabulous education.

To my loving family, for unfathomable, unshakable support. From delicious meals and babysitting to pep talks and warm hugs, you’ve made these four years of study and writing possible: Charlie Neebe; Brian and Danielle Combs; Katherine Neebe; my troop of my aunts, uncles, and cousins; Barb and John Daines; Weeje and Alan Neebe; and most especially my parents, Cathy and Steve Combs.

Finally, I dedicate this dissertation to my husband, Steven Neebe, who believed in me from the start, and who never stopped being my greatest champion.

Table of Contents

Dissertation Abstract	ii
Signature Page	iv
Acknowledgements	v
Table of Contents	vi
List of Tables	viii
List of Figures	ix
Chapter I: Introduction	1
Statement of the Problem	1
Purpose of the Study	10
Significance of the Study	11
Theoretical Framework	12
Cognitive Load Theory	13
Worked Examples Principle	20
Worked Examples Principle in Multimedia Learning	22
Technological Pedagogical Content Knowledge (TPACK)	22
Background and Need	25
Analytical Reading Comprehension as a Complex Cognitive Process	26
Improving Analytical Reading Comprehension Through Schema	30
Learning Through Modeling	36
The Need for the Study	38
Research Questions	42
Definition of Terms	43
Chapter II: Review of the Literature	48
Think-Alouds in Literacy Instruction	48
First-Generation Worked Examples Research	55
Seminal Studies	56
Self-Explanation in Worked Examples	64
Implications for Second-Generation Research	66
Second-Generation Worked Examples Research	67
Studies of Worked Examples in Ill-Defined Domains	69
Design Principles for Multimedia Worked Examples	79
Synthesis of Worked Examples and Multimedia Principles	80
Summary of the Literature	83
Chapter III: Methodology	85
Research Questions	85
Research Design	86

Sample	88
Protection of Human Subjects	89
Treatment Description	90
Instrumentation	94
Analytical Reading Comprehension	95
Mental Effort	96
Student Experience	111
Procedures	113
Data Analyses	118
 Chapter IV: Results	 120
Research Question 1: Analytical Reading Comprehension Results	121
Research Question 2: Mental Effort Results	125
Research Question 3: Student Experience Qualitative Results	127
Passage Annotations	127
Student Interviews	137
Ancillary Results	150
Analytical Reading Comprehension Sub-Scales	151
Struggling Readers	155
Summary of Results	159
 Chapter V: Discussion of Results	 161
Summary of the Study	161
Summary of Findings	164
Limitations	166
Discussion of Findings	169
Schema Acquisition for Adolescent Literacy	169
Worked Examples with Ill-Defined Problems	171
Differentiation and the Expertise Reversal Effect	173
Literacy Learning in the Zone of Proximal Development	175
Measuring Mental Effort	178
Conclusions	179
Implications for Research	180
Implications for Practice	183
 References	 186
 Appendices	
Appendix A: Principal Letter of Support	210
Appendix B: Informed Consent Letter	213
Appendix C: Multimedia Think-Aloud Worked Example Video	216
Appendix D: Traditional Text-Based Questions Video	218
Appendix E: Passage Annotation Handouts	220
Appendix F: Analytical Reading Comprehension Rubric	228
Appendix G: Mental Effort Rating Scale for Adolescent Analytical Reading	230

List of Tables

Table 1: Qualitative Factors of Text Complexity	28
Table 2: Cognitive Processes in Reading Fluency and Comprehension	31
Table 3: Cognitive Strategies to Model in Think-Alouds	53
Table 4: Instruments	94
Table 5: Reliability Statistics, Task Irrelevant (TI) Behaviors	106
Table 6: Reliability Statistics, Task Relevant (TR) Behaviors	107
Table 7: Reliability Statistics, Element Interactivity (EI)	107
Table 8: Reliability Statistics, Overall Mental Effort and Understanding (TOT)	108
Table 9: Construct Validity for Task Relevant (TR) Behaviors	108
Table 10: Overview of Procedures by Class Period	115
Table 11: Data Analyses by Research Question	118
Table 12: Analytical Reading Comprehension Results	122
Table 13: Gain Score Results in Analytical Reading Comprehension	124
Table 14: Mental Effort Rating Results	126
Table 15: Student Annotation Coding System	129
Table 16: Student Interview In Vivo Coding System	138
Table 17: Textual Observations Sub-Scale Results	151
Table 18: Inferences Sub-Scale Results	153
Table 19: Broader Implications Sub-Scale Results	154
Table 20: Analytical Reading Comprehension Scores for Struggling Readers	156
Table 21: Analytical Reading Comprehension Gain Scores for Struggling Readers	158

List of Figures

Figure 1: Model think-aloud	6
Figure 2: Working memory and cognitive load	17
Figure 3: Technological pedagogical content knowledge	23
Figure 4: Paas (1992) Mental Effort Rating Scale	62
Figure 5: Research design schematic overview	87
Figure 6: Box plot of group demographics for HSPT	114
Figure 7: Line graph of analytical reading comprehension scores	123
Figure 8: Bar chart of quantity of surface-level annotations	131
Figure 9: Bar chart of quantity of mid-level annotations	132
Figure 10: Bar chart of quantity of deepest-level annotations	134
Figure 11: Line graph of comprehension scores for struggling readers	157

CHAPTER I

INTRODUCTION

Statement of the Problem

Independent analytical reading comprehension is a core competency required for success inside the classroom and out, and it is one of the best predictors of college readiness, not to mention long-term success in college or career (Biancarosa & Snow, 2006; Gallagher, 2009; Lattimer, 2014). Beyond basic comprehension of vocabulary and plot, *analytical* reading comprehension is the process of thinking by which students extract and construct meaning from a text (RAND, 2002), actively building an understanding of the subtextual elements of literature to include theme and the development of thematic ideas; character development, motivation, and conflict; denotative, connotative, and figurative meaning of language; structure, pacing, and organization of a story; and an author's use of language in relation to tone and a sense of time and place (Common Core State Standards Initiative, 2012).

Despite its critical importance, many secondary students struggle with this skill, and the national statistics around proficiency in analytical reading comprehension are grim. When placed in heterogeneous classes and a one-size-fits-all curriculum, these struggling readers lose access to valuable instruction and invariably fall behind (Tomlinson, 2014; Tovani, 2000). Even after years of reading intervention and literacy programs, recent research into the analytical reading proficiency of American teenagers suggests that an alarming percentage of students, by national and international measures, do not have the basic skills expected of them by educators and employers alike (Biancarosa & Snow, 2006; Casner-Lotto & Benner, 2006). Over the past two decades, analytical reading test scores have been sagging in a steady pattern of decline, with

reading scores on the ACT dropping 5 percentage points since 1999, the NAEP dropping 5 points since 1992, and the PISA slowly slipping at an annualized rate of -0.3 points per year since 2003 (ACT, 2006; NCES 2015a; OECD, 2012). In 2015, only 37% of high school seniors scored at or above the “proficient” mark in analytical reading comprehension on the National Assessment of Educational Progress exam (US Dept. of Education, 2015; NCES, 2015a). In California, where the present study took place, a mere 26% of students passed the reading comprehension subsection of the most recent California High School Exit Exam (California Department of Education, 2015). These statistics are even more alarming for low-income students and students of color, whose scores illustrate an ever-widening achievement gap in analytical reading comprehension and whose experiences point to systemically inequitable access to high-quality curriculum and instruction (Education Trust, 2013, 2015; Gallagher, 2009; NCES, 2015b). According to the international scale from the Programme for International Student Assessment (PISA), the United States ranks 17th in reading comprehension worldwide, well behind countries of nonequivalent spending on education (OECD, 2012). When it comes to adolescent reading, this is still a nation at risk (National Endowment for the Arts, 2007).

Over the past 30 years, the direction of adolescent literacy instruction has remained in alignment with the national push toward educational standards through the Standards and Accountability Movement of the early 1990s, No Child Left Behind in the early 2000s, and Race to the Top and the Common Core Initiative in the 2010s (American Institutes for Research, 2013). To date, educators and legislators have determined a rigorous list of 10 comprehensive and critical skills for adolescent readers of literature to master, highlighting how complex a task it is for a high schooler to become proficient in analytical reading (Common Core State Standards Initiative, 2012):

1. Cite strong and thorough textual evidence to support analysis of what the text says explicitly as well as inferences drawn from the text.
2. Determine a theme or central idea of a text and analyze in detail its development over the course of the text, including how it emerges and is shaped and refined by specific details; provide an objective summary of the text.
3. Analyze how complex characters (e.g., those with multiple or conflicting motivations) develop over the course of a text, interact with other characters, and advance the plot or develop the theme.
4. Determine the meaning of words and phrases as they are used in the text, including figurative and connotative meanings; analyze the cumulative impact of specific word choices on meaning and tone (e.g., how the language evokes a sense of time and place; how it sets a formal or informal tone).
5. Analyze how an author's choices concerning how to structure a text, order events within it (e.g., parallel plots), and manipulate time (e.g., pacing, flashbacks) create such effects as mystery, tension, or surprise.
6. Analyze a particular point of view or cultural experience reflected in a work of literature from outside the United States, drawing on a wide reading of world literature.
7. Analyze the representation of a subject or a key scene in two different artistic mediums, including what is emphasized or absent in each treatment.
8. Delineate and evaluate the argument and specific claims in a text, including the validity of the reasoning as well as the relevance and sufficiency of the evidence [in informational texts].

9. Analyze how an author draws on and transforms source material in a specific work (e.g., how Shakespeare treats a theme or topic from Ovid or the Bible, or how a later author draws on a play by Shakespeare).
10. Read and comprehend literature, including stories, dramas, and poems, independently and proficiently.

In response to declining test scores and increasing federal pressure to “turn around” our schools and meet the many standards set for our students, educators have implemented myriad interventions targeting struggling readers: districts have created supplementary curriculum for remedial readers and launched reading recovery programs; schools have hired literacy coaches and increased the rigor of their course offerings; and teachers have learned how to implement higher standards and have applied new instructional strategies to support an ever-changing student population (US Department of Education, 2015). One specific analytical reading instructional strategy that has gained traction, particularly in elementary school literacy circles, is the teacher-led think-aloud, which has become a standard practice in the teaching of reading (Appleman, 2010; Beers, 2003; Frey & Fisher, 2013; Wilhelm, 2001) and is described by literacy experts Keene and Zimmermann as “the single most important teaching tactic at our disposal” (Keene & Zimmermann, 2007, p. 146).

The think-aloud is an instructional strategy for supporting developing readers in which the teacher models the strategies of a successful analytical reader, such as activating prior knowledge, decoding text at multiple levels, making predictions, visualizing, summarizing, asking questions, making connections, and monitoring and clarifying understanding (Appleman, 2010; Wilhelm, 2001). In a think-aloud, the teacher typically begins by stating the purpose for reading and asking the students to listen for the teacher’s use of reading strategies. He or she then

reads a selection of the text for the students, pausing frequently to literally think out loud, thus presenting an audible representation of the internal dialogue that takes place between the mind of a skilled reader and the text (Beers, 2003; Keene & Zimmermann, 2007; Wilhelm, 2001). During the think-aloud, the teacher provides a verbal or visual cue to signal to students the transition from reading to thinking, and clarifies how a particular strategy helps the teacher make meaning in a passage (Beers, 2003; Keene & Zimmermann, 2007). The goal of the teacher think-aloud is for students to adopt the habits of successful readers by hearing those habits modeled enough times that they can internalize the cognitive processes required for effective analytical reading (Frey & Fisher, 2013; Keene & Zimmermann, 2007). Figure 1 presents a classic think-aloud script (Frey & Fisher, 2013, pp. 21–22), based on the text *Number the Stars* (Lowry, 1989, pp. 46–47), which is often taught in the fifth grade.

Text	Think-Aloud
<p>"Your names?" the officer barked. "Annemarie Johansen. And this is my sister —" "Quiet! Let her speak for herself. Your name?" He was glaring at Ellen.</p>	<p>Today we're going to think about the specific words that the author has chosen and practice asking questions while we read. I am going to share some of my thinking with you, especially about the author's words and the questions I have when I am reading.</p>
<p>Ellen swallowed. "Lise," she said, and cleared her throat. "Lise Johansen."</p>	<p>"Barked" and "glaring" are words the author is using to help me imagine this Nazi officer's anger. The author's word choice is important and in this case helps me think about the seriousness of the situation.</p>
<p>The officer stared at them grimly. "Now," Mama said in a strong voice, "you have seen that we are not hiding anything. May my children go back to bed?"</p>	<p>I know why she's lying. She doesn't want them to know she's Jewish so she pretends to be a member of the family. I'll bet the Johansens will also lie for Ellen.</p>
<p>The officer ignored her. Suddenly he grabbed a handful of Ellen's hair. Ellen winced.</p>	<p>I was right. That's just what Mrs. Johansen did.</p>
<p>He laughed scornfully. "You have a blond child sleeping in the other room. And you have this blond daughter —" he gestured toward Annemarie with his head. "Where did you get the dark-haired one?"</p>	<p>Why would he pull Ellen's hair?</p>
	<p>Oh, no! What will the Johansens say? Now I understand why the title of this chapter is "Who Is the Dark-Haired One?"</p>
<p>Think-Aloud for <i>Number the Stars</i> (Lowry, 1989, p. 46-47). Excerpt from <i>Rigorous Reading</i> (Frey & Fisher, 2013, p. 21-22).</p>	

Figure 1. Model think-aloud (Frey & Fisher, 2013, pp. 21–22)

The analytical thinking modeled in Figure 1 moves beyond surface-level comprehension, such as decoding and vocabulary recognition, into the subtext of the story, such as character motivation, internal and external conflict, and tone. Instead of focusing on the *what* of the story, the think-aloud delves into the *how* and *why* of the literature.

As with any instructional strategy, think-alouds cannot solve all problems. Two central challenges currently limiting the efficacy of think-alouds in the secondary English classroom are timing and differentiation. First, timing: As students transition from elementary to secondary school, a parallel transition occurs in the purpose and place of reading. In English language arts and other disciplines such as social studies and science, this marks the shift from “learning to read” to “reading to learn,” in which disciplinary knowledge is taught predominantly through text (Allington, 2002; Conley, 2009; Lattimer, 2010). Teachers often assign content reading as homework and make use of class time for more hands-on activities, meaning-making, and, of course, assessment (Burke, 2012; Christenbury, 2006; Milner, Milner, & Mitchell, 2011; Smagorinsky, 2001). Unfortunately, in this standard model, struggling readers often do not have the support they need at home to access a challenging text, and instead they spend their time reinforcing misunderstanding and building frustration (Gallagher, 2004; Miller, 2009). Though the think-aloud process provides a useful structure for unpacking a complex text, it is not the just-in-time support that many students need when they are stumbling through independent reading at home.

Second, the current manifestation of the think-aloud as a tool for differentiation is limited. In the midst of classes with advanced independent readers, there is invariably a group of analytical reading “novices” who have difficulty following basic fiction structures and identifying plot elements in novels and short stories (Lewis & Ferretti, 2009). Often these students give up and quit reading altogether (Schoenbach, Greenleaf, & Murphy, 2012). Their challenges pervade across environments and disciplines, as they struggle any time reading occurs without the explicit guidance of a teacher reading aloud and directing their attention (Griffin, Wiley, & Thiede, 2008; Rupley, Blair, & Nichols, 2009; Unsworth & McMillan, 2013). Because

the think-aloud is an in-class, teacher-presented strategy, it limits the teacher to two instructional options: either (a) present the lesson to all learners, regardless of their prior knowledge, reading readiness, or need; or (b) present the lesson to a small group while assigning independent work to other students, thus making it exceptionally clear which students struggle and which students excel. To make the think-aloud a more effective instructional strategy, teachers need the flexibility to do two things: (a) present it only to students who need it, and (b) do so during the time in which they are reading. The influx of technology into schools provides an opening for improving a standard classroom practice, because it addresses the limitations of timing and differentiation.

During the same years in which the national education trend moved toward more rigorous standards and measurable outcomes for all learners, the education system experienced a dramatic change with the introduction of instructional technology into schools. As classrooms have become wired and connected spaces, and the Internet and digital devices have quickly become as ubiquitous for this generation of students as landlines and printed encyclopedias once were for their predecessors, today's educators have been given a unique opportunity to leverage technology in support of meaningful learning (Casap, 2015; ISTE, 2008). A technology-rich environment—and, in particular, a one-to-one (1:1) environment in which each student has access to a laptop, tablet, or smartphone in school and at home—offers teachers a diversity of options for how they work to meet the needs of all learners, including the potential to present a lesson in a manner that is both timely and targeted. With 1:1 technology, the teacher-led think-aloud could be presented digitally as part of a student's reading homework and discretely assigned only to those who need the additional scaffolding to be successful. Readers would receive just-in-time support, and no one would be the wiser about who received extra instruction.

Access to technology gives teachers the tools for “differentiation with dignity” (Neebe & Roberts, 2015, p. 106). Moreover, when all students can make sense of the independent reading assigned to them, they all can enter class on the same page, prepared to participate and learn. This model of using video-based technology to deliver some portion of direct instruction as homework, which frees up face-to-face time for more interactive and/or individualized experiences, is called “flipped teaching” or “blended learning” (Khan, 2011; Neebe & Roberts, 2015; Sams & Bergmann, 2013; Tucker, 2012).

Technology, of course, is no panacea. Simply placing tools in the hands of educators in hopes of transforming adolescent literacy instruction would be ill-considered. Indeed, the promise of change based on cutting-edge technology is nothing new in the education landscape. Consider the introduction of the motion picture, the Internet, the computer lab, and the classroom television. Each promised to revolutionize education, and each failed to deliver any significant improvement in teaching and learning (Mayer, 2009, 2014a). Until the publication of recent multimedia theories and studies (Mayer, 2005, 2014b), classroom teachers have had to rely upon trial and error to determine the utility of their blended instruction.

Fortunately, there is a robust body of new research literature that details the design principles of effective multimedia learning, providing clear guidance to instructional designers in technology-rich environments. To date, video-based instruction has been used most commonly in science, technology, engineering, and mathematics education to teach skills and procedures, or to present worked-out examples of problems and proofs (Mayer, 2014b; Renkl, 2014). Of particular interest to this study is the multimedia worked example (Renkl, 2014; Renkl & Atkinson, 2010), a well-researched, statistically effective, multimedia-based instructional strategy that parallels the practitioner process of the analytical reading think-aloud, albeit in other disciplines. A

multimedia worked example presents an expert's step-by-step thought process to solving a complex problem for students to study and emulate (Atkinson, Derry, Renkl, & Wortham, 2000). As evidenced by the list of standards for analytical reading and the current state of adolescent literacy, analytical reading comprehension is a complex skill for students to master, and teachers are strapped for solutions. This study aimed to apply a parallel solution to the problem at hand: to reimagine the face-to-face, teacher-facilitated think-aloud as a multimedia worked example that could be leveraged for differentiated, blended instruction.

Purpose of the Study

The purpose of this study was to investigate how multimedia worked examples that explicitly model the reading habits of successful readers through teacher think-alouds could facilitate effective differentiated analytical reading instruction for high school English language arts students who have access to 1:1 technology. In this experiment—an explanatory sequential mixed-methodology study—34 sophomore English students were randomly assigned to either the worked examples treatment condition or the traditional instruction comparison condition. Using a classic treatment-comparison repeat measures pretest-posttest design, students' analytical reading comprehension and perceived mental effort was assessed. This study addressed the essential quandary that most English language arts teachers face: how to provide an appropriate level of analytical reading instruction for all learners, support novice readers without inhibiting advanced readers, and challenge advanced readers without ignoring novice readers. Currently this problem is being addressed through either individual tutoring, which is neither time effective nor scalable, or whole-class instruction, which denies stronger readers the challenge and instructional pace they need.

Significance of the Study

Previous research into teacher-led think-alouds as an instructional strategy has centered on early literacy instruction and almost exclusively featured qualitative case studies of kindergarten through fifth grade (Baumann, Jones, & Seifert-Kessell, 1993; Beers, 2003; Davey, 1983; Frey & Fisher, 2013; Olshavsky, 1977). Little research has been done using the teacher-led think-aloud method with adolescent readers, and very little quantitative research has been done in this area. Conversely, much quantitative research attention has been given to worked examples of *well*-defined problems, in areas such as mathematics and science (Atkinson et al., 2010; Renkl, 2014; Renkl & Atkinson, 2010; Sweller & Cooper, 1985), yet relatively little attention has been given to applying worked examples to *ill*-defined problems (Kyun, Kalyuga, & Sweller, 2013; Nievelstein, van Gog, van Dijck, & Boshuizen, 2013; Rourke & Sweller, 2009). There are no known studies applying multimedia worked examples either to analytical reading comprehension or to triple-content worked examples (Renkl, Hilbert, & Schworm, 2009) in which students must contend with three domains: the learning domain (e.g., analytical reading), the exemplifying domain (e.g., the text being read), and the cognitive strategy domain (e.g., heuristic strategies and schema for approaching analytical reading). This mixed-methods study of multimedia think-aloud worked examples is significant because it bridges the gaps in both bodies of literature.

This study extends previous think-aloud research by (a) taking the elementary classroom practice of the teacher-facilitated think-aloud for early literacy instruction and placing it in the high school classroom setting as a scaffold for analytical reading comprehension, and (b) applying empirical, quantitative methodology to a classroom practice that has been studied and

reported predominantly through professional development texts and practitioner journals, or as a measurement protocol rather than an instructional method (Kucan & Beck, 1997).

Moreover, this study extends previous worked examples research by (a) situating the quasi-experiment in the naturalistic setting of a high school English language arts classroom instead of the traditional controlled laboratory setting, and (b) applying the model of worked examples to the ill-defined task of analytical reading to determine if worked examples of analytical reading effectively support differentiated instruction in English language arts, a rare triple-content worked example.

Beyond theoretical importance, the present study contributes an innovative, research-based solution to an age-old practical dilemma. By digging deeply into one instructional move, this study has the potential to equip educators to become more technologically skilled at their pedagogical craft. Practitioners in the field of English education specifically, or reading instruction broadly, should be emboldened to leverage technology, access to which is becoming pervasive, to support readers of varying readiness levels within a heterogeneous class or beyond the four walls of the classroom.

Theoretical Framework

Educational theories provide a lens through which to observe a problem or phenomenon. Certainly there are many ways to consider the challenge of supporting students with analytical reading comprehension. Of the four predominant learning theories—behaviorism, socioculturalism, constructivism, and cognitivism—this study takes a cognitivist approach. The foundation for the present study is Sweller’s (1988, 2010a) cognitive load theory, a psychological theory and set of instructional design principles based on an understanding of human cognition. Nested within that theory is the worked examples principle (Sweller, 2006;

Sweller & Cooper, 1985; Renkl & Atkinson, 2010), which is the specific instructional method used in this study to reduce cognitive overload for students engaging in the complex task of analytical reading. From there, the study builds upon developments made in worked examples research by considering the worked examples principle in multimedia learning (Renkl, 2014). Finally, this study situates worked examples within the practitioner context, drawing upon Mishra and Koehler's (2006) Technological Pedagogical Content Knowledge (TPACK) model as a guide.

Cognitive Load Theory

Cognitive load theory (Sweller, 1988, 2010a) is a relatively new psychological theory that explains the way in which instructional design affects learning by considering how the brain processes new information and retains knowledge (Moreno & Park, 2010). Human cognitive architecture is composed of the working memory and the long-term memory. The brain learns by relying on an information-processing system that can store knowledge, use the knowledge it has stored, create new knowledge, change existing knowledge to respond to the outside world, and accomplish all of these functions effectively (Sweller, 2012). Cognitive load theory is based on three fundamental assumptions about human cognitive architecture that have been consistently supported by cognitive psychology research.

Working memory. The first assumption of cognitive load theory is that the working memory—the part of the brain responsible for consciously processing and assimilating new information—has a fixed capacity and can manage only a finite number of elements at once (Moreno & Park, 2010). Miller (1956) estimated this number to be “seven, plus or minus two” independent pieces of information that a person can reliably hold simultaneously (p. 81). For educators, this first assumption has significant implications: instructional design must

acknowledge and accommodate for the constraints of students' working memory. If students are asked to make sense of a lesson, instructional materials, or content that extends them beyond the bounds of their working memory, it is unlikely that learning will take place (Sweller, van Merriënboer, & Paas, 1998). Students attempting to juggle too many competing demands on their attention will be unable to process the new information and transfer it to their long-term memory.

Long-term memory. The second assumption is that the long-term memory has a vast, potentially infinite capacity for storing complex networks of information (Moreno & Park, 2010), in contrast to the working memory, which serves as a temporary holding place for information as it is being learned. Of particular importance to the instructional designer, the long-term memory houses a student's prior knowledge, the concepts and skills that the student has stored from the process of learning, which support his or her future learning by integrating new knowledge in the working memory with prior knowledge in the long-term memory.

Educational psychologist David Ausubel (1968) said of the relationship between prior knowledge and meaningful learning that "the most important single factor influencing learning is what the learner already knows" (p. iv). From the perspective of cognitive load theory, this is because prior knowledge determines a student's level of expertise within a particular discipline (Kalyuga, Chandler, & Sweller, 1998), and thus it influences the amount of working memory energy a student must expend to make sense of new information (Plass, Kalyuga, & Leutner, 2010).

Schema. The third assumption of cognitive load theory is that schemas are necessary for transferring information from the working memory to the long-term memory and thus freeing up working memory capacity (Moreno & Park, 2010; Sweller, 1994). Rather than storing discrete

pieces of information in the long-term memory, humans have an associative memory that stores information in connected webs called schemas (Anderson & Bower, 1983; Kalyuga, 2010).

Schemas make sense of information in complex sets of connected facts that enable a person to process related information as one element instead of separately as all of its individual parts.

Sweller, van Merriënboer, and Paas (1998) explained schema construction in light of the process of reading:

A schema categorizes elements into information according to the manner in which they will be used ... When reading, we can derive meaning from an infinite variety of marks on a page because we have schemas that allow us to appropriately categorize letters, words, and combinations of words ... In early school years, children construct schemas for letters that allow them to classify an infinite variety of shapes (as occurs in handwriting) into a very limited number of categories. These schemas provide the elements for higher order schemas that are combined into words that in turn can be combined into phrases, and so forth. Ultimately, this process allows readers to rapidly scan a page filled with a hugely complex array of squiggles and derive meaning from it.
(p. 255)

As with the example of reading, well-developed schemas bypass the working memory and are processed automatically. The adept reader does not need to think about the process of translating markings into letters and letters into words. Schema automation makes room in the working memory for new learning, and it is this process of automation that separates an expert from a novice.

Research into schema theory started, surprisingly, by studying chess grand masters to determine what separated them in skill and speed from amateur chess players (De Groot, 1966).

Chess grand masters did not have innately more capacious working memories; instead they had a robust store of chessboard configuration schemas stored in their long-term memories that automated their processing when approaching a new game of chess, leaving more working memory processing power available for problem solving and creative solutions (Kalyuga, 2010). The difference between expert and novice, therefore, is prior knowledge, or schema. Without any prior knowledge to help make sense of new information, the natural default for the working memory is to process that new information through a sluggish, error-prone, guess-and-check procedure in which the working memory continues to randomly generate and test one solution at a time until an optimal solution is reached (Renkl, 2014; Sweller, 2012). Clearly, using schema is a far more efficient approach to learning. The implication for educators is that instructional design should free the learner to focus on schema acquisition, instead of filtering through other competing, extraneous cognitive demands (Moreno & Park, 2010). One way to accomplish this goal in instructional design is through worked examples, which are discussed in the subsequent section. First, however, it is imperative to clarify how the working memory processes information by exploring the three types of cognitive “loads” on the working memory: extraneous load, intrinsic load, and germane load (see Figure 2).

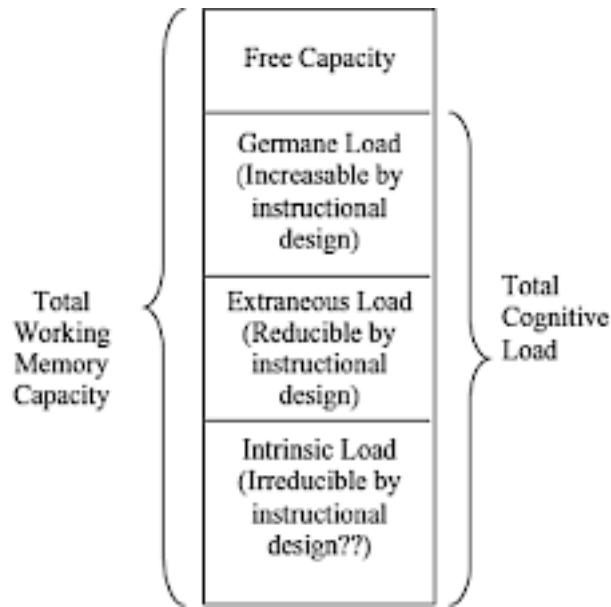


Figure 2. Working memory and cognitive load (Moreno & Park, 2010, p. 18)

Extraneous load. The first of the three loads to take shape in the theoretical model of cognitive load was extraneous load. This load inhibits working memory capacity by forcing the learner to attend to unnecessary cognitive demands in the learning process—that is, any demands that do not lead to the acquisition and automation of schema (Sweller & Chandler, 1994). Extraneous load is attributed to suboptimal instructional design (Sweller, 2010b) and can be caused by any of the following four scenarios: (a) instruction or learning materials that do not properly support a student who may lack sufficient prior knowledge to comprehend the lesson, thereby triggering random processing in the learner; (b) instruction or learning materials that ignore preexisting prior knowledge and overteach the content, leading to an overlap in the learner’s existing schema and the one presented by the teacher, and causing the student to contend with additional material to process; (c) instruction or learning materials that present too much information at once; and/or (d) instruction or learning materials that present information in a distracting or disorganized manner, requiring students to exert effort to mentally reintegrate

parts or search for components that belong together, such as graphs and labels that are separate from one another instead of integrated (Kalyuga, 2010).

Intrinsic load. Intrinsic cognitive load is the amount of load imposed by the inherent difficulty of the information being learned (Sweller, 1994). Because difficulty can be relative, intrinsic load comprises both the complexity of the content and the prior knowledge of the learner (Moreno & Park, 2010). Complexity is credited to the amount of element interactivity in a given task (Pollock, Chandler, & Sweller, 2002), which is the extent to which understanding one piece of information depends on understanding another related piece of information, or the extent to which understanding the “whole” requires understanding a number of subordinate parts. Given that the working memory can handle only a finite number of elements at once, the amount of intrinsic load placed on the working memory is determined in large part by the extent to which those elements interact, and thus, the extent to which the working memory must process multiple elements simultaneously (Pollock et al., 2002). Though some recent studies suggest that there are ways to manage intrinsic load by artificially reducing element interactivity (Mayer & Moreno, 2010; Paas, Renkl, & Sweller, 2003), most findings support the notion that intrinsic load is not something that instructional designers can control (Moreno & Park, 2010). Importantly, intrinsic load and extraneous load are additive, which means that as material becomes more complex and carries a higher intrinsic load, the learner has less capacity in the working memory for processing extraneous load (Sweller & Chandler, 1994). Though the techniques suggested by cognitive load theorists would be effective when learning low-intrinsic load material, they are crucial for comprehending high-intrinsic load material, and thus the work of cognitive load theory is focused predominantly on supporting students learning inherently complex content (Moreno & Park, 2010).

Germane load. The third source of cognitive load is germane load, which refers to the mental effort dedicated to developing and automating schema (Moreno & Park, 2010). According to cognitive load theory, this is the central goal of learning and therefore most “germane” to the process of moving knowledge from the working memory to the long-term memory. Whereas extraneous load and intrinsic load impose a penalty on the working memory, germane load is beneficial because it improves the learning process (Sweller et al., 1998). Increasing germane load is the opportunity that teachers create by reducing extraneous load. Sweller (2010a) explained that “reducing extraneous cognitive load would have little function if the working memory resources so freed were not used for productive learning” (p. 43). Thus, effective instructional design should not only reduce or eliminate extraneous cognitive load, but also direct the learner to integrate new learning with prior knowledge and engage the learner in creating or borrowing structures and organizational models for making sense of new information.

By understanding the principles of cognitive load theory, educators can design learning experiences that consider the relative stress that each activity creates and maximize the learning that happens in students’ limited information processing systems (Sweller, 2010a; Sweller et al., 1998). Research suggests that students’ working memory capacity is diminished when dealing with unfamiliar material (Kalyuga, 2006). It is imperative for educators to have a solution for supporting the learning process and the acquisition of new knowledge, which is often the preponderance of the work that happens in classroom instruction. Instructional design that respects this theory seeks to optimize students’ cognitive load by reducing extraneous load, managing intrinsic load, and increasing schema acquisition, or germane load (Moreno & Mayer, 2010). Within the domain of cognitive load theory research, one of the most consistent and

significant effects for mitigating distracting extraneous load and focusing the learner on schema acquisition is the worked examples principle (Sweller, 2006).

Worked Examples Principle

Simply stated, worked examples are instructional scaffolds that depict an expert's detailed solution to a problem for students to study and emulate (Atkinson et al., 2000). A traditional worked example presents a problem to be solved in addition to step-by-step written work that applies a strategy or principle to solve the problem in a systematic way (Sweller, 2006). Students study the worked-out example and then actively process what they have studied by engaging in a similar problem or procedure to gain feedback on what they understand (Sweller, 2006). They repeat this process several times with additional problems (Sweller, 1994). Statistically speaking, worked examples are effective. Educational researcher John Hattie (2009) reports from his meta-analysis of more than 800 studies that among the many moderating factors in student achievement, conventional worked examples have a moderate effect size of $d = .57$. Effect size is a statistical measure that describes "how big the difference is between group means, or how impressive the relationship is between variables," and thus provides a sense of the practical importance of a given phenomenon (Mitchell, 2015). Effect size is most commonly measured using Cohen's d , which—much like a z-score—uses standard deviation units to quantify the difference between two group means (Cohen, 1988). An effect size of $d > .80$ (eight tenths of a standard deviation difference) is considered a large effect; $d > .50$ (half of a standard deviation difference) is considered a moderate effect; $d > .20$ (two tenths of a standard deviation difference) is considered a small effect; and $d < .20$ is considered as having no effect.

According to the worked examples principle, worked examples are effective because they support novice students in developing schemas, which are necessary for students to solve

cognitive problems (Sweller, 1994). Worked examples free up cognitive capacity by allowing learners to focus specifically on the most relevant structural features of a problem, rather than trying to consider many different solution options or work through all of the additional steps required to solve a problem beyond the new skill set being learned (Renkl & Atkinson, 2010). Instead of relying on the brain's natural, randomized trial-and-error process, which burdens or overloads the working memory's capacity, worked examples allow the learner to temporarily "borrow" the schema of someone more expert, thus reducing extraneous load and returning working memory function to schema building (Renkl, 2014; Sweller, 2006). Specifically, worked examples support germane load by increasing natural self-explanation, which is one of the strongest predictors of student performance (Renkl, 1997).

Research into worked examples started three decades ago with mathematics instruction, during a time in which the prevailing pedagogy was to give students an abundance of problem-solving practice. Sweller and Cooper (1985; Cooper & Sweller, 1987) noticed that students in these conditions continued to use guess-and-check methods despite their repeated practice with more streamlined strategies. When Sweller and Cooper (1985; Cooper & Sweller, 1987) introduced the worked example as an alternative method for students to study, they found that students emulating worked examples were more efficient and more effective. Although the first generation of worked examples tested the principle with well-defined problems from mathematics and science in controlled, laboratory-like settings (Atkinson et al., 2000), more recent research has successfully applied the principle to ill-defined problems such as legal reasoning (Nieveelstein et al., 2013), artistic design (Rourke & Sweller, 2009), rhetoric (Schworm & Renkl, 2007), and writing (Kyun et al., 2013).

Worked Examples Principle in Multimedia Learning

Twenty years after the initial worked examples principle was penned, the cognitive theory of multimedia learning emerged and reinvigorated the principle with multimedia worked examples (Mayer, 2014a). The updated theory brought two advances to worked examples research: (a) it expanded researchers' understanding of human cognitive architecture by describing the dual-channel nature of the brain, in which people take in information through two separate channels—a verbal channel and a visual channel (Paivio, 1986; Renkl, 2014); and (b) it helped to shape the evolution of worked examples by applying them to more complex domains and ill-defined problems (Renkl, 2014).

The cognitive theory of multimedia learning suggests that students learn best from a combination of words and pictures, rather than words alone, because this combination maximizes possible instructional input by using both information processing channels (Mayer, 2014a; Paas & Sweller, 2014). Accordingly, multimedia worked examples present an expert's detailed solution to a problem through a combination of words and pictures. At this point, the majority of worked examples in complex domains are multimedia worked examples (Renkl, 2014). Multimedia worked examples are an effective high-leverage practice for scaffolding learning of ill-defined problems precisely because they more fully support schema acquisition while easing the working memory demands placed on the learner during active learning.

Technological Pedagogical Content Knowledge (TPACK)

Worked examples represent one very small instructional “move” within an expansive teaching repertoire. The challenge of contemporary research into worked examples is situating them within the bustling, organic context of the classroom and within the knowledge bank of expert teachers. Mishra and Koehler (2006) introduced the TPACK model as a conceptual

framework to advance discussions around educational technology. The TPACK model is most often illustrated as a Venn diagram with three intersecting circles (Figure 3) representing the three bodies of knowledge from which teachers must draw for effective instruction, to include content knowledge, pedagogical knowledge, and technological knowledge.

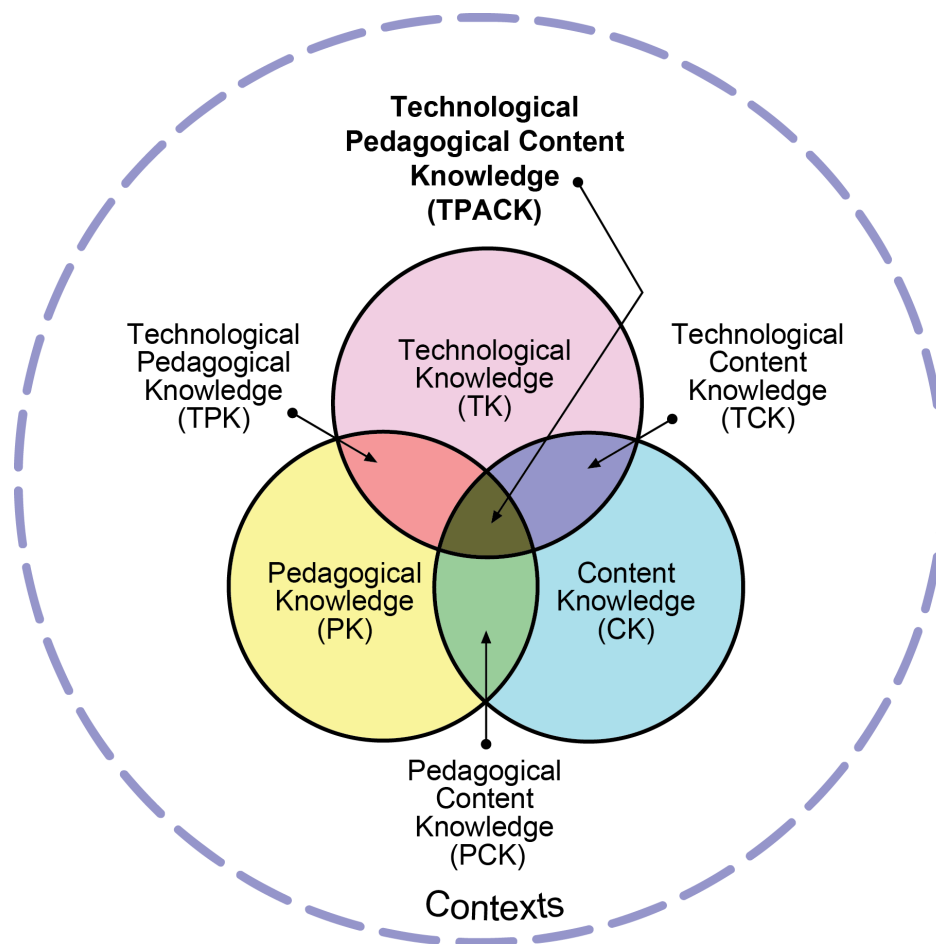


Figure 3. Technological pedagogical content knowledge. Reproduced by permission of the publisher, © 2012 by tpack.org

At the center is the synthesis of all three circles, technological pedagogical content knowledge, which Mishra and Koehler (2006) define as follows:

an understanding of the representation of concepts using technologies; pedagogical techniques that use technologies in constructive ways to teach content; knowledge of what makes concepts difficult or easy to learn and how technology can help redress some of the problems that students face; knowledge of students' prior knowledge and theories of epistemology; and knowledge of how technologies can be used to build on existing knowledge and to develop new epistemologies or strengthen old ones. (p. 1029)

In short, technological pedagogical content knowledge is a teacher's understanding of how and when to use a variety of technological tools to enhance teaching and improve learning within his or her content area.

Mishra and Koehler's (2006; Koehler & Mishra, 2009) work builds upon the seminal studies conducted by Lee Shulman (1986, 1987). Frustrated by the incessant pendulum swing in teacher education between content (a focus on the teacher's knowledge of his or her subject matter) and pedagogy (a focus on the teacher's knowledge of teaching methodology), Shulman (1986) called for a "more coherent" approach to identifying effective instruction and requisite knowledge for teaching (p. 9). Shulman argued that knowledge of what to teach and how to teach it must not be mutually exclusive. He proposed pedagogical content knowledge as that which separates mediocre teachers from masterful ones (Shulman, 1987). Pedagogical content knowledge includes, "for the most regularly taught topics in one's subject area, the most useful forms of representation of those ideas, the most powerful analogies, illustrations, examples, explanations, and demonstrations—in a word, the ways of representing and formulating the subject that make it comprehensible to others" (Shulman, 1986, p. 9).

Just as Shulman (1986, 1987) identified a problematic cleavage between knowledge of content and knowledge of pedagogy in teacher preparation, Mishra and Koehler (2006)

diagnosed a similar chasm between a teacher's knowledge of technology and knowledge of pedagogy and content. They affirmed the image of teachers as skilled craftspeople who draw upon a varied repertoire of methods and techniques, and who can assess and match the appropriate method to the learning situation and the learner. As disruptive technologies such as the Internet and personal computing devices have made their way into classrooms, Mishra and Koehler (2006; Koehler & Mishra, 2009) contested that knowledge of teaching and knowledge of technology could no longer remain a dichotomous pair. Just as a major scientific discovery or breakthrough in research would rightly force an educator to rethink the content he or she teaches, so too should the arrival of a technological innovation "[force] educators to think about core pedagogical issues" (Mishra & Koehler, 2006, p. 1029).

The present study is deeply embedded within the practitioner experience and within the ever-changing landscape of the high school classroom. It builds upon the early theoretical research into cognitive load and the predominantly clinical research into worked examples. As part of the natural evolution of research, this study joins with the many others that came before it to build a bridge between theory and practice. The goal of this study, therefore, is both to add to the theoretical research base around worked examples and to contribute to classroom teachers' technological knowledge repertoire to enhance the pedagogical content knowledge they already possess.

Background and Need

Studies of reading comprehension are some of the oldest in the field of educational research (Frey & Fisher, 2013; Webb, 1856). So much has been written about the subject of reading comprehension in the past 160 years that combing through it to tease out connections and solutions requires reading through a lens. This section sets the backdrop for thinking about

reading comprehension as a challenge deeply rooted in the functions and limitations of the human cognitive architecture. It brings into alignment what is known about why analytical reading is complex and what habits proficient readers possess with the theoretical framework of cognitive load theory, element interactivity, and schema acquisition. Furthermore, it connects think-alouds and multimedia worked examples as similar scaffolds for learning through modeling and advances the need for the present study.

Analytical Reading Comprehension as a Complex Cognitive Process

Acknowledging the importance of reading comprehension for students and the apparent gap between desired skill level and actual performance, it is imperative first to explore what makes analytical reading so challenging for adolescent readers. At the heart of the challenge is the fact that analytical reading is a complex cognitive task (Appleman, 2010; Frey & Fisher, 2013; National Reading Panel, 2000; Schoenbach et al., 2012) that requires the reader to creatively build mental structures by “select[ing] relevant information, mentally organiz[ing] it into a coherent structure, and integrat[ing] it with relevant prior knowledge activated from long-term memory” (Smagorinsky & Mayer, 2014, p. 611). Analytical reading of literary texts demands that students have command over the meaning of a text and that they be able to critically evaluate the message, remember the content, and apply the newfound knowledge flexibly (Alfassi, 2004). More specifically, true comprehension requires students to have mastered a set of basic reading fluency skills, such as recognizing phonemes, decoding words, developing automaticity in decoding, and accessing word meaning from the long-term memory; to use proficiently a set of more advanced reading comprehension skills, such as activating prior knowledge, identifying and using prose structures, making inferences, and using metacognitive knowledge; and to have command over task-specific reading skills, such as leveraging genre-

specific interpretative strategies (Kintsch, 1988; Smagorinsky & Mayer, 2014; van Dijk & Kintsch, 1983). Literacy researchers Freebody and Luke (1990) describe this process as the transition from “code breaker” to “meaning maker” and eventually to “text critic” and “text user.” For students to be successful in the high school English classroom, they must be able to engage in all four steps of this critical literacy process (Lattimer, 2010).

Moreover, analytical reading of literary texts requires students to use these cognitive processes to make meaning of texts with ever-increasing complexity (Appleman, 2010; Common Core State Standards Initiative, 2012; Frey & Fisher, 2013). From a cognitive load theory perspective, complexity is attributed to element interactivity (Pollock et al., 2002). In any work of literature, there are a number of moving parts that are inextricable. Students must monitor both text and subtext—what a text *says* and what a text *means* (Appleman, 2010). Text complexity arises from factors such as text purpose, levels of meaning, text structure, language conventionality and clarity, and the knowledge demands that it places on the reader (Frey & Fisher, 2013). See Table 1 for a comprehensive list of qualitative factors that contribute to text complexity.

Table 1

Qualitative Factors of Text Complexity

Component	Aspect	When a Text is Complex
Levels of Meaning and Purpose	Density and Complexity	Many ideas come at the reader, or there are multiple levels of meaning, some of which are not clearly stated.
	Figurative Language	There are many literary devices (e.g., metaphors, personification) or devices with which the reader is not familiar (e.g., symbolism, irony), as well as idioms or clichés.
	Purpose	The purpose is either not stated or purposefully withheld. The reader has to determine the theme or message.
Structure	Genre	The genre is unfamiliar or the author bends the rules of the genre.
	Organization	It does not follow traditional structures such as problem/solution, cause/effect, compare/contrast, sequence or chronology, and rich descriptions.
	Narration	The narrator is unreliable, changes during the course of the text, or has a limited perspective for the reader.
	Text Features	Fewer signposts such as headings, bold words, margin notes, font changes, or footnotes are used.
	Graphics	Visual information is not repeated in the text itself, but the graphics or illustrations are essential to understanding the main ideas.
Language Conventinality and Clarity	Standard English and Variations	Variations of standard English, such as regional dialects or vernaculars with which the reader is not familiar, are included.
	Register	It is archaic, formal, scholarly, or fixed in time.
Knowledge Demands	Background Knowledge	The demands on the reader extend well beyond his or her personal life experience.
	Prior Knowledge	The demands on the reader extend well beyond what he or she has been formally taught in school.
	Cultural Knowledge	The demands on the reader extend well beyond his or her cultural experiences and may include references to archaic or historical cultures.
	Vocabulary	The words used are representations of complex ideas that are unfamiliar to the reader, or they are domain specific and not easily understood using context clues or morphological knowledge.

Note. See Frey & Fisher, 2010, p. 9.

Beyond managing the cognitive load created by the text itself, adept readers are in constant reflective dialogue with the text, making personal connections that help to guide their reading (Beers, 2003). Alfassi (2004) calls this the dual cognitive process: students first make connections within the text, such as linking events and character relationships, and then make connections to their own lives, prior experience, and other texts they have read. The interplay between these two processes—deconstructing the text and then reconstructing it in connection to the reader—is where and how students make meaning (Alfassi, 2004). Even novice readers must simultaneously manage plot sequence, characters, motifs, symbols, archetypes, and themes, all in connection to one another and the purpose of the text as a whole (Barnhouse & Vinton, 2012; College Board, 2016). It is for this reason that Langer (2002) refers to analytical reading of literary texts as “high literacy,” because it requires students to access “deeper knowledge of the ways in which reading, writing, language, and content work together” (p. 3). Each of these connected elements places a strain on the working memory and limits students’ capacity for creating schemas, processing new information, and integrating it with prior knowledge. In short, analytical reading is a high element-interactivity task and thus carries high intrinsic cognitive load.

Cognitive load theorists argue that effective instructional design liberates the learner to focus on schema acquisition instead of sifting through extraneous cognitive demands (Sweller, 1988, 2010a). For the teacher of reading, this means freeing up students’ working memory capacity by reducing extraneous load. Swanson (1993) asserted that “skilled readers can be characterized by their ability to retain information in memory while simultaneously processing other information” (p. 286). He suggested that this quality is important for reading because “incoming information must be temporarily preserved while other information is being acquired

or manipulated” (p. 286). Since the working memory is critical to reading comprehension, strategies that optimize cognitive load are essential for supporting novice readers. By balancing the load on students’ working memory, educators can equip students to focus on the components that require maximum attention and alleviate the multiple cognitive burdens of the task of reading. Much of the research into reading comprehension looks into determining the strategies employed by more adept readers and considering why those strategies work (Alfassi, 2004; Cuevas, Russell, & Irving, 2012; Swanson, 1993). What follows is a brief overview of the cognitive processes needed for developing analytical reading schema at the high school level.

Improving Analytical Reading Comprehension Through Schema Acquisition

Typically, by the time a student reaches a high school English class, he or she has made the transition from *learning to read* to *reading to learn*, or from reading fluency to reading comprehension (Smagorinsky & Mayer, 2014). Though much of the language acquisition process occurs naturally outside of the formal classroom setting (Vygotsky, 1930/1978), school-based instruction around *learning to read* takes place between kindergarten and third or fourth grade, during which time students practice making phonemic sounds, decoding or sounding out syllables and whole words, building fluency in reading aloud, and connecting a word’s sounds to the word’s meaning (Smagorinsky & Mayer, 2014). After fourth grade, standard literacy instruction shifts to *reading to learn*, as students are expected by this point to have automated necessary schema for reading fluency, thus freeing up working memory capacity for processing more complex literacy tasks. Research into the intersection of cognitive science and literacy reveals that four critical cognitive processes occur during reading comprehension for more advanced readers that support schema acquisition: using prior knowledge, using prose structure,

making inferences, and using metacognitive knowledge (Smagorinsky & Mayer, 2014; see Table 2).

Table 2

Cognitive Processes in Reading Fluency and Comprehension

Name	Example Task
Recognizing phonemes	Substitution of first phoneme: You hear the word <i>ball</i> and are asked to change the /b/ sound into a /t/ sound.
Decoding words	Word identification: Pronounce the printed word <i>cat</i> . Word attack: Pronounce the printed word <i>blud</i> .
Decoding words fluently	Read a paragraph aloud fast and without error.
Accessing word meaning	Give a definition for a word and use it in a sentence.
Using prior knowledge	Reorganize the material to fit with an existing schema.
Using prose structure	Determine what information is important in a passage.
Making inferences	Attribute a motive to justify a character’s action.
Using metacognitive knowledge	Find a contradiction in a passage.

Note. See Smagorinsky & Mayer, 2014, pp. 608, 611.

Prior knowledge. First, advanced readers rely on prior knowledge to help them make sense of new material, using this knowledge “to guide how they select, organize, and integrate incoming information” with existing schemas (Smagorinsky & Mayer, 2014, p. 611). Activating prior knowledge often consists of readers making connections between the text at hand and other experiences they have had, other texts they have read, and observations of the world around them. Commonly, English teachers help prime prior knowledge by asking for “text to text, text to self, and text to world” connections (Keene & Zimmerman, 2007). This classroom practice is substantiated by research: In a synthesis of 17 meta-analyses of 3,607 studies ($n = 387,690$),

prior knowledge, as measured by proxy through prior achievement, accounts for a moderate effect size of $d = .67$ (Hattie, 2009; Fisher, Frey, & Hattie, 2016). Prior knowledge is an excellent predictor of future reading comprehension performance (Fisher et al., 2016; McNamara & Kintsch, 1996). Readers who possess and use relevant prior knowledge consistently outperform those who do not (Bransford & Johnson, 1972; Lipson, 1983; Pearson, Hansen, & Gordon, 1979; Pichert & Anderson, 1977). These readers have a “strong conceptional framework within which to situate new information and ideas” (Lattimer, 2010, p. 20) and they are much more likely to successfully register, recall, and retrieve new knowledge (Bransford, Brown, & Cocking, 2000). Thus, the importance of prior knowledge cannot be overstated. Prior knowledge acts as a schematic shortcut, linking new information to the long-term memory more efficiently than the working memory can (Keene, 2010).

Prose structure. Second, advanced readers rely on their knowledge of standard prose and genre-specific structures and use that knowledge within a new text to outline main points, summarize, recall, predict events, and evaluate details to determine which ones are most salient (Smagorinsky & Mayer, 2014). Studies of how disciplinary experts approach a new text within their field demonstrate that the simple task of reading changes dramatically based on the genre of the text, and that advanced readers require an additional, more specific form of prior knowledge: disciplinary knowledge (Shanahan, 2009). Disciplinary knowledge encompasses the formal rules, generalizations, and traditions of a discipline, and it includes the academic language germane to that discipline (Shanahan, 2009). In the case of literary reading, disciplinary knowledge also includes knowledge of specific genres and particular authors’ styles (Keene & Zimmermann, 2007). For example, advanced readers might begin reading a new novel with a repertoire of knowledge of literary prose structures, such as Freytag’s dramatic plot pyramid

(Freytag, 1863), the frame narrative, or the bildungsroman, and could leverage that knowledge to help organize new information into an existing schema for the order of events in a novel. Not surprisingly, students who can anticipate how a text will be organized or what it might include have a significant edge over those who are unfamiliar with the genre (Herber, 1978; Lattimer, 2010; Vaca & Vaca, 2008). Knowledge of prose structure affects comprehension. In studies of reading comprehension and text structure, more advanced readers are consistently more adept at culling information from a text and successfully placing it into a hierarchy of most to least important, whereas novice readers cannot distinguish between important and unimportant information (Brown & Smiley, 1977; Taylor, 1980). In a meta-analysis of 45 studies of students in Grades 2 through 12, the average weighted effect size for text structure instruction on reading comprehension was $d = .57$, which is typically interpreted as a moderate effect (Herbert, Bohaty, Nelson, & Brown, 2016).

Inferences. Third, advanced readers use inferencing strategies to construct meaning of a text while they read, testing new information against prior knowledge to shore up inconsistencies or holes in the text (Smagorinsky & Mayer, 2014). In classroom instruction, this practice is often described to students as “reading between the lines” and is evident in students’ ability to visualize what they are reading, fill in the figurative blanks beyond what the author literally presents, and formulate interpretations based on textual observations (Appleman, 2010). Advanced readers ask questions of the text, “make predictions about text, confirm their predictions, and test their developing meaning as they read on” to create their own interpretations of what they are reading (Keene & Zimmermann, 2007, p. 260). Novice readers often lack both (a) the prior knowledge requisite to make the cognitive leaps required by inferencing and (b) the additional working memory capacity to consider and integrate multiple pieces of information

from disparate locations in the text (Allington & McGill-Franzen, 2009). Oakhill and Yuill (1996) found that novice readers consistently struggled to answer questions that required inference-making, even when a text was available for review. In another study on inference-making, Paris and Lindauer (1976) found that whereas advanced readers were able to conclude what text suggested or implied when only implicit cues were presented (such as inferring from “our neighbor unlocked the door” that the neighbor had a key), novice readers were unable to decipher meaning unless explicit clues were presented. Thus, advanced readers have developed a cognitive flexibility that novice readers do not yet possess, in that advanced readers can “consider multiple elements of a text simultaneously and relate those text elements to prior knowledge,” thus making “multiple mental representations” of a text (Cartwright, 2009, p. 126).

Metacognitive knowledge. Fourth, advanced readers rely on a process of metacognition to monitor their comprehension, clarify meaning as they read, and revise meaning as needed (Appleman, 2010; Flavell, 1973; Greeno, Collins, & Resnick, 1996; Keene & Zimmermann, 2007; Smagorinsky & Mayer, 2014; van Dijk & Kitsch, 1983). Simply stated, metacognition is “thinking about the content and processes of one’s mind” (Winne & Azevedo, 2014, p. 63). In a synthesis of reading research on metacognition, Baker and Brown (1984) concluded that novice readers do not activate the same cognitive monitoring strategies as their more proficient peers. At the heart of metacognitive knowledge is the recognition that reading is an engaged process that requires the reader to interact with and make sense of the text—to ask questions, make connections, restate language from the text into one’s own words, and identify when and where meaning breaks down and actively work to repair it (Keene & Zimmerman, 2007; Lattimer, 2010; Tovani, 2000). Advanced readers evaluate their progress and monitor their comprehension as they read by silently asking themselves questions such as, “Is there something I don’t

understand? Am I learning the material? Are there any gaps in my knowledge or understanding? If I do find a gap in my knowledge, do I know what to do about it?" (Dole, Nokes, & Drits, 2009). It is no surprise that this process of reflection on meaning-making is effective in bringing about deeper comprehension of a text. In a meta-analysis of 63 studies ($n = 5,028$), Hattie (2009) reported that specifically training students to use metacognitive strategies in analytical reading accounted for a moderate effect size of $d = .69$ on tests of student achievement. Moreover, students who engage in metacognitive processes while reading are more likely to transfer what they are learning to new settings, underscoring the role of metacognition in schema development (Bransford et al., 2000; Palincsar & Brown, 1984).

As literary readers become more proficient at practicing the cognitive processes described above, they develop sets of complex cognitive maps that help them navigate new texts. These maps are the schemas that free up processing power in the working memory to comprehend more and more rigorous texts with greater ease. Advanced readers construct and refine schemas by activating prior knowledge, connecting to disciplinary text structure knowledge, drawing inferences, and monitoring comprehension and meaning. Not all readers will naturally adopt the four crucial cognitive processes necessary for analytical reading comprehension. Some readers need instructional scaffolds that explicitly support the construction of schemas. These scaffolds must reduce extraneous load, focus the learner on integrating new learning with prior knowledge, and engage the learner in creating or borrowing structures and organizational models for making sense of new information. One such scaffold is teacher modeling, which makes expert thinking visible.

Learning Through Modeling: From Think-Alouds to Multimedia Worked Examples

Research underscores the connection between improved reading comprehension and cognitive modeling from teachers, referred to in practitioner circles as thinking aloud. Rupley, Blair, and Nichols (2009) contend that teacher modeling is at the heart of high-quality reading instruction. Teachers who demonstrate what it looks like to read deeply in an authentic reading situation and who think aloud for their students help students “conceptualize reading skills and strategies, and how to apply them” (p. 127). The practitioner’s think-aloud has its roots in the Vygotskian theory of cognitive modeling by a more knowledgeable other, wherein students learn by imitating the actions of the adults around them (Vygotsky, 1930/1978). Instructional scaffolds and teacher guidance have the potential for moving students from their actual and independent developmental level to their potential developmental level, known as the Zone of Proximal Development (Vygotsky, 1930/1978, p. 85). According to Vygotsky, cognitive development in general, and language development in particular, are socially mediated processes (Vygotsky, 1930/1978), meaning that students learn best by “participating in activities with ‘more competent others’ who provide support for the parts that they cannot yet do by themselves” (Schoenbach et al., 2012, p. 21). Cognitive modeling provides students insight into the internalized thinking and self-regulation, or “inner speech,” that happens during the thinking process of a more knowledgeable other (Vygotsky, 1934/1962).

More specifically, modeling through a teacher think-aloud (Beers, 2003; Davey, 1983; Olshavsky, 1977; Wilhelm, 2001) works because it makes visible the thinking process that happens invisibly during reading, providing students with a clear schematic model to imitate and assimilate. Think-alouds support novice readers within the Zone of Proximal Development by providing them with the opportunity to access a complex text without being overburdened by the

cognitive load required to understand it independently (Frey & Fisher, 2013). During a think-aloud, “it is the teacher who is assuming most of the cognitive load” (Frey & Fisher, 2013, p. 26), because it is the teacher who is applying cognitive processes to become the code breaker, meaning maker, text critic, and text user on behalf of his or her students (Freebody & Luke, 1990). Think-alouds help the learner subvert the brain’s randomized guess-and-check process of meaning-making by temporarily borrowing the teacher’s analytical reading schema, thus reducing extraneous load and returning working memory function to schema building (Renkl, 2014; Sweller, 2006).

Though the think-aloud is lauded as the gold standard for elementary reading instruction (Keene & Zimmermann, 2007), there are some notable challenges with its current incarnation as a whole-class, teacher-paced, one-shot experience. When viewed from the perspective of cognitive load theory, the think-aloud poses potential problems for more advanced students in a heterogeneous class. First, when the think-aloud is used for whole-class instruction, stronger readers who have existing schemas for understanding the text may experience higher extraneous cognitive load, because learning a new schema that may interfere with their own existing schemas forces them to attend to unnecessary cognitive demands (Sweller, 2010a). This effect, in which the expert learner is disadvantaged by using the scaffolds that are necessary for the novice, is called the Expertise Reversal Effect (Kalyuga, Ayres, Chandler, & Sweller, 2003; Plass et al., 2010). Moreover, when think-alouds are teacher-paced and performed “live” in class, they deny struggling students the opportunity to slow down or revisit the teacher’s cognitive modeling, especially when struggling readers are reluctant to ask questions during class or to ask a teacher to repeat segments of a lesson (Tankersley, 2005). Research into cognitive load theory and instructional design for complex learning suggests that students benefit from retaining control

over the pace of the instruction they are processing (van Merriënboer & Kester, 2014). Teachers need a solution for modeling reading that respects the prior knowledge and learning needs of all students, and that considers the opportunities afforded by technological tools in addressing challenges of reading comprehension (Cuevas et al., 2012).

The Need for the Study

Reading researchers espouse the importance of instructional strategies that equip students to build analytical reading schema of their own and have relied on the classroom-tested practice of think-alouds for decades (Frey & Fisher, 2013). Despite the popularity of think-alouds in classroom instruction, and the pervasive recommendation toward think-alouds in practitioner texts (Appleman, 2010; Beers, 2003; Burke, 2000; Keene & Zimmermann, 2007; Milner et al., 2011; Wilhelm, 2001), studies into think-aloud methodology and efficacy have centered on early literacy instruction during the “learning to read” years, rather than on adolescent literacy during the “reading to learn” years (Baumann et al., 1993; Beers, 2003; Davey, 1983; Olshavsky, 1977). Little is known about the statistical significance of the think-aloud strategy as it pertains to student cognition or reading comprehension in adolescent readers.

Think-alouds are a natural match for the cognitive modeling presented in multimedia worked examples, which use 1:1 technology to allow for student-paced study of an expert’s detailed solution to a problem. Although multimedia worked examples have historically been used with math and science (Atkinson et al., 2000; Renkl & Atkinson, 2010; Sweller & Cooper, 1985), they are a statistically effective form of teacher modeling (Hattie, 2009) with a legacy of reducing extraneous load in students, optimizing cognitive load, and reducing the mental effort required of the working memory to comprehend complex, ill-defined subjects (Kyun et al., 2013; Nievelstein et al., 2013; Renkl, 2014; Renkl & Atkinson, 2010; Rourke & Sweller, 2009).

Thoughtful application of multimedia worked examples to the classroom practice of the think-aloud could yield similar benefits for students' reading comprehension, improving the teacher's options for effectively differentiating instruction for advanced and struggling readers alike.

A logical question, therefore, is why this application of worked examples to analytical reading think-alouds has not already been examined by researchers or teachers. One possible explanation is that worked examples research, especially with ill-defined problems or multimedia presentations, is a relatively new line of research; in an effort to define the theory as cleanly as possible, researchers may favor either problems that are unambiguous or settings that are more controlled and lab-like, such as a first-year university psychology class with a few hundred students who can study a series of computer models in a timed, succinct intervention (Renkl, 2014). For a number of reasons, such as the age and maturity of students, the nuanced nature of the topic, and the naturalistic setting, studying worked examples with adolescent reading comprehension is a messy proposition.

Another possible explanation for the dearth of research in this arena is the shift in perceived need around adolescent literacy instruction at the high school level. Many secondary English teachers assume that their role is to be a literature teacher, not a reading teacher, rationalizing that students learn how to read in the younger grades and no longer need this basic level of instruction (Appleman, 2010). The think-aloud has traditionally been a live, whole-class instructional strategy for early literacy development, and as such, may be dismissed because it takes up valuable class time that is needed for coverage of literary content. The current study may offer a solution for literature teachers to support students' cognitive processing of complex texts through a differentiated approach that does not compete for live class time.

Multimedia worked examples have a proven record of supporting students in processing complex information by reducing extraneous cognitive load, thus decreasing the total mental effort necessary to comprehend challenging content. In an early study of mental effort and classic worked examples, Paas (1992) hypothesized that learning through worked examples would require the student to expend less mental effort—or less of the available cognitive capacity in the working memory—than learning through traditional problem-solving. Subsequent studies of worked examples for both well-defined and ill-defined problems have applied the Paas (1992) model to measure mental effort, and their findings have confirmed Paas’s hypothesis. Of particular interest to the current study of reading comprehension is research into ill-defined problems, and more specifically, ill-defined problems in the humanities. Researchers in this narrowed field have found that studying worked examples of ill-defined problems, such as legal reasoning (Nivelstein et al., 2013) or English literature essay composition (Kyun et al., 2013), demands less mental effort than the comparison condition, wherein students learn by solving problems independently with the support of written process steps. Multimedia think-aloud worked examples have the potential to optimize cognitive load by reducing the strain on the working memory and returning student focus to schema construction or germane load processing. At this point in the history of worked examples research, it is unclear whether the expertise reversal effect (Kalyuga et al., 2003; Plass et al., 2010) applies to ill-defined problems such as the one in the present study. It is possible that multimedia worked examples with ill-defined problems benefit both advanced and novice students, as in the Nivelstein, van Gog, van Dijk, and Boshuizen (2013) study, but it is also possible that advanced students will experience these worked examples as excessive scaffolding that hinders their performance, as with the Kyun, Kalyuga, and Sweller (2013) study.

Furthermore, multimedia worked examples have a statistically significant effect on increasing students' comprehension and transfer (Cooper & Sweller, 1987; Hattie, 2009). In studies of ill-defined problems, such as writing learning journals (Hübner, Nückles, & Renkl, 2010), identifying designer styles (Rourke & Sweller, 2009), and interdisciplinary collaboration (Rummel, Spada, & Hauser, 2009), students who learned through worked examples, rather than the comparison condition of traditional problem-solving, scored significantly higher on direct transfer and near transfer tests. Much of what has been written about the classroom practice of the think-aloud is theoretical in nature (Kintsch, 1988; Pressley & Afflerbach, 1995), provides only anecdotal evidence of efficacy (Oster, 2001), or focuses on early elementary students learning how to read (Baumann et al., 1993). In contrast, studies of multimedia worked examples in ill-defined domains predominantly feature older adolescent students (Renkl, 2014) and draw upon a corpus of research from three decades of studies to substantiate claims of improving student comprehension. Applying the vast knowledge base of worked examples to a critical classroom practice of the think-aloud is an unprecedented and exhilarating step forward for both bodies of research.

Perhaps most important to the advance of think-aloud pedagogy, multimedia worked examples improve the teacher's opportunities for effective differentiation by offering just-in-time support to students who need it. Research into cognitive load theory and worked examples underscores the importance of appropriately matching instructional scaffolds to individual student learning needs. Although students with low levels of prior knowledge certainly benefit from teacher modeling, students with higher levels of prior knowledge are hindered by the additional support (Plass et al., 2010), and often, the worked example effect disappears entirely for these advanced students (Renkl & Atkinson, 2010). Because of their digital format,

multimedia worked examples give teachers the flexibility to electronically deliver scaffolded instruction to low priors without interfering in the cognitive processing of higher priors. Moreover, students benefit from the self-paced nature of multimedia worked examples (van Merriënboer & Kester, 2014). Because each individual student has control of his or her own laptop, tablet, or phone in a 1:1 environment, struggling students may listen multiple times, pause, slow down, and even repeat the lesson, allowing students who would never ask a clarifying question or request additional help in class the chance to access valuable instruction (Neebe & Roberts, 2015).

Thus, multimedia worked examples that explicitly model an expert reader's cognitive processes may prove to be a high-leverage strategy for educators interested in differentiated analytical reading instruction for adolescents. It is the aim of this study to determine if multimedia think-aloud worked examples could be an effective means of supporting struggling readers by increasing analytical reading comprehension while reducing working memory mental effort and optimizing cognitive load.

Research Questions

This study considered three research questions. The first two questions are quantitative in nature. Given the small sample size ($n = 34$) of the study, the final qualitative question is designed to help explain any differences that are not detected by the quantitative measures.

1. Analytical Reading Comprehension. To what extent is there an effect of multimedia think-aloud worked examples on analytical reading comprehension in a secondary English language arts class, as measured by open-ended written responses to literature?
2. Mental Effort. To what extent is there an effect of multimedia think-aloud worked examples on cognitive load, as measured by self-reported mental effort (Paas, 1992)?

3. Student Experience. What do student annotations and interviews reveal about the participant experience of studying multimedia think-aloud worked examples?

Definition of Terms

1:1 Technology: The ratio of one-to-one indicates that there is one device, be it a computer, tablet, or laptop, for every student in the class. There are three typical versions of 1:1 programs in schools: some schools use a device-cart model in which students retrieve devices at the start of a class period and return them at the end; other schools have devices assigned to students all school-year long so that students may use the devices outside of a given class period; other schools follow a BYOD (Bring Your Own Device) model in which students supply their own devices and bring them to class each day (Neebe & Roberts, 2015).

Analytical Reading Comprehension: The RAND Reading Study Group (2002) defines reading comprehension as the “process of simultaneously extracting and constructing meaning through interaction and involvement with written language” (p. 11). Analytical reading comprehension specifies the type of thinking in which students engage during that meaning-making process. Specifically, adolescent analytical reading comprehension of literary texts includes an understanding of the subtextual elements of literature, to include theme and the development of thematic ideas; character development, motivation, and conflict; denotative, connotative, and figurative meaning of language; structure, pacing, and organization of a story; author’s use of language in relation to tone, a sense of time and place; and so on (Common Core State Standards Initiative, 2012).

Cognitive Load Theory: A psychological theory and set of instructional principles based on knowledge of human cognitive architecture as an information processing system, in

which the working memory is limited and the long-term memory is essentially infinite (Moreno & Park, 2010; Paas & Sweller, 2014).

Cognitive Theory of Multimedia Learning: A theory for instructional design based on the central premise that people learn more deeply from words and pictures than from words alone. The theory stipulates that people have two separate channels for processing verbal and visual material (dual-channel assumption), and that each channel is limited in how much it can process (limited capacity assumption). Meaningful learning requires active engagement in the appropriate cognitive processes during learning (active processing assumption) (Mayer, 2014a).

Differentiation: An instructional practice that flexibly adapts content, process, and product for individual students or groups of students to effectively reach learners who range in readiness, interest, and experience (Tomlinson, 2014).

Effect Size: A statistical measure that describes “how big the difference is between group means, or how impressive the relationship is between variables,” and thus provides a sense of the practical importance of a given phenomenon (Mitchell, 2015). Effect size is most commonly measured using Cohen’s d , which—much like a z-score—uses standard deviation units to quantify the difference between two group means (Cohen, 1988). An effect size of $d > .80$ (eight tenths of a standard deviation difference) is considered a large effect; $d > .50$ (half of a standard deviation difference) is considered a moderate effect; $d > .20$ (two tenths of a standard deviation difference) is considered a small effect; and $d < .20$ is considered as having no effect.

Element Interactivity: The extent to which understanding one piece of information depends on understanding another related piece of information, or the extent to which understanding the “whole” requires understanding a number of subordinate parts. (Pollock et al., 2002).

Expertise Reversal Effect: An effect that occurs when an expert learner is disadvantaged by using the scaffolds that are necessary for the novice (Kalyuga et al., 2003; Plass et al., 2010).

Extraneous Cognitive Load: The form of cognitive load that is imposed by unnecessary cognitive demands in the learning process, such as nonessential tasks that do not lead to the acquisition and automation of schemas (Sweller & Chandler, 1994). Extraneous load is attributed to suboptimal instructional design (Sweller, 2010b).

Germane Cognitive Load: The form of cognitive load that is dedicated to developing and automating schemas (Moreno & Park, 2010).

Intrinsic Cognitive Load: The form of cognitive load that is imposed by the inherent difficulty of the information being learned (Sweller, 1994).

Long-Term Memory: In human cognitive architecture, the long-term memory is the vast, potentially infinite storage base for complex networks of information (Sweller, 2012).

Mental Effort: A proxy measure for cognitive load that asks learners to rate their perceived cognitive load on a subjective Likert scale by discerning the amount of mental effort expended on processing a particular task (Brünken, Seufert, & Paas, 2010).

Metacognition: Actively monitoring and regulating one’s own cognitive processes during learning (Flavell, 1973).

Multimedia Worked Example: Computer-based presentations that use 1:1 technology to allow for student-paced study of an expert’s detailed solution to a problem during the initial phase of cognitive skill acquisition (Renkl, 2014).

Pedagogical Content Knowledge: Teachers' knowledge of powerful analogies, illustrations, examples, explanations, demonstrations, and other forms of representation of regularly taught topics in their subject area, such that they are skilled in making the subject matter comprehensible to others (Shulman, 1986).

Self-Explanation: Explanations provided by learners that articulate the "inner speech" that happens during the thinking process (Vygotsky, 1934/1962). Specifically as they relate to worked examples, "self-explanations contain information that is not directly given in the learning materials and that refers to the worked example features and the reasons for them" (Renkl, 2014, p. 407).

Scaffolding: Instructional supports provided during the learning process that embed guidance in context (Reiser & Tabak, 2014); instructional supports are slowly removed over the course of skill acquisition, which is what makes them akin to scaffolding in building construction (Wood, Bruner, & Ross, 1976).

Schema: Representations of knowledge that are necessary for transferring information from the working memory to the long-term memory (Moreno & Park, 2010). They connect knowledge in "stable patterns of relationships between elements" and "can be linked together and organized into hierarchical structures" (Kalyuga, 2010, p. 48).

Technological Pedagogical Content Knowledge: A teacher's understanding of how and when to use a variety of technological tools to enhance teaching and improve learning within his or her content area (Mishra & Koehler, 2006).

Think-Aloud: An instructional strategy for supporting developing readers in which the teacher models the strategies of a successful analytical reader, such as activating prior knowledge, decoding text at multiple levels, making predictions, visualizing,

summarizing, asking questions, making connections, and monitoring and clarifying understanding (Appleman, 2010; Wilhelm, 2001).

Worked Example: Instructional scaffolds that depict an expert's detailed solution to a problem for students to study and emulate (Atkinson et al., 2000).

Working Memory: The part of the brain responsible for consciously processing and assimilating new information. The working memory has a fixed capacity and can manage only a finite number of elements at once (Moreno & Park, 2010).

CHAPTER II

REVIEW OF THE LITERATURE

This study of multimedia think-aloud worked examples naturally has its roots in two lines of research: think-alouds and worked examples. The former provides the content-specific strategies that will be utilized in the instructional intervention around analytical reading, whereas the latter provides the conceptual framework for the design of the study. The review of the literature presented here begins by (a) briefly identifying the standard protocols used in think-alouds for reading instruction and (b) reviewing the origins of the think-aloud as a product of a developing understanding of the cognition process during reading. It then (c) traces the lineage of worked examples research, from the earliest studies of well-defined problems through more contemporary studies of ill-defined problems and multimedia worked examples, culminating in a discussion of design principles for effective worked examples of ill-defined problems using a multimedia platform.

Think-Alouds in Literacy Instruction

In the domain of reading comprehension instruction, language and literacy learning has historically been situated within a sociocultural theoretical context in which language learning is defined as a socially mediated activity; learning develops when students engage with teachers, peers, parents, and more knowledgeable others (Bruner, 1978; Rogoff, 1995; Vygotsky, 1930/1978). In part, it is from this tradition that the classroom practice of the think-aloud was born. As an instructional scaffold, the think-aloud relies heavily upon “an aspect of social interaction” between the student and the teacher (Kucan & Beck, 1997, p. 272). Thus, from the surface it might seem that a cognitivist approach to literacy instruction, such as the one taken in

the present study, is out of step with the tradition of the research in this area. However, there also exists a long line of thinking, about the cognitive processes that occur during reading, that bridges the gap between the sociocultural theoretical tradition of literacy research and the cognitivist theoretical framework of this study.

Beyond its sociocultural roots, the think-aloud came about in large part as a response to the increasing influence of cognitive science on the study of education, which is evident in the increasing interest in defining the cognitive process of reading, starting especially in the 1950s. Since the early twentieth century, literacy researchers have sought to explain the internalized action that occurs during the process of reading, in an effort to improve reading instruction and student support (Kucan & Beck, 1997). Starting with Huey (1908), reading has been represented in ever-evolving terms: reading as thinking (Huey, 1908); reading as reasoning (Thorndike, 1917); reading as response to literature (Rosenblatt, 1938); reading as interpretation (Piekarz, 1956); reading as information processing (Gough, 1972); reading as problem solving with strategies (Olshavsky, 1977); reading as micro- and macro-structure discourse processing (Kintsch & van Dijk, 1978); reading as a function of schema (Anderson & Pearson, 1984); reading a collaborative meaning construction (Palinscar & Brown, 1984, 1988; Palinscar & Klenk, 1992); and engaged reading as the “joint functioning of motivational processes and cognitive strategies” (Wigfield et al., 2008, p. 432). What has remained constant is the interest in isolating, studying, and defining the invisible action of cognitive processing that takes place during reading, rather than approximating that action by studying reading products, such as interviews and notes recalling what was read (Kucan & Beck, 1997; Simons, 1971).

Think-alouds were first used in literacy studies as a measurement protocol to identify the strategies used by expert readers (Afflerbach & Johnston, 1984; Olson, Duffy, & Mack, 1984;

Pressley & Afflerbach, 1995). In a seminal text, *Protocol Analysis: Verbal Reports as Data*, authors Ericsson and Simon (1984/1993) outlined the procedures for using verbal self-report data in empirical studies with relatively well-defined tasks. They argued that people can effectively describe the workings of their short-term memory, because this is where conscious processing occurs; however, people cannot effectively describe the workings of their long-term memory. Ericsson and Simon's (1984/1993) protocol analysis was subsequently used to investigate the cognitive processes of expert readers in a series of think-aloud studies (Afflerbach, 1990; Caron, 1989; Hare, 1981; Hartman, 1995; Langer, 1990; Olshavsky, 1977). These studies considered reading as a product of problem solving and looked at expert think-alouds to help define the strategies that novice readers should adopt in order to become proficient.

In a meta-analysis of 38 think-aloud studies on reading comprehension, Pressley and Afflerbach (1995) categorized the findings into three kinds of reading strategies—before reading, during reading, and after reading—and identified an extensive list of strategies consistently used by expert readers. These strategies included, for example, setting a purpose for reading, previewing a text, activating prior knowledge, identifying important information, re-reading, restating or explaining text, constructing a main idea or gist, looking for patterns in the text, making predictions, adjusting previous hypotheses based on new information, making inferences, integrating different parts of the text, and interpreting the text (Pressley & Afflerbach, 1995, pp. 30–59). From this process of investigating “essentially invisible mental processes,” researchers such as Afflerbach and Johnston (1984) concluded that think-alouds “offer a unique, if sometimes less than transparent, window for viewing cognitive processes” during reading (p. 320). In turn, the expert strategies identified in the studies of the late 1970s to early 1990s became the basis of comprehension strategy instruction in the teaching of reading.

Once think-alouds were successfully employed with expert readers as a measurement tool, researchers applied the verbal protocol to studies of developing readers as a means of gauging the efficacy of specific instructional interventions (Pressley, Graham, & Harris, 2006). Think-alouds have been used to measure students' response to intervention in areas such as reader-text interest (Wade, Buxton, & Kelly, 1999) and strategic reading among young readers (Schellings, Aarnoutse, & van Leeuwe, 2006). In addition, think-alouds themselves have been used as an intervention. Researchers have tested the think-aloud protocol as a strategy for increasing students' metacognition and comprehension monitoring (Kucan & Beck, 1997). Studies have included, for example, think-alouds for increasing metacognition and comprehension monitoring in second-language middle-schoolers (McKeown & Gentilucci, 2007); think-alouds for prior knowledge and strategy use in college students (Crain-Thoreson, Lippman, & McClendon-Magnuson, 1997); think-alouds to practice applying text-processing strategies in sixth graders (Silven & Vauras, 1992); and think-alouds to enhance children's comprehension monitoring abilities (Baumann et al., 1993).

Most important to the present study, out of this line of empirical research that uses think-alouds as a measurement protocol, and based in large part on the expert reader strategy findings, the think-aloud protocol was adapted by practitioners for classroom use. The teacher performs the verbal protocol while reading a passage from a shared text as a way of modeling the cognitive comprehension process (Bereiter & Bird, 1985; Davey, 1983; Keene & Zimmermann, 2007; Olson & Land, 2007). It is important to understand the instructional practice of think-alouds as the offspring of the measurement protocol of think-alouds and within the context of comprehension strategy instruction. Simply stated, literacy educators took a measurement tool and turned it into a teaching tool (Kucan & Beck, 1997).

Research into think-alouds for reading instruction is more limited. Much of the academic literature focuses on generalized suggestions for performing think-alouds, as in Collins and Smith's (1982) chapter on teaching reading comprehension: "As the text is being read, the teacher interrupts maybe once or twice a paragraph to make comments about . . . different aspects of the comprehension process" (p. 182). Many of the recommendations for *how* to perform think-alouds draw upon research into expert strategies and actually discuss *what* to perform in think-alouds: make predictions, visualize, activate prior knowledge, monitor comprehension, and use fix-up strategies when comprehension breaks down (Davey, 1983). Researchers have justifiably focused first on the content of reading comprehension instruction; however, much less is known about the specific moves that create an excellent think-aloud.

Despite the gap in theoretical knowledge, there exists a tremendous body of literature on think-alouds that documents the wisdom of practice. The classroom protocol begins with the teacher stating a purpose for reading and cuing students to listen for the teacher's use of specific reading strategies, such as asking questions or making predictions (Appleman, 2010; Wilhelm, 2001). The teacher then reads aloud a short section from a shared text, pausing frequently to vocalize her thoughts, making accessible to students the internal dialogue that takes place between the mind of a skilled reader and the text (Beers, 2003; Keene & Zimmermann, 2007; Wilhelm, 2001). During the think-aloud, the teacher provides a verbal or visual cue to signal the transition from reading to thinking, and clarifies how a particular strategy helps the teacher to make meaning in a passage (Beers, 2003; Keene & Zimmermann, 2007). The goal of the teacher think-aloud is for students to adopt the habits of successful readers by hearing those habits modeled enough times that they can internalize the cognitive processes required for effective analytical reading (Frey & Fisher, 2013; Keene & Zimmermann, 2007). In many classrooms,

teachers then ask students to participate in thinking-aloud in pairs or small groups to externally reinforce the internal practice of self-explanation (Beers, 2003; Wilhelm, 2001). Table 3 depicts a selection of cognitive strategies used by expert readers with sentence stems for teachers to use in thinking-aloud.

Table 3

Cognitive Strategies to Model in Think-Alouds

Strategy	Description	Sentence Stem
Using Prior Knowledge	Activate relevant prior knowledge to help assimilate information from the text into existing schema; use schema to relate the text to other texts, personal experiences, or knowledge of the world; use schema for specific authors and their styles to better understand text.	I already know that ...
		This reminds me of ...
		This relates to ...
		I experienced this once when ...
Monitoring Comprehension	Identify when the text makes sense or does not; re-read for clarity; use fix-up strategies to repair meaning; check and revise evolving interpretations; pause to consider the text and reflect on understandings.	I got lost here because ...
		I'm guessing that this means ... but I still need to know ...
		At first I thought ... but now I ...
		I think what this is saying is ...
Asking Questions	Spontaneously generate questions before, during, and after reading; ask questions for different purposes, such as to clarify, predict, determine style, content, format, etc.; use questions to focus attention on important elements of the text.	I wonder why ...?
		What if ...?
		How come ...?
Drawing Inferences	Make conclusions, form unique interpretations, make predictions and test those	I'll bet that ...
		What this means to me is ...

	predictions, use text in combination with background knowledge to seek answers to questions.	I think this represents ...
Using Sensory and Emotional Images	Create visual, auditory, or other sensory representations of the text that draw upon emotional connections and prior knowledge; use images to draw conclusions, clarify or enhance comprehension, and to engage in the text.	I can picture ... In my mind I see ... If this were a movie ...
Determining What is Important in Text	Identify themes and key ideas; distinguish between important and unimportant information; use text features and text structures as a guide; prioritize focus on important parts.	So, the big idea is ... A conclusion I'm drawing is ... I think this part is important because ...
Synthesizing Information	Monitor overall meaning, key concepts, and themes; consider how text elements fit together to create an overarching idea, such as symbols, motifs, conflict; synthesize the text beyond the literal meaning to the inferential level; share, recommend, and critique.	The most important message from this text is ... I noticed a pattern with ... and think it means ... I like/don't like ___ because ...

Note. Adapted from *Mosaic of Thought* (Keene & Zimmermann, 2007, pp. 255–272), and “A Cognitive Strategies Approach to Reading and Writing Instruction for English Language Learners in Secondary School” (Olson & Land, 2007, p. 280).

Though the think-aloud has been most popular in elementary school literacy instruction, it holds tremendous promise as an instructional tool for adolescent literacy because it situates reading strategy instruction within the context of content instruction. In the same way that Shulman (1986, 1987) and Mishra and Koehler (2006) identified points of departure between the body of research supporting the need for teachers’ pedagogical knowledge, content knowledge, and technological knowledge, reading researchers have identified similar areas of friction within

the literacy community between those who advocate for supporting adolescent readers through a skills-based approach and those who promote a content-based approach (Learned, Stockdill, & Moje, 2011). Some assert that disciplinary knowledge is paramount to success, so they focus on teaching *content*, whereas others maintain that knowledge of reading strategy is necessary for understanding content, and thus they focus on teaching reading *process*.

More recently, though, there has been a call to integrate these seemingly dichotomous instructional aims (Appleman, 2010; Learned, Stockdill, & Moje, 2011). Literacy researchers and teachers of reading have pushed back against approaches that decontextualize discrete literacy skills from the process of reading interesting and challenging material (Lattimer, 2014; Schoenbach et al., 2012). They decry the skills-in-a-box solutions peddled by publishing companies and adopted by “failing” schools, criticizing them as stripped-down imitations of reading that unfortunately leave struggling readers even farther behind (Schoenbach et al., 2012). Analytical reading is complex exactly because it cannot be reduced to independent and objective skills; it has to be taught in context and with all its complex connected parts intact. It is for these reasons that think-alouds, in their incarnation as an instructional tool, are particularly well suited for supporting adolescent readers as they deepen their understanding of both literary content and literacy strategies. Given the roots of the think-aloud in the cognitivist line of thinking, as well as the need for more empirical research around the structures of a successful think-aloud, it is a sensible next step to link the classroom application of the think-aloud to a well-studied, high-leverage practice for cognitive modeling: the worked example.

First-Generation Worked Examples Research

Of the classic effects of cognitive load theory, worked examples are one of the strongest and best studied (Sweller, 2006). Research in this field began in earnest in the mid-1980s,

building off earlier studies into the distinction between experts' and novices' cognitive architecture (Chi, Feltovich, & Glaser, 1981; Chi, Glaser, & Rees, 1982; De Groot, 1966). These early studies established the basis for understanding expertise as the result of schemas, rather than as the result of a faster and more capacious working memory. Expert problem solvers, whether in chess or physics, relied on available mental constructs that equipped the participants to recognize patterns in problems, categorize new problems according to those patterns, and solve these new problems more efficiently than novices (Chi et al., 1981; Chi et al., 1982; De Groot, 1966).

Moreover, contemporaneous research surrounding the discovery learning movement revealed that even an abundance of problem solving practice did not yield schema acquisition (Sweller & Cooper, 1985). It seemed that learners could focus their attention on either (a) solving problems or (b) studying patterns and problem types that would help them develop and internalize schemas, but they could not do both simultaneously (Sweller & Cooper, 1985). Given the foundation of this early research into the expert-novice difference and the process by which a novice becomes an expert, Sweller and Cooper (1985; Cooper & Sweller, 1987) designed a series of experiments that used worked examples to target schema acquisition in students. These seminal studies are the first to expose a worked example effect, making a clear case for using worked examples to support meaningful learning.

Seminal Studies

Sweller and Cooper (1985) published the first empirical studies of worked examples and proposed that studying worked examples was a superior alternative to problem solving. In a series of five experiments, the researchers sought to determine if studying worked examples, instead of engaging in conventional problem solving, could increase students' schema

acquisition, increase their efficiency in solving near and far transfer problems, and improve their accuracy in solving problems. The first experiment was conducted simply to confirm the novice-expert distinction in the population of high school and university level mathematics students that the researchers would study in subsequent experiments. Students from three disparate levels of mathematics underwent a series of non-mathematics-related memory questions and a series of algebraic questions. The results indicated that, although all students had relatively similar scores on the random memory questions, the lowest-level mathematics students had the poorest memory of mathematics questions. These results confirmed previous research suggesting that the expert-novice divide rests along the line of schema, not working memory; that is, schemas are “an essential component of expertise” and knowledge acquisition depends on the extent to which a learner can acquire schemas (Sweller & Cooper, 1985, p. 69).

Experiments two through four each presented variations on the same core study: each sample included between 20 and 40 eighth- or ninth-grade algebra students; each experiment contrasted a worked examples treatment condition against a conventional problem solving comparison condition; each condition included between 10 and 12 participants; and each experiment was measured using the dependent variables of time taken and number of errors made. All students received the same initial instruction on the concepts, procedures, and rules needed to solve the problems in the test set. During the acquisition phase of the experiments, students in the comparison condition were asked to simply solve the problems, whereas students in the treatment condition were permitted to study the worked example before solving the problems. Despite small changes in study design—such as the number of questions assigned, the level of challenge in the worked examples studied, and the level of similarity between the

worked examples and the test problems—each experiment yielded similar, statistically significant results in favor of a worked example effect.

In the fifth and most rigorous of the experiments, Sweller and Cooper (1985) divided students into two blocks by matching students in pairs based on prior mathematics achievement. They then randomly assigned one block to the treatment condition and the other to the comparison condition. They found that studying worked examples went much faster ($\bar{x} = 58$ seconds) than solving conventional problems ($\bar{x} = 232$ seconds). More importantly, results from a Wilcoxon matched pairs signed ranks test suggest that students who studied worked examples, rather than practiced problem-solving, were able to solve similar, near transfer problems more quickly than their counterparts ($T(12) = 8$) and with fewer errors ($T(7) = 3$). Students in the worked examples condition were also less likely to rely on the inefficient strategy of algebraic expansion (in which the expression $c(a + b)$ unnecessarily becomes $ca + cb$) than their problem-solving peers ($T(5) = 0$). Sweller and Cooper's (1985) initial research called into question the "practice makes perfect" maxim supported by proponents of problem-solving as learning, and sparked subsequent studies of worked examples in mathematics as well as other well-defined disciplines.

Expertise reversal effect. Sweller and Cooper's (1985) first set of experiments did not, however, yield a statistically significant difference in efficiency or accuracy when it came to far transfer problems. In a later series of experiments, they investigated this shortcoming, hypothesizing that transfer did not occur in the 1985 study because of the extensive list of rules students had to learn in a limited window of time in order to perform well on the problems presented. Thus, the second series of experiments (Cooper & Sweller, 1987) was designed to better understand the conditions under which worked examples are effective. Most notably, in

their second experiment, Cooper and Sweller (1987) amassed a sample of 104 eighth graders, whom they split into eight groups of 13 students to run a full factorial experimental design along three independent variables: ability (high versus low), practice duration (short versus long), and instructional format during acquisition of content (worked examples versus conventional problem solving). They confirmed that students whose learning process featured a “heavy emphasis” (p. 347) on worked examples were more efficient and effective than their problem-solving peers. Also, they extended their own previous findings by noting that the worked example effect is most marked for students in the “low ability” group, and interestingly that effect “largely disappear[s]” when the high ability group is examined (p. 353).

Sweller and Cooper’s (1985; Cooper & Sweller, 1987) seminal studies of worked examples with unambiguous, well-defined algebra problems highlighted three defining traits that shaped the direction for the first generation of worked examples research. First, despite instruction presenting better, more efficient methods, students in the problem-solving condition in each of the eight experiments persisted in using the “guess and check” method rather than what they had been taught (Cooper & Sweller, 1987; Renkl & Atkinson, 2010; Sweller & Cooper, 1985). Second, students benefited from studying multiple worked examples during the skill acquisition process rather than a single worked example (Cooper & Sweller, 1987; Sweller, 2006; Sweller & Cooper, 1985). Third, worked examples appeared to be most effective for novices in the “early phases of skill acquisition” (Renkl & Atkinson, 2010, p. 93), rather than for students who already had developed schemas and automated rules for algebraic problem solving (Cooper & Sweller, 1987).

Cooper and Sweller’s (1987) insight into the efficacy of worked examples for low versus high ability groups laid the groundwork for later research into the expertise reversal effect, which

some studies of ill-defined problems have challenged (Niveau et al., 2013) and which this study considers. The expertise reversal effect occurs when instructional processes that benefit novices hinder expert learners. Plass, Kalyuga, and Leutner (2010) explain how expertise reversal functions in light of cognitive load theory:

The magnitude of mental load in learning depends on the schemas that have been previously acquired by the learner ... Although experts in a particular domain do not possess larger working memory capacities, they experience a decreased working memory load because they have larger organized knowledge structures stored in long-term memory. [Thus,] expertise may actually trigger additional cognitive load because experts have to process information that, given their high level of expertise in the given domain, is unnecessary for them to assure successful learning. The expertise reversal effect occurs when an instructional method that is effective for novices becomes ineffective for more knowledgeable learners. (pp. 67–68)

Simply stated, worked examples become less effective when students already have schema formed (Kim, 2005).

Multiple empirical studies of worked examples for well-defined problems, which compare students with low prior knowledge against students with high prior knowledge, reveal disordinal interactions for high prior students in the worked example condition, suggesting the existence of the expertise reversal effect (Kalyuga, 2007). Tuovinen and Sweller (1999) found an effect size of $-.31$ for high prior students in their study of worked examples versus exploratory-based learning; Kalyuga and Sweller (2004) found an effect size of $-.36$ for high prior students in their study of worked examples versus problem solving; Reisslein, Atkinson, Seeling, and

Reisslein (2006) found an effect size of $-.26$ for high prior students in their study of example-problem pairs versus faded worked examples.

Measuring the effect of worked examples. As worked examples research continued to develop, researchers sought a way to measure the efficacy of the intervention on reducing extraneous load, beyond proxy measures such as time on task or outcome measures such as number of errors made, as was the case in the Sweller and Cooper (1985; Cooper & Sweller, 1987) experiments. Paas (1992) introduced the Mental Effort Rating Scale in the *Journal of Educational Psychology* in “Training Strategies for Attaining Transfer of Problem-Solving Skill in Statistics: A Cognitive Load Approach.” Much like the seminal studies of Sweller and Cooper (1985; Cooper & Sweller, 1987), Paas (1992) studied the worked example effect using very well-defined problems in statistics.

The purpose of the Paas (1992) study was to test the difference in transfer performance, problem-solving time, and perceived mental effort in solving complex statistical problems between a conventional problem-solving method and problem solving using worked examples or completing partially worked examples. Paas (1992) hypothesized that the worked example and completion conditions would “require less mental effort and less instructional time than would the conventional condition,” thus resulting in higher transfer performance (p. 430). The major constructs investigated in this study include the independent variable: worked examples in two forms; and the dependent variables: mental effort, transfer, and speed of problem solving.

To measure perceived mental effort, Paas (1992) presented students with a symmetrical scale illustrated on a physical number line, with numbers ranging from 1 to 9 and corresponding Likert-style labels from “very, very low mental effort” to “very, very high mental effort” (see Figure 4). Upon completion of each of the 12 problems, students were asked to “rate the

perceived amount of mental effort invested in the problem” (Paas, 1992, p. 432). The Paas (1992) scale is a modified version of an earlier scale, also a 9-point symmetrical item, meant to measure a respondent’s perception of the difficulty of a task (Bratfisch, Borg, & Dornic, 1972). In the original study, participants were asked to estimate the difficulty of each item on a test; these estimations were later correlated with standard scores corresponding to the frequency of correct solutions. Bratfisch, Borg, and Dornic (1972) found a linear relationship between the two and reported a strong positive Spearman rank order correlation ($r \geq .92$). Paas (1992) reported similarly high reliability between the condition-independent instructional problems and the test problems, estimated with Cronbach’s coefficient alpha ($\alpha \geq .90$).

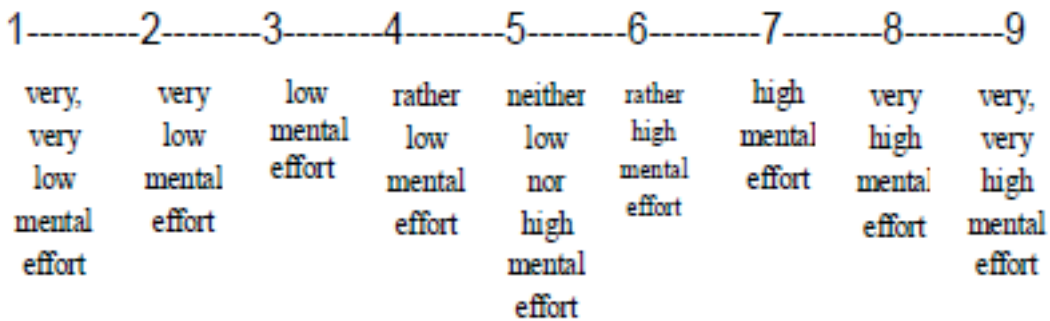


Figure 4. Paas (1992) Mental Effort Rating Scale

The sample for Paas’s (1992) quasi-experimental study included 16- to 18-year-old students ($n = 46$) attending a Dutch technical school; all but one student in the sample were male. Students in this convenience sample were recruited to participate as part of a business administration course. The population to which the Paas (1992) study could reasonably be generalized is male secondary school students in a pre-technical program; however, as this is a non-probability sample, generalizability is limited in the statistical sense. Students were randomly assigned to either the control group (conventional problem solving) or one of two

treatment groups (worked examples or completion of partially worked examples). All students were instructed to solve, study, or complete a series of 12 problems, and then immediately to rate their perceived mental effort. The computerized program that delivered the problems logged the time spent per problem. At the end of the experiment, all students took the same transfer posttest, which included 12 near-transfer problems and 12 far-transfer problems. The computerized program successfully gathered data for all but four students, whose results were excluded from the study.

The results of the study confirmed two of the three hypotheses. Paas (1992) used the omnibus Fisher F -test to detect between-group variability among the treatment and control groups on each of the dependent variables. He reported mean scores, standard error, and F statistics, but he did not report standard deviations or effect sizes (d) for the six dependent variables that were analyzed: time on general instruction, time on specific instruction, number of incorrect solutions generated, time on test, test performance, and perceived mental effort. According to Paas's analysis, students in the treatment groups outperformed the control group on the assessment of test transfer ($F = 13.55$) and took less time to solve problems ($F = 6.86$), suggesting that studying a worked example is a more effective and efficient way to learn. Post hoc analysis yielded significant condition effects on both the near-transfer ($F = 6.17$) and far-transfer problems ($F = 8.32$). Contrary to the hypothesis, the means in mental effort scores did not differ across the three conditions ($F = 0.74$).

Paas (1992) concluded that his study confirmed previous research into the worked example effect, despite the lack of variability in mental effort scores. Given that his conclusions were derived from a relatively small sample (13 in the conventional condition, 14 in the worked example condition, and 15 in the completion condition), which was made smaller by some

missing data from four students who had to be excluded from the initial sample, it is possible that some condition effects on the mental effort dependent variable were not detected. He did suggest that more research is needed on the topic of subjective measures of perceived mental effort. In the years since this study, many in the field of cognitive load research have employed the Paas scale in their own studies of worked examples to measure mental effort (Paas & van Merriënboer, 1993, 1994; Paas, Tuovinen, Tabbers, & Van Gerven, 2003), finding significant worked example effects on reducing mental effort.

Self-Explanation in Worked Examples

One critique of these initial studies was that learners in the treatment condition had the opportunity to remain fairly passive while studying the worked examples (Renkl & Atkinson, 2010). Educational research consistently highlights that meaningful learning requires active cognitive engagement in the meaning-making process (Mayer, 2014b). From the perspective of cognitive load theory, meaningful learning is about increasing generative processing or germane load (Mayer, 2014a). It was problematic, therefore, that in many of the early studies of worked examples, learners were characterized as superficially going through the motions of studying, but they were not actively engaged in germane load activities (Renkl, 1997). This perhaps explains the limitations in finding statistically significant main effects on far transfer tests. The existence of a worked example effect theoretically means that students studying worked examples are supported in the learning process by a reduction in extraneous load and a reciprocal increase in germane load; however, research suggests that, practically speaking, many students did not make productive use of the extra working-memory capacity freed up by the worked examples, and they failed to engage in schema development and acquisition (Renkl & Atkinson, 2010).

As a result, a series of studies was conducted to test whether worked examples that prompted or trained students to explain their thinking while learning could be more effective and more transferrable than standard worked examples in well-defined domains, such as physics (Chi et al., 1989), computer programming (Pirulli & Recker, 1994), and statistics (Atkinson, Renkl, & Merrill, 2003). In one such study, Chi, Bassok, Lewis, Reimann, and Glaser (1989) required students to verbally process their own natural thinking processes as they studied a set of worked examples of physics problems. The researchers then recorded, coded, and analyzed the students' verbal self-explanations. They found that students who could articulate inferences about how and why the worked examples progressed from one step to the next were far better equipped to apply the same procedure to a non-isomorphic problem (Chi et al., 1989). Stronger-performing students produced more abundant and more relevant self-explanations, such as ones that commented on the worked example's structure or content, paraphrased or restated what the example line said, or demonstrated the student's active monitoring of his or her understanding or comprehension failures. In confirmation of the Chi et al. (1989) findings, Renkl (1997) categorized students who supplied effective self-explanations as either anticipative reasoners, who anticipate the next step and cross-check against the example, or as principle-based explainers, who explain the principles and goals of an example and can elaborate on those principles. Chi (2000) later revised her conclusions to clarify the nature of effective self-explanations as a dual-process involving both making inferences (as stated in the initial study) and repairing one's own mental model when it conflicts with the model presented in the worked example (Chi, 2000).

In a more recent study of worked examples with self-explanation, the research team built upon previous studies by embedding self-explanation prompts within worked examples of

statistical probability problems. By encouraging students to describe the underlying patterns and principles present in each solution step, they fostered the inference and repair-making process (Atkinson et al., 2003). The sample included 40 high school students in advanced algebra who were randomly assigned to one of two conditions: worked examples with prompts for self-explanation, or worked examples without prompts for self-explanation. In the self-explanation condition, students were presented a list of the probability principles introduced prior to the experiment, and then were asked at the end of each step to select the principle being used. Both groups were tested on problems similar to (near transfer) and slightly different from (far transfer) those studied in the worked examples. The researchers found a statistically significant main effect for self-explanation on both near and far transfer tests, suggesting that students who studied worked examples with self-explanation prompting outperformed their unprompted peers. For near transfer, Atkinson, Renkl, and Merrill (2003) reported a Cohen's f statistic of .42, which corresponds with a large effect size. For far transfer, the researchers reported a Cohen's f statistic of .37, which also corresponds with a large effect size. These studies underscore the importance of supplementing worked examples with self-explanation for greatest impact.

Implications for Second-Generation Research

The first generation of worked examples research focused exclusively on supporting students' schema acquisition of the discrete solution steps needed to solve well-defined problems in domains such as algebra, physics, statistics, and computer programming. One clear question unaddressed by early worked examples research was whether this form of instructional intervention would be effective for teaching students how to solve ill-defined problems that didn't have established "algorithmic solution procedures" (Renkl & Atkinson, 2010, p. 106). If worked examples could be applied to ill-defined problems, three additional questions remained:

1. How might a worked example of an ill-defined problem describe the solution steps required to “solve” the problem?
2. How might the instructor structure the transition from studying worked examples to problem solving so that students practice both near- and far-transfer problems?
3. Would the expertise reversal effect still apply to problems with a potentially more flexible range of what it means to have expertise? (Renkl & Atkinson, 2010).

Each of these questions has been taken up by the second generation of worked examples research, which builds upon the foundation described above, while leaving open the door for the present study of multimedia think-aloud worked examples for the ill-defined problem of analytical reading.

Second-Generation Worked Examples Research

Recently, research has sought to extend the discussion of worked examples beyond well-defined problems to include ill-defined problems. Although past studies have articulated the effect of worked examples on well-structured tasks in fields such as mathematics, science, and mechanics, these studies have not looked as closely into ill-structured tasks for fields such as law, medicine, and the humanities. The difference is critical, since many professions require that people can think critically about how they choose which concept to apply, reason through unclear problems, and grapple with multiple possible outcomes to the same scenario. Although a worked example for a well-structured task can rely on concrete concepts to find an algorithmic solution, worked examples for ill-structured tasks cannot. Classic worked examples focus on reducing novice learners’ extraneous load by modeling solution steps for unambiguous, highly structured tasks.

Worked examples of well-defined problems “rely on the application of a limited number of concepts, rules, and principles” that have “a clear goal and an algorithmic solution path” (Nievelein et al., 2013, p. 199). In contrast, ill-defined problems—the types of problems often considered by the humanities—do not have such definitive solution steps upon which to rely. In disciplines such as literature, writing, and the arts, the problem is often less clear-cut, the path to solution less linear, the rules less established, and the end result less definitive (Nievelein et al., 2013). Given this critical distinction, researchers have rightly asked whether worked examples, which identify and model discrete solution steps, are still effective in ill-defined domains.

In a review of the handful of studies on worked examples of ill-defined problems, Renkl, Hilbert, and Schworm (2009) argued that the preliminary findings in this new field of research suggest that learning from complex worked examples in “heuristic domains” is effective (p. 68). Contemporary studies of worked examples for ill-defined problems have ranged from teaching argumentation (Schworm & Renkl, 2007), to identifying designers’ styles (Rourke & Sweller, 2009), to writing effective learning journals (Hübner et al., 2010), to reasoning about legal cases (Nievelein et al., 2013), to composing essays on English literature (Kyun et al., 2013). Each of these studies found a significant main effect for worked examples, highlighting how worked examples may be effective not only in concrete disciplines such as math and science, but also in more abstract disciplines such as design, literature, and the law.

It is possible that the reason worked examples are effective in a broad range of disciplines is that “there are no grounds for assuming that problem-solving skills acquired in well-defined domains through studying worked examples differ substantially in any way from those acquired in ill-defined domains” (Rourke & Sweller, 2009, p. 186). In truth, worked examples may be even more important and more effective for reducing extraneous load in students studying

problems in ill-defined domains, because these problems are typically more complex and require more prior knowledge to solve (Renkl et al., 2009; van Merriënboer & Sweller, 2005). Students stand to benefit even more from the load-reducing effects of worked examples in ill-defined domains.

Studies of Worked Examples in Ill-Defined Domains

This review of the literature will highlight three crucial studies of multimedia worked examples in ill-defined domains that most closely align with the objectives and setting of the present study. Though other studies exist in this line of research, these three studies—of legal reasoning, journal writing, and argumentation—tackle problems most closely related to the cognitive processing skills required in analytical reading. They also are most closely aligned with the present study, since they use multimedia tools to present the worked examples.

The first experiment (Nivelstein et al., 2013) considers worked examples of legal reasoning with a focus on better understanding the expertise reversal effect in ill-defined domains. The second experiment (Hübner et al., 2010) examines worked examples of writing learning journals with a focus on better understanding the conditions that lead to transfer of learning from the training session to similar tasks. The third experiment (Schworm & Renkl, 2007) tests worked examples of argumentation skills with a focus on ways to foster self-explanation. Each of these studies work toward answering many of the questions posed by researchers of worked examples in well-defined domains, and together they present a model for the methodology of the present study.

Study one. In a study by Nivelstein et al. (2013), the researchers examined the strength of the worked example effect on the ill-defined problem of legal reasoning and challenged the influence of the expertise reversal effect in this domain. The Nivelstein et al. (2013) study,

grounded in cognitive load theory, seeks to do two things: (a) expand the definitions set forth by Sweller (1988) for what makes worked examples effective in reducing extraneous load, and (b) better understand what sparks the expertise reversal effect. The researchers noted that previous studies found an expertise reversal effect in learners with high prior knowledge, and as a result, these studies suggested that worked examples are most useful for novices. However, some studies hinted at the possibility that the expertise reversal effect may only affect those studying well-structured tasks (Schmidt, Loyens, van Gog, & Paas, 2007). Therefore, the purpose of the Nievelstein et al. (2013) study was to determine whether novice and advanced law students' legal reasoning would improve when given worked examples to study in place of a more independent, systematic learning activity, such as step-by-step instructions.

This experimental study used a 2x2x2 factorial design, with resulting in four possible test conditions for each of two possible expertise levels. The factors were worked examples (yes vs. no), process steps (yes vs. no) and student expertise (novice vs. advanced). The sample included 75 novice law students and 36 advanced law students, all of whom volunteered to participate. Students were randomly assigned to one of four categories, in each of which worked examples and process steps were the treatment conditions and independent problem solving was the comparison condition. The four categories included worked example with process steps, worked example without process steps, independent problem solving with process steps, and independent problem solving without process steps. To ensure equality of variance between groups, all participants were assessed using a baseline prior knowledge test that measured conceptual knowledge of property law.

During the intervention phase, students either studied or attempted to solve a practice civil law case dealing with property transfer. Students in the problem solving condition without

process steps had to solve the case by relying upon prior learning in the course and the textbook of civil codes. Students in the process steps condition received a five-step guide outlining the general protocol for property law cases, which they were permitted to use in attempting the practice case. Students in the worked example condition received a multimedia worked example detailing an expert's legal reasoning process of the practice case. Students in the worked example plus process step condition received a multimedia worked example that detailed an expert's legal reasoning process of the practice case and named each of the five steps in the process guide.

Upon completion of the intervention phase, all students wrote their own legal reasoning for a similar civil property transfer case. After students completed the test task, they responded to a mental effort rating scale (Paas, 1992) in which they self-reported their mental effort on a scale of one to nine. The free responses from the test task were scored using a model created by a private law professor. The data from each expertise level group were separately analyzed using a 2x2 Analysis of Variance (ANOVA) to examine the differences between the treatment and comparison condition at each level. In order to contrast the data across level groups, a nonparametric Mann-Whitney U test was used.

The ANOVA of the novice students yielded a significant worked example effect: novice students in the worked example condition performed better on the test task than novice students in the independent problem solving condition. Effect sizes were reported using Cohen's f (Cohen, 1988), which ranges in strength from small ($f < 0.10$) to medium ($0.10 > f < 0.25$) to large ($f > 0.40$). Moreover, novice students in the worked example condition were able to spend less time learning the material ($f = .75$) and exert less mental effort ($f = .42$) than did novice students in the independent problem solving condition. The analysis also revealed the effect of

process steps for novice students: there was no significant effect on the test task, but novice students with process steps spent more time on the learning task ($f = 0.37$). Additionally, there was a significant effect on mental energy invested, as novice students with process steps invested more mental energy in the learning phase ($f = .24$).

Similarly and more surprisingly, the ANOVA of the advanced students also yielded a significant worked example effect ($f = 1.02$), suggesting that advanced students in the worked example condition performed better on the test task than advanced students in the independent problem solving condition, a finding that stands in stark contrast to previous research on the expertise reversal effect. The analysis of time spent on learning task revealed that advanced students in the worked example condition were more efficient in learning the material than students in the independent problem solving condition. There was no significant effect of worked examples on mental effort invested in the advanced student group. There was also no significant effect for the process steps condition in the advanced students, across test task, time on task, and mental effort exerted. Interestingly, a nonparametric Mann-Whitney U test revealed no significant difference between the advanced and novice students when it came to time spent learning.

The data suggest a significant worked example effect for both novices and more advanced students solving an ill-defined problem. Contrary to the outcome of studying worked examples of well-defined problems, the expertise reversal effect may not hinder students with high prior knowledge studying worked examples of ill-defined problems. The Nievelein et al. (2013) study plays an important role in the research literature in the field of worked examples. By drawing upon previous studies around both well-structured and ill-structured tasks and modeling the current study after previous designs, the authors were able to extend the research

knowledge base around the expertise reversal effect and to confirm previous findings about the expanded scope of worked examples, to include ill-structured tasks. They clarified the way in which the expertise reversal effect inhibits students studying well-structured tasks but may not be detrimental to those studying ill-structured tasks.

Study two. In an earlier study by Hübner, Nückles, and Renkl (2010), the researchers investigated the effect of multimedia worked examples with prompting for self-explanation on students' use of specific cognitive and metacognitive strategies in writing learning journals. Although they had conducted previous studies around strategy instruction in university-level courses and found that students were able to apply instructional prompts to their own work (Nückles, Hübner, & Renkl, 2009), they conversely found that younger, high school-aged students were unable to transfer instructions to practice (Glogger, Holzapfel, Schwonke, Nückles, & Renkl, 2009). Thus Hübner et al. (2010) conducted this study in order to better understand the conditions that lead to transfer of cognitive and metacognitive writing strategies for high schoolers writing learning journals, a critical practice for writing-to-learn pedagogy.

The theoretical framework for the study builds on previous work around writing-to-learn and self-regulated learning, whereas the methodological framework for the study draws upon research into worked examples with self-explanation (Chi et al., 1989). The experimental study used a 2x2 factorial design, resulting in four possible conditions around two independent variables: worked examples and informed prompting. The sample included 70 students from a German secondary school who were randomly assigned to one of the following conditions: informed prompting ($n = 18$), learning journal multimedia worked example ($n = 17$), both informed prompting and learning journal multimedia worked example ($n = 18$), or no instructions in addition to the assignment topic ($n = 17$). The average age of students in the study

was 17.62 years (equivalent to a second-semester eleventh-grader or first-semester twelfth-grader in American high schools). To ensure the experimental groups were comparable, all participants were assessed using a topic-specific prior knowledge pretest on the topics presented in the lecture. Following the pretest, students watched a videotaped lecture on a topic of social psychology, during which they were not permitted to take notes.

During the intervention phase, students in the informed prompting condition attended a standardized presentation on the utility and value of using cognitive and metacognitive strategies, such as organization of information, elaboration on content, monitoring of learning, and planning of remediation strategies. Students in the learning journal multimedia worked example condition watched and listened to a recorded PowerPoint presentation that narrated an exemplar learning journal. At several places in the presentation, students were prompted to self-explain. More specifically, the presentation was self-paced; students pushed a button to start the presentation and paused the presentation when they were asked to explain which cognitive or metacognitive strategy the exemplar journal was using. Students assigned the passage from the journal entry to the corresponding prompt from the list and received feedback as to whether or not their selection was correct. In the third experimental condition, students received both informed prompting and the worked example. In the fourth condition, students received neither intervention. After completing the intervention phase, students responded to an open-ended posttest that measured comprehension of the topic from the lecture. Additionally, learning journals were assessed for frequency and quality of application of cognitive and metacognitive strategies.

The data were analyzed using a 2x2 ANOVA to contrast the differences among the four conditions for eight separate subtests: organization cognitive strategies, elaboration cognitive

strategies, metacognitive strategies, and learning outcomes for both the intervention phase and the transfer phase. In four of eight subtests, the researchers found a statistically significant strong worked example effect, all of which were reported using the effect size partial eta-squared (Cohen, 1988), which ranges in strength from weak ($\eta_p^2 < 0.06$) to medium ($0.06 > \eta_p^2 < 0.13$) to strong ($\eta_p^2 > 0.13$). Students in the worked example condition performed significantly better than their peers in any of the other three conditions with respect to elaboration strategies in the intervention phase ($\eta_p^2 = 0.17$), elaboration strategies in the transfer phase ($\eta_p^2 = 0.15$), metacognitive strategies in the intervention phase ($\eta_p^2 = 0.18$), and comprehension of the learning material in the transfer phase ($\eta_p^2 = 0.15$).

The data highlight two critical insights for worked examples research. First, the strong main effects for worked examples are confirmation of the utility of worked examples for ill-defined problems, such as writing to learn. Much like the worked examples presented in the Nievelstein et al. (2013) study, Hübner et al. (2010) successfully deconstructed the ambiguous, ill-structured tasks of writing for a learning journal by providing a heuristic technique for using cognitive and metacognitive strategies that students could emulate while writing. Second, the data suggest that when it comes to optimizing conditions for students to effectively apply their learning to a task, worked examples with built-in opportunities for self-explanation is superior to informed prompting—and even to worked examples with informed prompting. Contrary to the hypotheses of the researchers, “informed prompting did not enhance the amount of cognitive and metacognitive strategies, neither in the immediate training [intervention] session, nor in the transfer session seven days later” (Hübner et al., 2010, p. 26). These additional steps appear not to be necessary to increase student learning outcomes for high school students studying in an ill-defined domain.

Study three. Given the critical importance of self-explanation in worked examples, the third study (Schworm & Renkl, 2007) presented in this review of the literature focuses on how best to foster self-explanation while studying a worked example of an ill-defined problem. In the case of the Schworm and Renkl (2007) study, that ill-defined problem is rhetoric—or skills for argumentation. The researchers note that in previous studies of worked examples with well-defined problems, students needed to understand only one domain—the learning domain. However, with ill-defined problems, students must understand a second and sometimes a third domain in order to access the learning domain. In the case of argumentation, as in the Schworm and Renkl (2007) study, the learning domain was argumentative structure based on a theoretical model that students needed to understand. Additionally, students needed to understand a second domain, called the exemplifying domain, in which they contend with how that model was presented through an example that applied the theoretical model. Whereas traditional, algorithmic worked examples rely on only one domain (single-content examples), worked examples of ill-defined problems rely on at least two domains (double-content examples). Thus, it was the purpose of this study to determine whether double-content worked examples of ill-defined problems would still yield a significant main effect on student learning, and if so, when prompting for self-explanation, whether those prompts should elicit further thinking about the learning domain, the content domain, or both.

This experimental study used a one-factor design with four instructional conditions: worked example with no self-explanation prompts; worked example with eight self-explanation prompts in the exemplifying domain; worked example with eight self-explanation prompts in the learning domain; and worked example with eight self-explanation prompts, four of which were in the exemplifying domain and four in the learning domain. The sample included 72 student

teachers (mean age = 23.4 years) who were randomly assigned to one of the four experimental conditions ($n = 18$). Participants completed several pretests—an assessment of typing skills, an assessment of prior content knowledge of the explanatory domains, and two brief assessments on declarative knowledge of the learning domain, argumentation—leaving the researchers to conclude that the differences between the experimental conditions were minimal.

The intervention phase took place in a computer-based environment. Students in all four conditions received a multimedia worked example that was presented via video. Each video began with a brief, six-minute dialogue about stem cell research and achievement differences by gender in science and mathematics (the explanatory domain) that applied the principles of argumentation being studied (the learning domain). First the dialogue was played in full from start to finish, and then it was replayed in four shorter segments that aligned with the four components of the argumentative theory being learned. At these segment breaks, students in the self-explanation conditions were prompted to respond to various questions by typing inside a box on the screen. Questions ranged from learning domain prompts, to exemplifying domain prompts, to a combination of the two. Students could continue the worked example only after responding to the prompt. Upon completion of the intervention phase, all students took a series of three posttests: assessments of recall from the video lesson, declarative knowledge about argumentation, and applied argumentation skills. In addition to these measurements, the researchers also coded and counted the quality and quantity of written self-explanations.

The data were analyzed through a series of 4x2 and 2x2 mixed factor ANOVAs in which instructional condition was the between-participants factor and time was the within-participant factor. Partial eta-squared (η^2) was reported for the effect size of each analysis. Among the many findings in the Schworm and Renkl (2007) study, the researchers reported four findings in

particular which are material to the present study. First, they found a statistically significant strong main effect for worked examples across all four instructional conditions ($\eta p^2 = .60$), suggesting that multimedia double-content worked examples can be powerful tools for fostering student learning in ill-defined domains. Second, prompting for self-explanation of the learning domain (in this case, argumentation) increases the number ($\eta p^2 = .26$) and quality ($\eta p^2 = .46$) of student self-explanations related to the contents of argumentation. Both of these are interpreted as strong effect sizes. Third, prompting for self-explanation of the exemplifying domain (in this case, stem cell research and gender differences in math and science achievement) increases the quality ($\eta p^2 = .22$) of student self-explanations related to the exemplifying domain, but does not significantly change the quantity of such explanations. Fourth, there was no significant interaction effect between groups with specific learning or exemplifying-domain prompts and groups with both sets of prompts, suggesting that it makes no difference whether students are presented solely with learning-domain prompts or their learning-domain prompts are combined with exemplifying-domain prompts (Schworm & Renkl, 2007).

Though the findings from this study extend the scope of worked examples with prompts for self-explanation beyond well-structured tasks, the researchers note that their results reflect only double-content examples in an ill-defined domain. In a call for future research, to which the present study responds, Schworm & Renkl (2007) assert, “A challenging task for further research would be to investigate how the approach of learning with self-explaining [worked] examples can be expanded to examples that contain cognitive modeling. Such examples are often used in teaching cognitive strategies” (p. 294). The present study of multimedia think-aloud worked examples with prompts for self-explanation is an instance of the under-researched triple-content worked example (Renkl et al., 2009), in which students must contend with three domains: the

learning domain (analytical reading), the exemplifying domain (the text being read), and the cognitive strategy domain (heuristic strategies and schema for approaching analytical reading).

Taken together, these three studies highlight the promise of the present study, which (a) applies the design principles of worked examples with self-explanation to a new and untested ill-defined domain, and (b) expands on the body of research around worked examples by considering a complex discipline with three levels of content for students to process in order to reach a deep level of understanding. These studies suggest that worked examples of unstructured, ambiguous tasks should be effective in reducing the extraneous cognitive load that burdens students in assimilating schema, while increasing students' comprehension of challenging material.

Design Principles for Multimedia Worked Examples

Beyond extending the conversation around type of task—ill-structured or well-structured—some researchers have explored the format of delivery and, in particular, the effectiveness of computer-based learning in delivering worked examples (Crippen & Earl, 2007; Gerjets, Scheiter, & Schuh, 2008; Jung & Kim, 2006; Kim, 2005; Kyun & Lee, 2009; Renkl, 2014). These researchers are among a growing group of people who are providing evidence to support the transfer of worked examples from the classic paper-and-pen model to a computer-based learning environment that is in tune with the rapid changes in education. In making this transition, however, it is imperative to consider the ensuing challenges of reducing extraneous cognitive load in a learning environment that is potentially even more distracting. Renkl et al. (2009) caution that “worked examples that do not reduce extraneous cognitive load ... do not foster learning” (p. 68). Therefore, what follows is a synthesis of salient design principles for

worked examples, from the first generation to the present, that take into account the principles of multimedia learning (Mayer, 2014a).

Synthesis of Worked Examples and Multimedia Principles

Studies of learning from worked examples and learning with multimedia have produced an abundance of design principles for constructing effective instructional materials (Atkinson et al., 2000; Mayer, 2014b; Mayer & Moreno, 2010; Renkl, 2014; van Merriënboer & Kester, 2014). Though many of the recommendations are most useful in designing for algorithmic tasks in well-structured domains, a selection of these principles is particularly important for effectively designing multimedia worked examples of ill-defined problems.

Self-explanation principle. Self-explanation is a process in which learners actively engage in studying the worked example by commenting on its structure or content, anticipating the next step and cross-checking against the example, explaining the principles and goals of an example and elaborating on those principles, and monitoring understanding or comprehension failures (Chi et al., 1989; Renkl, 1997). Self-explanation is a successful strategy because it increases germane cognitive load.

Multiple examples principle. Students learn best from worked examples when they are presented with multiple worked examples to study (Atkinson et al., 2000; Sweller, 1994). Findings from a study by Reed and Bolstad (1991) confirmed that two examples are better than one, and that students with a simple and a complex example outperformed their counterparts who had only one example or one example plus written procedure steps.

Example-problem pairs principle. In addition to providing students with multiple opportunities to study similar worked examples, Atkinson, Derry, Renkl, and Wortham (2000) assert that students learn best when examples are paired with alternating opportunities for

practice. In the case of well-defined worked examples, expert solution should be matched with a similar problem for students to solve.

Multiple modalities principle. In keeping with one of the core tenets of multimedia learning—that students learn better from a combination of words and pictures, rather than from words alone (Mayer, 2014a)—studies of multimedia worked examples suggest that students learn best from worked examples when expert cases are presented using “simultaneous multiple modalities” to include “aural explanation overlaid on video” (Atkinson et al., 2000, p. 207). Students given mixed-mode instructional media outperform their peers in single-mode experimental conditions (Mousavi, Low, & Sweller, 1995).

Coherence principle. Instructional materials are most effective when they include only content that is relevant to and essential for student learning (Mayer & Moreno, 2010). Though it may be tempting to “spice up [a] lesson” using multimedia tools (Mayer & Fiorella, 2014, p. 284), the better lesson is the more straightforward one. Interesting but unnecessary material is a distraction; omitting superfluous content reduces extraneous load.

Signaling principle. Students learn best when they have cues that direct their attention to the most important parts of a lesson (Mayer & Moreno, 2010). The signaling principle stipulates that instructors should include vocal emphasis, pointer words, headings, and initial outlines to help students identify and retain the essential material, which thus reduces extraneous cognitive load (Mayer & Fiorella, 2014). Similarly, Atkinson et al. (2000) recommend chunking worked examples into smaller steps, with labeled subgoals and visual separation between steps, to help students comprehend complex content. The signaling principle is also referred to as the “meaningful building blocks” principle in some research of multimedia worked examples (Renkl, 2014).

Split-attention principle. Poorly designed instructional materials sometimes require the learner to integrate different sources of information, such as diagrams and text, on their own, which causes an unnecessary increase in extraneous load. Instead, the split-attention principle states that instructors should present corresponding elements (narration, text, images) close to one another, both visually and temporally, to reduce extraneous cognitive load (Mayer & Fiorella, 2014). The split-attention principle is often divided into the spatial contiguity and temporal contiguity principles (Mayer & Moreno, 2010).

Segmenting principle. Similar to the signaling principle, students learn best when they are in control of the pace with which they progress through a multimedia presentation, such as a worked example. The segmenting principle suggests breaking the multimedia presentation into manageable segments to allow students to progress and interact with media at their own pace. This principle is in keeping with van Merriënboer and Kester's (2014) research on the self-pacing principle, which also intimates the importance of self-pacing for facilitating students' elaboration on and deep processing of information. Segmenting is theorized to help students manage intrinsic load.

Modality principle. Students learn best from multimedia instructional materials, such as worked examples, when the words are narrated rather than written (Mayer, 2014a). Presenting words in spoken form may help to manage the intrinsic load of the learning activity "by distributing the cognitive processing across both information processing channels" (Mayer & Moreno, 2010, p. 147). This principle is particularly important when students are studying complex and unfamiliar content.

Personalization and voice principle. Multimedia instructional materials are most effective when the spoken words are presented in an approachable, conversational style, rather

than a formal style (Mayer, 2014b). Moreover, students prefer a human voice rather than a computerized voice. Personalization and voice are theorized to promote “active processing of new information,” because students are more likely to “relate the material to themselves, thus creating deeper memories of the learning experience” (Moreno & Mayer, 2010, p. 160). As a result, personalization and voice increase germane load.

Taken together, these principles define the parameters for effective multimedia worked examples that have the potential to reduce extraneous cognitive load, manage intrinsic cognitive load, and support the learner in fostering germane cognitive load to build schema and transfer new learning into long-term memory.

Summary of the Literature

Learning by example is nothing new to the education landscape. Research into think-aloud protocols, though often used as a form of assessment rather than instruction, provides practical, content-specific literacy strategies for crafting the instructional intervention in this study. Additionally, research into worked examples—from the earliest pen-and-paper studies of algebra and statistics, to more recent multimedia studies of legal reasoning and writing—shapes the conceptual framework for the present study by detailing a series of design principles for effective multimedia worked examples. The combined literature around instruction that leverages *multimedia*, *think-alouds*, and *worked examples* defines a clear path for this study. An effective multimedia think-aloud worked example will (a) follow the design principles of multimedia learning, to include the coherence principle, the signaling principle, the split-attention principle, the segmenting principle, the modality principle, and the personalization and voice principle; (b) apply the known best practices in performing think-alouds, to include naming a strategy or skill and describing its purpose, reading a selection of text and pausing to verbally

apply the strategy or skill, and reflecting after the think-aloud on the usefulness of the strategy or skill; and (c) follow the design principles of effective ill-defined worked examples, to include the self-explanation principle, the multiple examples principle, the example-problem pairs principle, and the multiple modalities principle.

CHAPTER III

METHODOLOGY

This mixed-methods study tested the worked example effect on the ill-defined problem of adolescent analytical reading comprehension, drawing upon practitioner-based literacy research into teacher think-alouds as a guide. This study extends previous worked examples research by (a) situating the quasi-experiment in the naturalistic setting of a high school English language arts classroom and (b) applying the model of worked examples to the ill-defined task of analytical reading. This chapter reviews the research questions; articulates the research design; describes the setting, sample, and procedures for protecting human subjects; and discusses the instrumentation, treatment, procedures, pilot, and data analyses. The chapter closes with an investigation of the limitations of the study design.

Research Questions

This study considered three research questions. The first two questions are quantitative in nature. Given the small sample size ($n = 34$) of the study, the final qualitative question is designed to help explain any differences that are not detected by the quantitative measures.

1. Analytical Reading Comprehension. To what extent is there an effect of multimedia think-aloud worked examples on analytical reading comprehension in a secondary English language arts class, as measured by open-ended written responses to literature?
2. Mental Effort. To what extent is there an effect of multimedia think-aloud worked examples on cognitive load, as measured by self-reported mental effort (Paas, 1992)?
3. Student Experience. What do student annotations and interviews reveal about the participant experience of studying multimedia think-aloud worked examples?

Research Design

This quasi-experimental study draws upon an explanatory sequential mixed-methodology design (Creswell, 2015). The first (quantitative) strand of the design was a classic treatment-comparison repeat measures pretest-posttest study and used a matched pairs randomized block design to ensure a relatively similar sample around reading comprehension between the treatment and comparison groups. The independent variable for this study was the instructional intervention, which had two levels: (a) multimedia think-aloud worked examples and (b) traditional text-based questions, equivalent to the “traditional problem-solving” comparison condition in worked examples research around well-defined problems (Atkinson et al., 2000; Renkl & Atkinson, 2010; Sweller, 2006). The two dependent variables for this study were mental effort and reading comprehension.

The second (qualitative) strand of the design engaged students in a series of open-ended interviews and written annotations to help explain the differences that were detected by the quantitative measures in the first strand and identify differences that may not have been detected by the quantitative instruments given the small sample size. The study lasted three weeks during a six-week unit on *The Great Gatsby* (Fitzgerald, 1925), beginning at the start of the second semester in January 2017. For the purpose of mitigating researcher bias, the teacher-researcher partnered with a co-researcher who is an expert in scientific research design. The co-researcher holds an EdD in instructional technology, is the director of technology for the school at which the study took place, and has taught graduate-level courses in research methodology. A biology teacher by trade, she is well versed in the scientific method. For a schematic overview of the research design, see Figure 5 on the following page.

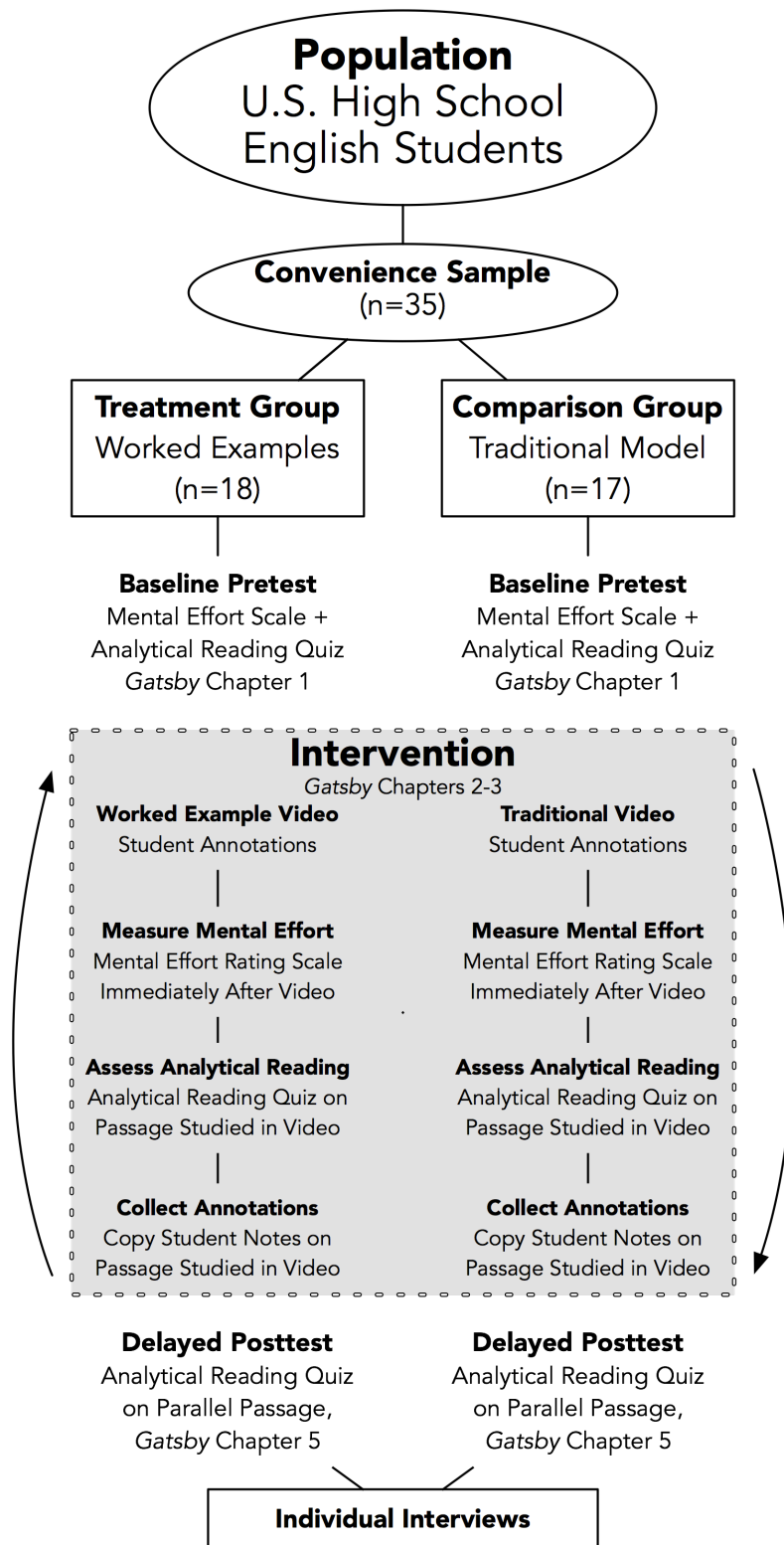


Figure 5. Research design schematic overview

Sample

The sample came from an independent, college-preparatory high school in California's Silicon Valley, a hub of innovation in educational technology. The school has been recognized as an Apple Distinguished Program for their one-to-one iPad program, making it particularly well suited for a study of an instructional intervention that leverages multimedia tools. The total enrollment of the school is approximately 600 students, with 28% of students identifying as students of color and the majority identifying as white. The student body is composed of an even number of males and females. This particular high school is one of the few independent schools in the area that offer a special education program, heightening the school's need for differentiated instruction and suggesting greater generalizability of the sample to students outside of the school. Approximately one quarter of students enrolled receive accommodations for documented learning differences.

The sample for the study was a convenience sample that included 35 students from two heterogeneous 10th-grade United States literature classes taught by the primary researcher. The choice of 10th-grade classes was based on the single-track nature of the English program at that grade level and the need for teachers to support a wide range of reading skill. (No honors, AP, or remediation courses are offered in English at the 10th-grade level.) The sample made up 20% of the sophomore class, suggesting that it should reflect the academic and demographic diversity of the school. Moreover, the sample represented a wide range of analytical reading skill. Standardized national percentile scores on the Scholastic High School Placement Test (HSPT 2C, 2014) in analytical reading for this group span from the 23rd percentile all the way up to the 97th percentile, illustrating the wide range of need for instructional support. The treatment and comparison groups were composed of a similar range of prior reading skill scores.

During the interview process, it became apparent that the data for one student needed to be dropped from the study, as the student fell asleep during the second phase of the instructional intervention. Thus the total sample size for the study was 34 students, with 18 students participating in the treatment group, and 16 students participating in the comparison condition.

Protection of Human Subjects

The study was approved as exempt research by the University of San Francisco Institutional Review Board for the Protection of Human Subjects, as it involved minimal risk to subjects according to 45CFR46.101(b). There were no known or anticipated risks to participants in the dissertation study. By participating in the study, students had the opportunity to learn new reading comprehension strategies and better understand their own analytical reading process. In conducting this study, the researcher and co-researcher observed all ethical standards and policies of the University of San Francisco Institutional Review Board for the Protection of Human Subjects and all human research protection regulations of the American Psychological Association (2010). The study was vetted and endorsed by the school principal, also an educational researcher who holds a Ph.D. in education. A letter of support from the principal is included in Appendix A.

The following precautions were taken to protect the rights of students participating in the study:

1. The researcher explained the purpose of the study to her students and gave them multiple opportunities to ask questions and understand their role in the research.
2. Signed parental consent forms were collected from every participant in the study. The consent form included an introductory letter stating the purpose of the study and describing the procedures for the intervention and data collection (Appendix B). Every student enrolled

in the teacher-researcher's US Literature courses was given signed permission by his or her parent or guardian.

3. To protect each student's confidentiality, names and identifying factors did not appear with test results. The co-researcher had access to student names and test scores for the sole purpose of coding the data, which ensured anonymity during the study. Information obtained during this study was shared only with the researcher, co-researcher, and University of San Francisco faculty sponsor.
4. All audio obtained during the qualitative phase of the study was transcribed by a professional transcriptionist and coded so that no personally identifying information was available from them. During the interviews, students were identified by research number, not by name. Students were asked for assent to record and transcribe their interviews. The recordings were kept in a secure place and were listened to only for research purposes by the researcher. Recordings were retained for the duration of the dissertation process and will be erased immediately after the dissertation defense is completed.
5. At the end of the research study and before the culminating assessment for the unit, all students were granted access to each of the videos presented during the study.

Treatment Description

The independent variable for this study was the instructional intervention, which had two levels: (a) multimedia think-aloud worked examples and (b) traditional text-based questions. Students participated in either the treatment or the comparison condition for the instructional intervention, which was based on the second and third chapters of *The Great Gatsby* (Fitzgerald, 1925).

A natural learning curve occurs in reading any novel. The highest natural cognitive load happens in the early chapters as students are gaining familiarity with the characters, language, text structure, and basic plot elements. In order to test student comprehension and mental effort after the intervention in a chapter of relatively high cognitive load, the teacher-researcher, under the advice of the content and construct validity panelists, selected the second and third chapters of *The Great Gatsby* as the content for the intervention. The baseline assessment—testing for homogeneity of variance and group equivalence—took place at the very end of the first chapter, and the delayed posttest took place after Chapter 5. *The Great Gatsby* (Fitzgerald, 1925) has nine chapters.

The instructional intervention began with students reading the assigned chapter independently in class to assure that all students read without the assistance of outside resources. During the subsequent class period, students logged in to Schoology, the course learning management system, to access the materials for the intervention. United States Literature is a yearlong course, which means that by the time of the study, students were very familiar with accessing course content via Schoology. In the first phase of the instructional intervention, students viewed one of two videos for Chapter 2 of *The Great Gatsby*—either the worked example or the traditional text-based questions—depending on their randomly assigned group. In the second phase of the instructional intervention, students viewed the video for Chapter 3 that was from the same condition (worked example or traditional) as the first video they watched.

A total of four instructional videos were created by the teacher-researcher: a worked example video and a text-based questions video for each chapter studied during the intervention (Chapters 2 and 3). Each video began by setting an analytical purpose for re-reading the passage, a practice that is consistent with research on literacy instruction (Beers, 2003; Frey & Fisher,

2013). Then the videos narrated the passage from start to finish without interruptions. The worked example video and the text-based questions video focused on the identical passage within each chapter. After the reading of the passage, the videos diverged; the worked example video presented a think-aloud deconstruction of the passage, whereas the text-based questions video presented a series of close reading questions that led students to the same content and depth of analysis as the think-aloud. Videos were compared for equivalence and validated by an outside expert—an English teacher with 12 years of middle and high school teaching experience, a master’s degree in literature, and expertise in supporting struggling readers.

The worked example videos drew upon a think-aloud framework for processing through the passage (Frey & Fisher, 2013). The teacher performing the think-aloud read the passage aloud, pausing to verbally apply the analytical strategy or skill named at the start of the video, and reflecting after the think-aloud on the usefulness of the strategy or skill. In keeping with the interactive nature of think-alouds, in which the learner is guided to reflect on the example provided by the teacher, the treatment condition videos followed the model of worked examples with self-explanation (Chi, 2000; Chi et al., 1989; Renkl et al., 2009; Schworm & Renkl, 2007). Students in the treatment condition were prompted periodically to comment on the structure and content of the think-aloud, explaining the method the teacher used to break down the passage that was previewed in setting the analytical purpose for reading. A selection of screenshots from the worked example videos are rendered in Appendix C. The Chapter 2 multimedia think-aloud worked example video is available online at <http://tinyurl.com/neebe-worked-example>. In contrast, the traditional videos presented students with a series of text-based questions, the literature equivalent of independent problem-solving questions, which is the standard practice for a comparison condition in worked example research (Atkinson et al., 2000; Renkl & Atkinson,

2010; Sweller, 2006). Students were prompted to pause the video after each question to annotate the text before moving to the next question. A selection of screenshots from the traditional text-based questions videos are rendered in Appendix D, and the video for Chapter 2 is available online at <http://tinyurl.com/neeбетraditional>.

Students in both conditions watched the videos while making notes on a passage annotation handout containing the passage, surrounded by ample margin space for note-taking. Student annotation handouts with selected passages are included in Appendix E. Students were permitted to pause and rewind the videos as needed to annotate. By this point in the school year, students were very familiar with the process of annotating a text to preserve the analytical thinking in which they engaged during reading.

Both the treatment and the comparison conditions were presented with videos of approximately fifteen minutes in length that used identical color schemes and fonts. All videos were accessed online through Schoology. To ensure independence of the treatment and comparison conditions, videos were available only to students on Schoology and only for the duration of the class period. They were made unavailable the moment students walked out the door. Additionally, all students wore headphones to further limit participant bias. Headphones were provided for students who forgot to bring their own. The physical setting for the study was the teacher-researcher's classroom, which was a large space for the 17 to 18 students in the room at any given time. Students worked at modular tables, which had dividers between students for the duration of the intervention so that students could not see one another's screens.

Each video adhered to applicable design principles for multimedia learning (Mayer, 2009, 2014a). Videos in both the treatment and comparison conditions employed clear and colloquial "aural explanation" (Atkinson et al., 2000, p. 207) and only included content and

multimedia features that were essential for student learning. Videos included cues, such as vocal emphasis, pointer words, headings, and initial outlines, to direct student attention to the most important parts of the lesson. Moreover, students had control over the pacing and progression of the videos, which were broken down into manageable segments.

Instrumentation

This study used two instruments in the quantitative strand to measure each of the dependent variables during the intervention. Analytical reading comprehension was assessed using an open-ended written analytical reading quiz, which was scored on a 15-point rubric. Mental effort was measured using the Mental Effort Rating Scale for Adolescent Analytical Reading, a seven-point scale. All instruments were deployed through Schoology, which students accessed via their school-issued iPads or personal laptops.

The study then engaged students in three types of open-ended measures in the qualitative strand to describe the experience of interacting with the treatment and also to explain any quantitative differences in achievement and mental effort between treatment and comparison or between advanced and struggling readers. Table 4 provides an overview of the instruments that were used in the study in sequential order.

Table 4

Instruments

Phases of Study	Instruments
Baseline	Mental Effort + Analytical Reading Assessment, Chapter 1
Intervention, Phase I	Mental Effort + Analytical Reading Assessment, Chapter 2
Intervention, Phase II	Mental Effort + Analytical Reading Assessment, Chapter 3
Delayed Posttest	Mental Effort + Analytical Reading Assessment, Chapter 5
Qualitative/Explanatory	Interviews + Focus Groups

Analytical Reading Comprehension

After each phase of the study, students submitted a brief, written response to a select passage from *The Great Gatsby*. In keeping with the standard English language arts practice of assessing analytical reading comprehension through written responses to literature, the quiz at the end of each chapter asked students to connect the author's use of language and literary conventions to broader implications. After reading a passage that describes the primary setting for one of the central characters of the novel, students were asked to write a paragraph responding to the following prompt: "What are the specific details that you notice about the setting in this passage? Based on these details, what inferences can you draw about the character(s) associated with that setting? Your response will have two parts: what I see, and what I infer. Start the first part of your response with, *In this passage, I see ...* Start the second part of your response with, *Therefore, I infer that ...* Try to include as many observations and inferences as you can, and make sure that your observations and inferences are clearly connected."

The four passages from the novel (Appendix E) were selected by the teacher-researcher, who has a degree in English, an advanced degree in curriculum design with an emphasis on literacy instruction, and 10 years of experience teaching *The Great Gatsby* within the context of a high school United States literature course. Passages and analytical reading quiz prompts were validated by a panel of two additional content experts, both teachers of United States literature, one of whom holds an advanced degree in literacy instruction and the other of whom holds a PhD in American Literature. In particular, experts identified the passages from the selected chapters that were richest in language, most similar in their indirect development of character through setting, and would be most conducive to close reading and literary analysis.

Responses were assessed using the analytical reading response rubric, which is composed of three sub-scales: textual observations, inferences, and broader implications (Appendix F). This rubric is an abridged version of a more comprehensive rubric that the teacher-researcher developed in collaboration with four English language arts teachers from two high-performing California high schools, one public and one independent. Three of the collaborating teachers have taught Advanced Placement English, and two have served as head of their schools' English departments. All four collaborating teachers hold degrees in English, and three of the four hold advanced degrees in curriculum design with an emphasis on literacy instruction.

The comprehensive rubric has been used for more than ten years, by at least six English language arts teachers, to teach more than 1,000 high school English students in Grades 9 through 12, which points to its content validity. To ensure the accuracy of the researcher's ratings of student responses using the rubric, one of the collaborators of the rubric served as a second reader, assessing a random selection of student responses and all responses that yielded outlier scores, and comparing scores with the teacher-researcher.

Mental Effort

Direct response from 10th-grade literature students on the Mental Effort Rating Scale for Adolescent Analytical Reading (Appendix G), a self-report survey, provided data to measure students' perception of mental effort after they completed the analytical reading task. Because the survey was taken immediately after the task (van Gog, Kirschner, Kester, & Paas, 2012), it has the potential to reveal whether or not the instructional intervention of multimedia think-aloud worked examples was successful in reducing excess strain on the working memory. In other words, the comparative data expose whether or not the scaffolds used to support reading comprehension were doing their job. The primary means of collecting this data was an electronic

survey delivered through Schoology. The instrument was self-administered upon completion of the analytical reading written response, after which the data was immediately and electronically analyzed.

The Mental Effort Rating Scale for Adolescent Analytical Reading was created in response to the teacher-researcher's concerns with the accuracy of the existing single-item scale in the context of the present study. Two popular versions of this extant scale are in circulation currently: (a) the Paas single-item, 9-point, ordinal Mental Effort Scale (Paas, 1992), which assesses perceived mental effort on a line-graph continuum from "very, very low mental effort" to "very, very high mental effort"; and (b) the Yeo and Neal (2004, 2008) adaptation of Paas's scale, which asks participants to rate how "hard" they were trying on an 11-point ordinal scale. Because the Paas (1992) scale is the standard of practice in worked examples research, it bears explanation why this study did not employ this scale in rating mental effort.

Despite the strong statistics suggesting high validity, three central elements in the Paas (1992) Mental Effort Rating Scale made it a potentially less appropriate instrument for this study of the ill-defined problem of analytical reading comprehension. First, the survey may be open to misinterpretation by the student sample being studied. The question presented in the Paas (1992) scale asks participants to provide an approximation of a subjective state, but it may leave room for error caused by students misunderstanding the terms in the question or having insufficient background knowledge of the construct to answer. Renowned survey methodologist Floyd Fowler (2014) notes that reliable questions provide "consistent measures in comparable situations." In other words, good questions mean the same thing to every respondent. Fowler (2014) cautions against asking questions with words that are not "understood universally" or that use "terms or concepts that can have multiple meanings" (p. 79). "Mental effort," the very

essence of the single-item Paas (1992) scale, is a label for a construct in cognitive psychology under the umbrella of cognitive load theory (Sweller, 1988), and it is a label not likely to be universally understood by the 15- or 16-year-old students who participated in this study. To test this assumption, the researcher conducted casual cognitive interviews with students during the pilot phase to create the instrument. The researcher asked a class of sophomores what the phrase “mental effort” meant to them; their answers ranged from mental effort being very positive (“Isn’t it good to put in effort?”) to very negative (“If something is really hard for me” or “It takes too much energy”). In essence, the Paas (1992) scale would present each respondent in the study with a unique survey, resulting in potentially inaccurate and inconsistent data.

Second, the single-item nature of the scale potentially weakens the reliability of the measure in the context of studying an ill-defined problem such as analytical reading comprehension. In studies of well-defined problems, like in the original Paas (1992) study, the single-item, mental-effort scale follows every individual problem-solving question in the set, thus increasing the overall reliability of the scale. In contrast, because of the nature of ill-defined problems in general and reading comprehension in particular, students may work on only one “problem,” and thus may have only one opportunity to rate their mental effort. To increase the reliability of the scale, Fowler (2014) promotes “ask[ing] multiple questions, with different question forms, that measure the same subjective state,” knowing that statistically speaking, “multiple questions help even out response idiosyncrasies and improve the validity of the measurement process” (p. 97). One strategy to strengthen reliability in survey design is to write questions that ask for observable, easy-to-recall behaviors that serve as a proxy for the construct, instead of asking students directly about the construct itself; the Mental Effort Rating Scale for Adolescent Analytical Reading follows this approach. Rather than allow each student to define

“mental effort” for him- or herself, Fowler (2014) suggests writing questions that ask students for “the information needed to classify their experiences” (p. 92) and rely on a validity panel of construct experts to align those experiences with the construct being studied.

Third, the nine-point continuum scale may offer too many response options for younger students. Students such as the 15- and 16-year-olds who participated in this study may not be able to reliably distinguish between the nine categories in the Paas (1992) scale, since the differences between the options are minimal. To a high school student, “very, very low mental effort” and “very low mental effort” are likely quite similar. Although Fowler (2014) notes that in questions that ask participants to order along a continuum, “it is probably better to have more categories than fewer,” he warns that “there is a limit . . . to the precision of discrimination that respondents can exercise in giving ordered ratings. When the number of categories exceeds the respondents’ ability to discriminate their feelings, numerous categories simply produce unreliable noise” (p. 97).

Paas himself suggests that although the scale is highly reliable and serves as an effective starting point for mental effort research, it is insensitive to the varying types of cognitive load and cannot yet distinguish between intrinsic load, extraneous load, and germane load (Brünken et al., 2010). There is a need for a more accurate instrument to measure students’ perception of mental effort after completing a given task.

Thus, the Mental Effort Rating Scale for Adolescent Analytical Reading was created to leverage the positive findings that Paas (1992) presents about reliability of self-report mental effort ratings, while adapting for an ill-defined task and a younger audience composed of more concrete thinkers. Following are the steps that were taken to construct the instrument, in accordance with Fowler’s (2014) methodology for survey creation and in compliance with the

University of San Francisco Institutional Review Board for the Protection of Human Subjects (exempt research, 45CFR46.101(b)).

Conducting focus groups. During the pilot of the instrument, students from three sections of 10th-grade United States Literature ($n = 45$) were placed into small focus groups of three or four participants each to describe their experience with analytical reading. At the start of the instructional unit, two weeks before starting the focus groups, students were placed into a seating arrangement that—unbeknownst to them—created natural smaller groupings of participants by reading level. These homogeneous grouping helped draw out stories about the challenges of reading by increasing the likelihood that students felt comfortable describing their struggles with analytical reading in front of peers facing similar challenges. Students' reading level was determined by standardized reading comprehension scores on the Scholastic High School Placement Test (HSPT 2C, 2014), the most recent standardized test of reading comprehension that all students had taken. In their small groups, students met with the teacher-researcher in the spring of 2016 for 10- to 15-minute conversations. The teacher-researcher drew upon the following open-ended questions to elicit students' stories about analytical reading:

- Today, I'm going to ask you to tell me about your reading experiences. What is it about reading for English class that is different from reading for fun?
- For our purposes today, we're going to focus on required reading for this class. What helps you understand what you are reading? What gets in the way of your understanding?
- When you read something in this class that's hard to understand, what tells you that it's hard? Describe some of the things going on in your mind when you're trying to read something really challenging.

- Describe a reading experience where you've found yourself shutting down, feeling distracted or totally overwhelmed. What happened in that experience? What triggered your response?
- What about the opposite? Could you describe a reading experience where you found yourself totally focused and in the reading "flow"? What helped you get over obstacles and get sucked into the text?
- What are some of the things we do in class that help you understand or stick with the "tough stuff"? Why are those things helpful? Describe what it's like in your head. How is that thing different from other things we do in class?

From the student responses, the teacher-researcher identified central themes that mapped onto the construct of mental effort within cognitive load theory (Brünken et al., 2010) and grouped students' commentary into three main categories: (a) task-irrelevant behaviors such as environmental and distraction factors that map onto extraneous load, (b) task-relevant behaviors such as specific challenges with the reading assignment that map onto intrinsic load and working memory function, and (c) task-relevant behaviors such as combining or compounding challenges that speak to the element interactivity effect.

Generating instrument questions. To create a scale that would be sensitive to the varying types of cognitive load and based in observable behaviors, 14 questions were developed using language that students used in their focus groups. Each of the questions retained the nine-point ordinal line format of the original Paas (1992) scale, with the expectation that adult and student panelists would comment on any necessary changes. The first six questions addressed the task-irrelevant behaviors that constitute potential barriers to engagement. For example, one question asked, "How much were you distracted by personal factors (hunger, exhaustion, stress,

drama)?" The next seven questions addressed task-relevant behaviors that constitute potential challenges when students become engaged in the reading task, including a question about element interactivity. For example, one task-relevant question asked, "How much did you have to think, to keep track of characters in the passage?" The question pointing to element interactivity was, "To what extent did the combination of all these factors (vocabulary, sentence structure, plot, character, literary elements) get in the way of your understanding of the passage?" The final question was the original Paas (1992) question, asking students to rate their overall mental effort in the passage. This final question was included as an additional external validity check.

Reviewing instrument questions. The scale was reviewed by a content validity panel of three literacy experts, all classroom teachers in English language arts. The first classroom teacher was the English department chair at the school from which the survey sample was taken. He has 18 years of classroom experience teaching English in Grades 9 through 12. The second classroom teacher has been teaching middle and high school English for 12 years and most recently has taught a specialized course for students who struggle with reading. The third classroom teacher, a published author, has been teaching middle and high school English for 20 years, holds an advanced degree in literacy instruction, and teaches part-time at the university level. The three panelists provided feedback on the wording of questions, critical concepts that appeared to be missing from the survey, the order of questions, and the format of response choices. Four key changes emerged from the content validity panel:

- changing the wording of the question stem from "How much did you have to think to ..." to "How much thinking did you have to do to ...";

- clarifying the distinction between the two questions that ask about re-reading, so that one is clearly re-reading because of a loss of focus and the other is re-reading to deepen understanding;
- modifying the question about “keeping track of characters” so that it is about the more challenging task of “keeping track of character relationships”; and
- adding an additional question that asks students to rate their overall understanding of the passage.

Two of the three panelists pointed out that nine options along a number line would be too many for students to differentiate between, but that five would be too few.

The scale was also reviewed by a construct validity panel of two experts in cognitive load theory. One is a full professor at the School of Education at the University of San Francisco who teaches the cognitive psychology course. The other is a former high school English teacher and current high school administrator and adjunct professor, who holds an EdD and whose dissertation focused on applying the principles of cognitive load theory to multimedia design. Both panelists provided feedback on the degree of alignment between (a) the questions and observable behaviors and (b) the constructs they sought to measure, specifically in regard to determining task-relevant and task-irrelevant burdens on the working memory and identifying element interactivity. Key suggestions that emerged from the construct validity panel included

- removing the question about “How much did you have to try to become or stay interested?” since it mapped onto a different construct—motivation and interest—rather than mental effort;
- clarifying the question about re-reading in the task-irrelevant section;

- adding a question, in student-friendly language, about how much “brainpower” it took to analyze the passage, as a parallel to the original Paas (1992) question; and
- changing the order of the opening sequence of questions.

The panelists agreed that nine choices was an overwhelming number for younger students, and thus recommended seven choices on the number line. After gathering all of the feedback from both panels, the teacher-researcher revised the scale and proceeded to conducting cognitive interviews.

Conducting cognitive interviews. The scale was then vetted by two students—a intermediate reader and a struggling reader—through a series of cognitive interviews. During this phase, the students read each question aloud, explained what the question meant to them in their own words, pointed out any language that was confusing to them, and offered suggestions for clarifying confusing questions. They explained how they would respond to each question and why they chose a particular number on the number line. The students agreed that the original text for the questions was too wordy and that, for their ease in responding, it would be preferable to use only one word or phrase to describe a behavior instead of two or three. For example, they suggested changing the question asking if they “zoned out, lost focus, or found their mind wandering” to simply asking about zoning out. “If all three phrases mean essentially the same thing,” one student suggested, “why bother saying it more than once?” Both student said that nine choices in the number line were too many response options for them and that they did not know what it would mean to answer “one” instead of “two” or “eight” instead of “nine.” When asked if seven options made more sense, they both said yes. Based on student feedback, revisions were made to the scale before implementing the pilot test of the instrument.

Pilot testing. Students in this pilot were just starting their study of *Lucy: A Novel*, by Jamaica Kincaid (1990). The timing of the survey within the *Lucy* unit mirrored the timing of the present study. The reading task assigned before using the mental effort rating scale was a passage from *Lucy* that includes many literary elements, particularly color imagery, that are similar to the literary patterns in *The Great Gatsby* (Fitzgerald, 1925). At the start of the class period, students spent an average of 10 minutes reading the passage, with the instruction that they should read critically, apply the analytical reading strategies they had been practicing all year, and look for stylistic connections between the excerpt and the novel as a whole.

Upon completion of the reading task, students were directed to Schoology to take the electronic survey. The instructions for the survey said, “Please focus your responses on the passage you just analyzed.” The Mental Effort for Adolescent Analytical Reading instrument was provided as a Google Form with radio buttons next to each of the scale options (<http://goo.gl/forms/VPx85PYVW1>). The teacher-researcher selected Google Forms as the software for collecting data because the school is a Google Apps for Education school, students are familiar with the Google suite, and students had taken surveys and quizzes using Google Forms earlier in the school year.

Students completed the survey within five minutes. At the end of the survey, they were asked for feedback and, in particular, whether they had any difficulty with any of the questions or if any of the wording was confusing. Of the 44 students present in three classes to take the survey, none expressed difficulty or confusion. Two students noted that some of the questions seemed to be asking the same thing. In a show of hands, the overwhelming majority of students agreed that the questions accurately represented the types of experiences they have when reading challenging texts.

Calculating reliability and validity. A spreadsheet of student responses was analyzed using the Statistics Package for the Social Sciences (SPSS). Questions were collapsed into scales, including Task Irrelevant (TI) behaviors, Task Relevant (TR) behaviors, Element Interactivity (EI), and Total (TOT). Within scales, student responses were coded as continuous variables ranging from one to seven; one represented very low mental effort or mental load on the working memory, and seven represented very high effort or load. Cronbach's Alpha was used to measure the internal consistency of each scale and thus the overall reliability of the survey. Three of the four scales exceeded the threshold for an adequate measure of reliability in an affective study with a Cronbach's Alpha above .70. The Cronbach's Alpha for Task Irrelevant behaviors was .84, for Task Relevant behaviors it was .80, and for Element Interactivity it was .74. In looking at the Item-Total Statistics, it is evident that each of items on the measure is rather strong, because none of them significantly improves the overall reliability statistic if deleted (Tables 5, 6, and 7).

Table 5
Reliability Statistics, Task Irrelevant (TI) Behaviors

	<i>M</i>	<i>SD</i>	Cronbach's Alpha if Item Deleted
TI Zone Out	2.91	1.59	.80
TI ReRead	2.86	1.92	.78
TI Classroom Distract	2.70	1.77	.85
TI Personal Factors	3.60	1.85	.81
TI Worry	4.23	1.90	.83
TI Interest	4.16	1.80	.80

Table 6

Reliability Statistics, Task Relevant (TR) Behaviors

	<i>M</i>	<i>SD</i>	Cronbach's Alpha if Item Deleted
TR Vocab	2.84	1.53	.79
TR Syntax	3.98	1.78	.76
TR Plot	3.47	1.58	.74
TR Character	3.37	1.89	.74
TR Lit Devices	5.05	1.41	.76
TR ReRead	4.30	1.77	.81

Table 7

Reliability Statistics, Element Interactivity (EI)

	<i>M</i>	<i>SD</i>	Cronbach's Alpha if Item Deleted
EI Layering	4.58	1.50	.(a)
EI Context	4.81	1.44	.(a)

a. Cronbach's Alpha for Element Interactivity is .74. Because this is a bivariate correlation, it is not possible to report internal consistency without one of the two items.

In the fourth scale, overall mental effort and understanding (TOT), the Cronbach's Alpha was .32, which can probably be attributed to including the unrelated item about *understanding* in the scale, an item that was called into question by the construct validity panel. With the item concerning *overall understanding* removed, the Cronbach's Alpha increases to .67, which is still below the .70 threshold. (See Table 8.) Thus it seems reasonable to remove this question and separate the remaining two questions in the TOT scale into individual questions, since they do not exhibit a strong enough measure of internal consistency. The question on brainpower is more strongly correlated with the intervention target area of task relevant behaviors ($r = .72$) than is the original Paas (1992) question on mental effort ($r = .48$), suggesting that if only one question is to be retained, it should be the overall question on brainpower.

Table 8

Reliability Statistics, Overall Mental Effort and Understanding (TOT)

	<i>M</i>	<i>SD</i>	Cronbach's Alpha if Item Deleted
TOT Brain Power	4.40	1.35	.53
TOT Understanding	5.00	.93	.67
TOT Paas Mental Effort	5.26	1.12	-.20

To ensure that the Mental Effort Rating Scale for Adolescent Analytical Reading measures the construct of mental effort accurately, validity was calculated in two ways: correlation to corresponding external measures and internal correlation. The data set was divided into equal thirds based on the external measure of the Scholastic High School Placement Test Reading Scale (HSPT 2C, 2014), to compare the mental effort ratings of the most advanced readers against those of the most novice readers, yielding a statistically significant, moderately weak negative correlation ($r = -.37$) between task relevant mental effort and externally determined reading level. The more advanced readers tended to experience lower cognitive load from the task relevant items, whereas the more novice readers tended to experience higher cognitive load from those same items. An independent samples t-test reveals a statistically significant difference between novice and advanced readers on the scale TR, with a mean difference of .76. (See Table 9.)

Table 9

Construct Validity for Task Relevant (TR) Behaviors

	Novice (<i>n</i> = 15)		Advanced (<i>n</i> = 14)		<i>r</i>	Mean Diff
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Task Relevant Behaviors	4.02	1.08	3.26	.87	-.37*	.76

*Correlation is significant at the .05 level (2-tailed).

Refining the instrument. As discussed in the calculation of reliability, the question asking students to rate their overall understanding will be removed. Frequencies on this question illustrate a strong social desirability effect, with more than 90% of students selecting 4, 5, or 6 for their level of understanding. Many of the limitations of the instrument stem from the age-old problem with the reliability of self-reported data. As the intercorrelations illustrate, self-reported experiences correlate with other self-reported experiences, but do not correlate strongly with external objective measures. The only way to counteract this effect is to augment the data collected by this instrument with additional data, such as objective tests and standardized scores. Additionally, on the advice of the construct validity panel, the question about becoming or staying interested in the passage will be removed, because it maps onto the construct of motivation rather than cognitive load. Because the question on brainpower is more strongly correlated with the intervention target area of task relevant behaviors than the question on mental effort is, only the brainpower question will be retained. Thus the final instrument contains 14 questions. The first section, titled “Barriers to Engagement,” contains five questions that target task-irrelevant behaviors and includes the following prompts:

1. How much did you find yourself zoning out while you were supposed to be analyzing the passage?
2. How much did you find yourself re-reading the passage because you zoned out?
3. How much were you distracted by the classroom environment (climate, noise, classmates)?
4. How much were you distracted by personal factors (hunger, exhaustion, stress, drama)?
5. How much did you find yourself worrying about not understanding the passage?

The second section is titled “Once You’re Engaged in the Passage.” The subsequent six questions target task-relevant behaviors:

1. How much thinking did you have to do to make sense of the vocabulary in the passage?
2. How much thinking did you have to do to make sense of the complex sentences in the passage?
3. How much thinking did you have to do to follow what happened in the plot of the passage?
4. How much thinking did you have to do to keep track of character relationships in the passage?
5. How much thinking did you have to do to understand the significance of the literary elements (symbols, motifs, theme, etc.) in the passage?
6. How much did you re-read to deepen your understanding of the passage (not because you zoned out)?

At the end of the second section are two questions aimed at measuring task relevant behaviors caused by element interactivity:

1. How much did the combination of all the factors above (vocabulary, complex sentences, plot, character relationships, literary elements) challenge your understanding of the passage?
2. How much thinking did you have to do to understand the significance of the passage in connection to the novel as a whole?

The third and final section is titled “Overall ...” and asks students to rate their total mental effort:

1. How much brainpower did it take for you to understand the passage you analyzed?

Student Experience

The qualitative strand of the study engaged students in a series of written annotations and open-ended interviews to supplement the findings from the first, quantitative strand of data collection. Passage annotations present a visible record of the student meaning-making process, whereas one-on-one interviews provide a retrospective account experience of engaging in studying literature with multimedia think-aloud worked examples.

Passage annotations. After each phase of the study, students submitted the passage annotation handout (Appendix E) containing their notes on the passages with which they had worked that day. Annotations were reviewed by the teacher-researcher to identify patterns in the quantity and quality of notes that students made during the intervention, focusing in particular on the depth of thinking present and the nature of students' engagement with the instructional materials. Depth of thinking was coded in relation to the level of cognitive complexity the annotation demonstrated (Keene & Zimmermann, 2007; Olson & Land, 2007). The teacher-researcher used an existing organizational structure for dividing annotation types into surface-level, mid-level, and deepest level, which is a play on the common classroom analogy of teaching reading comprehension like the parts of an iceberg (Gallagher, 2004; Hollenbeck & Saturnus, 2013). This analogy is so prevalent among middle and high school teachers that it has been fashioned into a popular classroom poster, and so true to the form of fiction that it most likely originated with one of the most famous writers in the American literary canon (Hemingway, 1932).

Typical annotations in a 10th-grade English class range from (a) surface-level annotations, such as marking plot points, defining vocabulary, and reacting to the plot; to (b) mid-level annotations, such as asking "why" questions about the passage, identifying literary

devices in the passage, and making character inferences based on the passage; to (c) deepest-level annotations, such as identifying literary patterns like motifs, symbols, and archetypes that exist beyond the parameters of the passage; describing character development, conflict, or motivation over time; and articulating themes, thematic ideas, or thematic dualities.

Student interviews. In addition to considering student annotations as a qualitative measure, the teacher-researcher interviewed every student in the treatment condition to better understand the experience of analyzing the novel with the support of the multimedia think-aloud worked example. Students met one-on-one with the teacher-researcher for a conversation lasting between five and seven minutes. Interviews drew upon these questions as a starting point:

1. When you've worked through passages on your own without my support, what has that process been like? Felt like?
2. Can you describe what's going through your mind when you get to a tricky passage in *Gatsby*? What strategies are you using to make sense of things you don't understand right away?
3. Now let's focus on the two days when you analyzed *Gatsby* using the videos of me thinking out loud. Walk me through your experience on those two days, as if I wasn't in the room and didn't know the process. What happened in the video?
4. Could you describe what it was like for you to study the passage using the videos? What was going on in your mind while you were playing the think-aloud?
5. How was studying with the video similar to, or different from, your experiences of re-reading and studying a passage on your own?
6. What do you think teachers should know if they are going to use these videos with their English students?

In some instances, other questions arose during the natural course of conversation or as a follow-up to student comments. All interviews were recorded for later review and transcribed by a professional transcriptionist. The teacher-researcher coded and analyzed the transcripts, using an in vivo coding process that adhered to the steps laid out in Gray's *Doing Research in the Real World* (Gray, 2014, p. 604), which are described in the data analysis section of this chapter. Because the qualitative strand of the study served purely as a supplement to the quantitative strand of the study, multiple raters were not utilized.

Procedures

After receiving approval from the University of San Francisco Institutional Review Board for the Protection of Human Subjects, the co-researcher and teacher-researcher began the process of creating matched pairs for the randomized block design using students' reading scores on the Scholastic High School Placement Test (HSPT 2C, 2014). In a previous unpublished study of the reliability of the Scholastic High School Placement Test, the teacher-researcher found the assessment to have adequate reliability scores for a norm-referenced test. Per the request of the test's publisher, Scholastic, Inc., the specific details of those findings cannot legally be included in this publication.

First, the co-researcher gathered participants' Scholastic High School Placement Test (HSPT 2C, 2014) reports from the school admissions office. She assigned each student a random number on the top of the report and keep a master roster matching student names and random numbers. Then she redacted all identifying elements from the report, such as name, address, gender, and birthday, so that the primary researcher, who is also the students' classroom teacher, would not know which report belonged to which student. Together, the co-researcher and teacher-researcher determined the closest possible pairs using the Reading Standard Score,

Reading Local Percentile, Reading National Percentile, and Reading Comprehension Raw Score from the HSPT (HSPT 2C, 2014) report card for comparison. Pairs were recorded in the master roster. Individuals within pairs were randomly assigned to either the treatment or comparison condition (Figure 6).

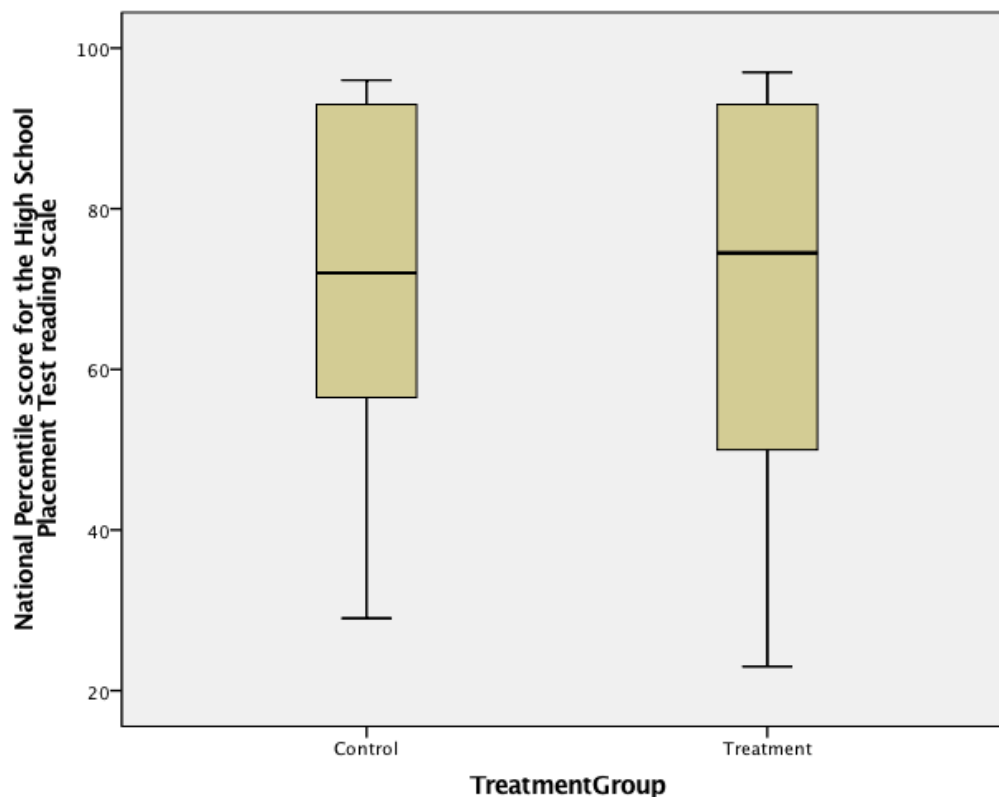


Figure 6. Box plot of group demographics for the national standardized percentile score in reading comprehension on the High School Placement Test (HSPT).

Next, the co-researcher created two identical course shells in Schoology, one for the treatment group and one for the comparison group. She gave temporary access to the teacher-researcher, who uploaded all of the group-specific content for the study—including videos, links to assessments, and instructions—and scheduled publication windows such that the materials would be available only for the duration of the class period during which the study took place.

When this step was complete, the co-researcher removed the teacher-researcher’s access to these courses, enrolled students into the appropriate section, and renamed the courses so that they had identical course names for all students participating in the study. Students did not have access to any materials until the intervention begins.

The intervention phase took place over a period of four consecutive hour-long class periods. For an overview of the intervention, see Table 10.

Table 10

Overview of Procedures by Class Period

Class Period	Description
1	Read Chapter 1, baseline mental effort assessment, baseline analytical reading comprehension assessment
2	Review Chapter 1, read Chapter 2 independently
3	Intervention phase one: administer video lesson, mental effort assessment, Chapter 2 analytical reading comprehension assessment
4	Chapter 2 whole-class activity, read Chapter 3 independently
5	Intervention phase two: administer video lesson, mental effort assessment, Chapter 3 analytical reading comprehension assessment
6	Read first half of Chapter 4 out loud together, complete independently
7	Read Chapter 5 independently, mental effort delayed posttest, Chapter 5 analytical reading comprehension delayed posttest
8–10	One-on-one interviews with students in the treatment group during independent reading time

Note. Each class period is 60 minutes long.

Day 1: On the first day of the study, students completed their reading of Chapter 1 of *The Great Gatsby* (Fitzgerald, 1925). All reading took place during class time to ensure that students read attentively and did not access online summary materials (such as *SparkNotes*, *Cliff’s Notes*, and *Shmoop*) instead of, or in addition to, their own reading. When they finished the chapter, they completed the baseline assessment, which included the mental effort rating scale and an analytical reading quiz, which used a standard prompt that appeared in each of the subsequent analytical reading quizzes.

Day 2: In preparation for the instructional intervention, students review Chapter 1 to clarify any confusion that naturally occurs at the beginning of a new novel. Then they spent the remainder of the class period reading Chapter 2, which took students between 20 and 30 minutes. Students who finished early had an alternative, unrelated assignment on which to work (ongoing personal narrative blogging project).

Day 3: When students came into class, they each received their random research number from the co-researcher on a small circular sticker, which they placed inside the back cover of their books. They were told that this number would be their only identifier for the work they submitted during the course of the study. Taking out their iPads and headphones, they then logged in to the course to which they have been randomly assigned in Schoology, where the intervention module laid out all of the instructions for the day. All students were given verbal instructions to watch the instructional video and take notes in their text. Students had three tasks to complete, all of which were self-paced within the one-hour period. Any students who finished before the bell had an alternative, unrelated assignment on which to work (silent reading time from their free-reading book). When students entered the intervention module on Schoology, they first clicked on the video. The video component of the lesson took most students between 20 and 25 minutes, depending on how many times the students paused or rewound the video. Upon completion of the video, students were directed to click the link in Schoology to take the Mental Effort Rating Scale for Adolescent Analytical Reading. The final assessment for the class period was the analytical reading quiz based on the passage studied in the video, which was also linked in Schoology. Each instrument included clear directions for student procedures. Both the ratings on the mental effort scale and the responses to the comprehension quiz were typed and collected via Google Forms. All responses, whether from students in the treatment or in the

comparison, funneled into a singular database to minimize bias in grading. Students were identified only by their random number, which kept student identities hidden from the teacher-researcher during the quantitative strand of the study. (Note that during data collection, the co-researcher automatically collected students' school-administered email address and cross-referenced each name against the research number to ensure accuracy in reporting; she then removed column containing email addresses prior to sending the final spreadsheet to the teacher-researcher.)

Days 4 and 5: For the second iteration of the worked examples intervention, students repeated the process outlined above for Days 2 and 3, focusing their study on Chapter 3 of *The Great Gatsby* (Fitzgerald, 1925) instead of Chapter 2.

Days 6 and 7: Following the two phases of the instructional intervention, and after reading Chapters 4 and 5, students received a parallel passage from Chapter 5 that contained many of the same literary elements and image patterns as the passages studied with the videos. Students were given instructions identical to those given during the baseline assessment; when they finished the chapter, they completed the posttest assessment, which included the mental effort rating scale and an analytical reading quiz. At this point, the intervention phase was complete and students carried on with the remainder of the unit, studying *The Great Gatsby* (Fitzgerald, 1925) as a whole class.

Days 8, 9, and 10: In the week after the intervention, the teacher-researcher conducted and recorded interviews with students in the treatment group.

Data Analyses

This explanatory sequential mixed-methodology experimental study included a quantitative and a qualitative strand. Descriptive statistics, including means and standard deviations, for student reading comprehension scores and mental effort scores are reported in Chapter IV. In conducting the statistical analyses, the significance level was set at 0.05 for each two-tailed test. Cohen’s *d*, a measure of practical importance which describes the standard deviation difference between group means, is reported in Chapter IV. What follows is an overview of the data analysis procedures by research question:

Table 11

Data Analyses by Research Question

RQ	Data Collected	Proposed Analysis
1. Analytical Reading Comprehension (COMP)	Baseline Analytical Comp Intervention Phase I Comp Intervention Phase II Comp Delayed Posttest Comp	<i>t</i> tests
2. Mental Effort (ME)	Baseline ME Intervention Phase I ME Intervention Phase II ME Delayed Posttest ME	<i>t</i> tests
3. Student Experience	Passage annotations, one-on-one interviews	Qualitative analysis, coding for themes and patterns

The first research question addresses the effect of multimedia think-aloud worked examples on measures of analytical reading comprehension. In order to verify that the paired block design was effective in creating two equivalent groups, Levene’s Test of Equality of Variances was conducted for the baseline Analytical Reading Comprehension quiz.

Subsequently, data were analyzed using a series of *t* tests between treatment and comparison groups for each phase of the study, as well as for gain scores from the baseline assessment to each of the intervention phases and to the posttest. The second research question addresses the effect of multimedia think-aloud worked examples on measures of mental effort. Again, *t* tests

were conducted between treatment and comparison groups for both phases of the intervention. The third research question supplements the first two by considering the participant experience of studying multimedia think-aloud worked examples through qualitative evidence, such as written annotations and individual interviews of participants. Qualitative data were processed, coded, and analyzed by the teacher-researcher in accordance with Gray's qualitative coding process (Gray, 2014, p. 604), which includes transcription of the data, high-level familiarization with the data, a focused reading of the data to begin marking key words and phrases, an initial coding of the data followed by a review and revision of codes, and identification of emerging themes and patterns.

CHAPTER IV

RESULTS

The purpose of this study was to investigate how multimedia worked examples that explicitly model the reading habits of successful readers through teacher think-alouds could facilitate effective analytical reading instruction for high school English language arts students who have access to 1:1 technology. In the quantitative strand of the study—research questions one and two—a classic treatment-comparison repeat measures pretest-posttest design was used, and students' analytical reading comprehension and perceived mental effort was assessed. In the qualitative strand of the study—research question three—the participant experience was described through interviews and annotations in order to more deeply understand the quantitative data presented.

All quantitative data were analyzed using a series of *t* tests between treatment and comparison groups for each variable at each phase of the study. In each of the *t* tests conducted, regardless of the statistical significance of the outcome, Cohen's *d* was calculated to quantify the magnitude of the difference between group means using standard deviation units (Cohen, 1988). Because of the relatively small sample size for each group (treatment $n = 18$, comparison $n = 16$), the benefit of including effect size—the difference in group means divided by the pooled standard deviation—is that it provides a detailed sense of what is happening with the data that is independent of sample size. An effect size of $d > .80$ (eight tenths of a standard deviation difference) is considered a large effect; $d > .50$ (half of a standard deviation difference) is considered a moderate effect; $d > .20$ (two tenths of a standard deviation difference) is considered a small effect; and $d < .20$ is considered as having no effect.

This chapter opens with the quantitative strand, presenting descriptive statistics, *t* test results, and effect sizes for research questions one and two. Then it discusses the qualitative data collected for research question three, describing the codes that were used and the themes that emerged from the interviews and annotations. Finally, the chapter closes with a section dedicated to ancillary results that were beyond the scope of the initial research questions, but that are germane to the greater discussion of worked examples and adolescent literacy development.

Research Question 1: Analytical Reading Comprehension Results

To what extent is there an effect of multimedia think-aloud worked examples on analytical reading comprehension in a secondary English language arts class, as measured by open-ended written responses to literature?

The first research question investigated the difference in mean scores between the treatment and comparison conditions on an assessment of analytical reading comprehension at four testing periods: the baseline assessment before any intervention had occurred, the first phase of the intervention, the second phase of the intervention, and the delayed posttest after the intervention was completed. Prior to running the independent-samples *t* tests for each testing period, Levene's test of Equality of Variances was conducted to verify the assumption of homogeneity of variance. The significance value was .72, which confirms that the two group variances can be treated as equal. The rubric used to measure analytical reading comprehension had fifteen total possible points. Scores ranged from four to fourteen. Table 12 presents descriptive statistics, independent-samples *t* test results, and effect sizes (Cohen's *d*) for the treatment and comparison conditions on the dependent variable of Analytical Reading Comprehension at four testing intervals.

Table 12

Analytical Reading Comprehension Results: Descriptive Statistics, Independent-Samples t Test Results, and Effect Size at Four Testing Intervals.

	Treatment (<i>n</i> = 18)		Comparison (<i>n</i> = 16)		<i>t</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Baseline Assessment	7.83	1.65	7.50	1.67	.58	.20
Intervention Phase I	9.22	2.39	7.25	2.44	2.38*	.82
Intervention Phase II	9.22	2.80	7.88	1.86	1.63	.58
Delayed Posttest	9.61	2.00	8.19	1.87	2.13*	.73

*Difference in means is significant at the .05 level (2-tailed).

The treatment group ($M = 7.83$, $SD = 1.65$) and comparison group ($M = 7.50$, $SD = 1.67$) began the study with relatively equal means ($d = .20$). After the first phase of the instructional intervention, the treatment group exhibited, on average, statistically significantly higher scores in analytical reading comprehension ($M = 9.22$, $SD = 2.39$) than the comparison group ($M = 7.25$, $SD = 2.44$), $t(32) = 2.38$, $p < .05$. The effect size ($d = .82$) was large, favoring the multimedia think-aloud worked example treatment condition over the traditional text-based questions comparison condition. After the second phase of the instructional intervention, the treatment group ($M = 9.22$, $SD = 2.80$) out-performed the comparison group ($M = 7.88$, $SD = 1.86$) in analytical reading comprehension, but this difference was not statistically significant. The effect size ($d = .58$) was moderate, favoring the multimedia think-aloud worked example treatment condition over the traditional text-based questions comparison condition. On the delayed posttest, after the conclusion of the instructional intervention, the treatment group exhibited statistically significantly higher mean scores in analytical reading comprehension ($M = 9.61$, $SD = 2.00$) than the comparison group ($M = 8.19$, $SD = 1.87$), $t(32) = 2.13$, $p < .05$. The effect size ($d = .73$) was moderately large, again favoring the multimedia think-aloud worked example treatment condition over the traditional text-based questions comparison condition. Figure 7

illustrates group means as a line chart of analytical reading comprehensions scores at the four time intervals for students in the treatment and the comparison condition.

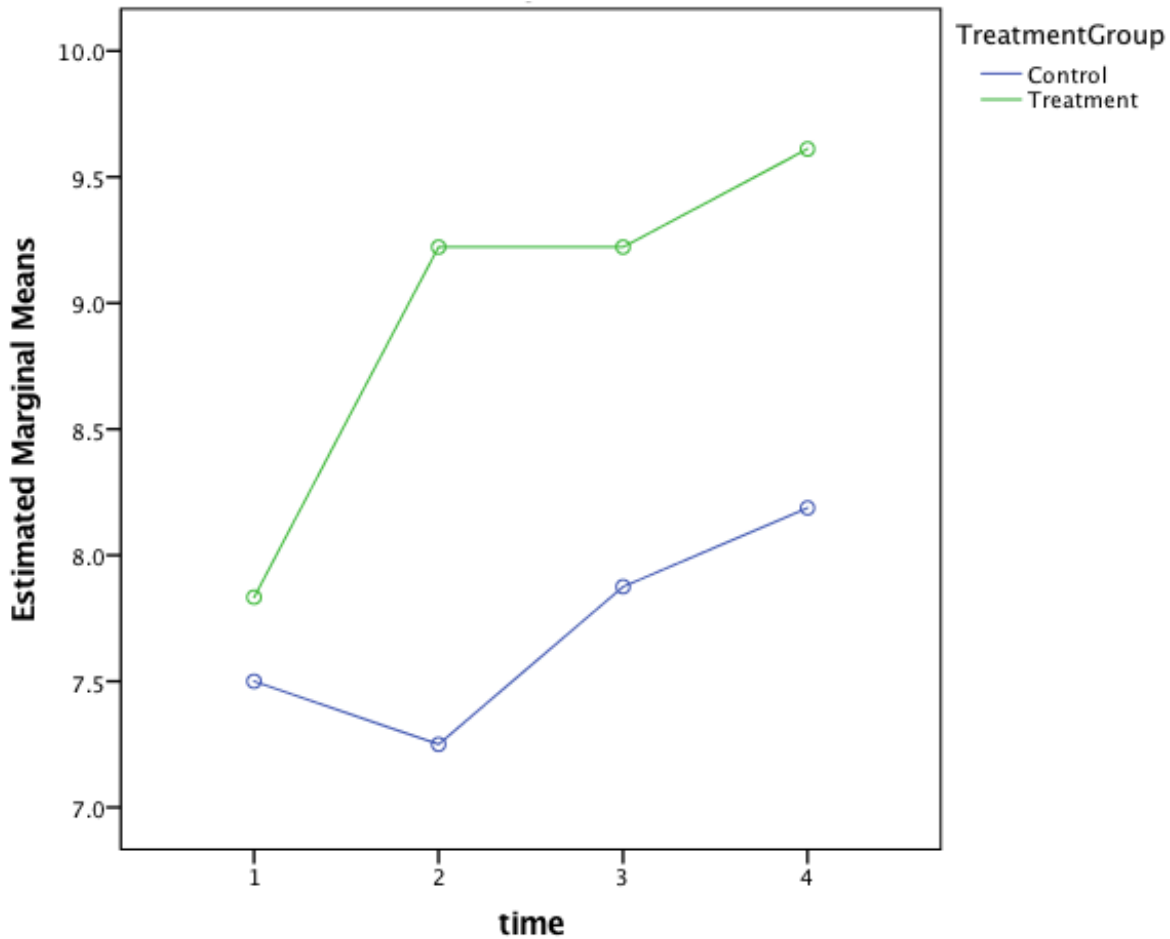


Figure 7. Line graph of analytical reading comprehension scores at four time intervals for students in treatment and control groups.

Figure 7 reveals a difference in mean scores, with the treatment group (top line) scoring higher than the comparison group (bottom line). Importantly, it illustrates the pattern of scores for both groups. The greatest leap in scores for students in the treatment group was between the baseline and the first phase of the intervention. After the second phase of the intervention, the scores for students in the treatment group stagnated. Scores for students in the treatment group increased again at the posttest. In contrast, scores for students in the comparison group dropped between

the baseline and the first phase of the intervention, and steadily rose after the second phase of the intervention and the posttest, though they remained significantly below the scores of the treatment group by the end of the study.

In order to further clarify the differences in mean scores between the treatment and comparison conditions, gain scores were calculated by subtracting each student's baseline score from his or her subsequent scores. Table 13 provides descriptive statistics, independent-samples *t* test results, and effect size calculations for gain scores in analytical reading comprehension at three testing intervals.

Table 13

*Gain Score Results in Analytical Reading Comprehension: Descriptive Statistics, Independent-Samples *t* Test Results, and Effect Size at Three Testing Intervals.*

	Treatment (<i>n</i> = 18)		Comparison (<i>n</i> = 16)		<i>t</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Gain: Intv. I – Baseline	1.39	1.96	-.25	2.02	2.39*	.82
Gain: Intv. II – Baseline	1.39	2.17	.38	1.26	1.64	.59
Gain: Posttest – Baseline	1.78	1.35	.69	1.08	2.58*	.90

* Difference in means is significant at the .05 level (2-tailed).

For the first gain score between Intervention Phase I and the baseline assessment, the treatment group exhibited significantly higher mean scores in analytical reading comprehension ($M = 1.39$, $SD = 1.96$) than the comparison group ($M = -.25$, $SD = 2.02$), $t(32) = 2.39$, $p < .05$. The effect size ($d = .82$) was large, favoring the multimedia think-aloud worked example treatment condition over the traditional text-based questions comparison condition. For the second gain score between Intervention Phase II and the baseline assessment, the treatment group ($M = 1.39$, $SD = 2.17$) performed higher than the comparison group ($M = .38$, $SD = 1.26$) in analytical reading comprehension, but this difference was not statistically significant. The effect size ($d =$

.59) was moderate, favoring the multimedia think-aloud worked example treatment condition over the traditional text-based questions comparison condition. For the third gain score between the delayed posttest and the baseline assessment, the treatment group exhibited significantly higher mean scores in analytical reading comprehension ($M = 1.78$, $SD = 1.35$) than the comparison group ($M = .69$, $SD = 1.08$), $t(32) = 2.58$, $p < .05$. The effect size ($d = .90$) was large, favoring the multimedia think-aloud worked example treatment condition over the traditional text-based questions comparison condition.

Research Question 2: Mental Effort Results

To what extent is there an effect of multimedia think-aloud worked examples on cognitive load, as measured by self-reported mental effort (Paas, 1992)?

The second research question investigated the difference in mean scores between the treatment and comparison condition on a scale of perceived mental effort at four testing periods: the baseline assessment before any intervention had occurred, the first phase of the intervention, the second phase of the intervention, and the delayed posttest after the intervention was completed. Prior to running the independent-samples t tests for each testing period, Levene's test of Equality of Variances was conducted to verify the assumption of homogeneity of variance. The significance value was .97, which confirms that the two group variances can be treated as equal. To report perceived mental effort, students used a seven-point scale that started at one and ended at seven with a midpoint of four. Scores ranged from one to five. Table 14 presents descriptive statistics, independent-samples t test results, and effect sizes (Cohen's d) for the treatment and comparison conditions on the dependent variable of Mental Effort at four testing intervals.

Table 14

*Mental Effort Rating Results: Descriptive Statistics, Independent-Samples *t* Test Results, and Effect Size at Four Testing Intervals.*

	Treatment (<i>n</i> = 18)		Comparison (<i>n</i> = 16)		<i>t</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Baseline Assessment	2.91	.90	3.08	.90	-.56	.19
Intervention Phase I	2.96	.87	3.11	.90	-.50	.17
Intervention Phase II	3.10	.82	3.17	.95	-.24	.08
Delayed Posttest	2.93	.62	2.94	.65	-.03	.01

* Difference in means is significant at the .05 level (2-tailed).

The treatment group ($M = 2.91$, $SD = .90$) and comparison group ($M = 3.08$, $SD = .90$) began the study with relatively equal means ($d = .19$). After the first phase of the instructional intervention, the treatment group ($M = 2.96$, $SD = .87$) exhibited, on average, slightly lower mental effort scores than the comparison group ($M = 3.11$, $SD = .90$), but this difference was not statistically significant. The effect size ($d = .17$) was very small, slightly favoring the multimedia think-aloud worked example treatment condition over the traditional text-based questions comparison condition. After the second phase of the instructional intervention, the treatment group ($M = 3.10$, $SD = .82$) exhibited slightly lower mental effort mean scores than the comparison group ($M = 3.17$, $SD = .95$), but this difference was not statistically significant. The effect size ($d = .08$) was so small as to suggest that there was no effect. On the delayed posttest, after the conclusion of the instructional intervention, the treatment group ($M = 2.93$, $SD = .62$) exhibited mental effort mean scores that were practically equal to those of the comparison group ($M = 2.94$, $SD = .65$). The effect size ($d = .01$) was infinitesimal. Further statistical tests, which would have included line graphs and gain scores, were not necessary to clarify the patterns in scores and thus were not conducted.

Research Question 3: Student Experience Qualitative Results

What do student annotations and interviews reveal about the participant experience of studying multimedia think-aloud worked examples?

The second, qualitative strand of data collection engaged students in a series of written annotations and open-ended interviews to supplement the findings from the first, quantitative strand of data collection. These qualitative data offer a window into the student experience around analytical reading comprehension and mental effort, while also serving to augment the quantitative findings reported for the previous two research questions. Passage annotations presented a visible record of the student interpretation process, and one-on-one interviews provided a personal explanation of the cognitive and metacognitive experience of studying literature with multimedia think-aloud worked examples. Annotations and interview transcripts were processed, coded, and analyzed in accordance with Gray's qualitative coding process (Gray, 2014).

Passage Annotations

First, the teacher-researcher sorted the passage annotations by chapter and read through each document to familiarize herself with the students' comments. In the preliminary stage of data analysis, she made initial notes about general observations, such as the shift from students underlining large portions of the text without adding written annotations in Chapter 1, before the instructional intervention began, to students circling and boxing more specific lines and phrases with written annotations in the margins of later chapters.

In order to prioritize depth of analysis over breadth, the teacher-researcher selected five sets of matched student pairs as sources of variation and contrast for careful inspection. These pairs had been established at the start of the study when students were matched with the closest

possible partner using the Reading Standard Score, Reading Local Percentile, Reading National Percentile, and Reading Comprehension Raw Score from the HSPT (HSPT 2C, 2014), and then were randomly assigned to either the treatment or comparison condition. The pairs selected for the in-depth qualitative analysis represented a broad distribution of reading levels, including high (95th–96th percentile), medium-high (89th–90th percentile), medium (71st–76th percentile), medium-low (50th percentile), and low (46th percentile). The “medium” matched pair had a wider percentile range than the other pairs, but the raw score on the HSPT (HSPT 2C, 2014) reading comprehension test was identical. Otherwise, pairs scored one or zero percentile points apart.

Once the five matched pairs were selected, the teacher-researcher conducted a focused reading of the data, underlining key words and phrases in students’ annotations. Beyond considering sheer quantity of annotations, it was imperative to decipher any differences in the quality of annotations, as indicated by the depth of thinking depicted by the notes and scribbles on the page. In keeping with literacy research on the types of cognitive processes in which students engage during reading (Beers, 2003; Frey & Fisher, 2013; Keene & Zimmermann, 2007; Olson & Land, 2007), annotations were coded based on existing descriptions of analytical reading levels: surface level, mid-level, and deepest level annotations. Table 15 depicts these three categories along with examples of student remarks that exemplify the categories. Surface level annotations stayed at the plot and vocabulary level. Mid-level annotations moved beyond plot but were restricted to the parameters of the passage itself. Deepest level annotations moved beyond the passage to situate it within the context of the novel as a whole.

Table 15

Student Annotation Coding System

Level	Description	Example
Surface Level Annotations	Marking plot points	Underlines, check marks, stars, and hearts
	Defining vocabulary	Prodigality means excessive, extravagant waste
	Reacting to the plot	Daisy is crying ☹
	Expressing confusion	Where are they?
Mid-Level Annotations	Asking “why” questions	Why is there so much personification here?
	Making connections to other parts of the selected passage	Gatsby seems to have changed from that paragraph [arrow up to previous paragraph]
	Making predictions based on the passage	Is something bad going to happen? Lots of dark imagery.
	Drawing character inferences based on the passage	Gatsby is just trying to show off and look rich.
	Identifying literary devices in the passage	Motif of weather here.
Deepest Level Annotations	Explaining the effect of literary elements in the passage	Polysyndeton! So many ANDs! Makes reader overwhelmed by all the elements of the party.
	Connecting patterns in this passage, such as motifs, symbols, and archetypes to other passages in the novel or other texts students had read	Juxtaposition of rich vs. poor. Valley of Ashes (grey, depressing, dusty) here is really different from East Egg in last chapter (good weather, white, gold, glowy, happy, carefree).
	Describing character development, character conflict, or character motivation over time	Gatsby is usually green/new money, but this is gold/old money. Maybe Gatsby is trying to become old money so he can start a new life with Daisy [marks “dull gold”].
	Teasing out themes, thematic ideas, or thematic dualities	He doesn’t even describe the people—they’re just described as their belongings. Is Gatsby more interested in the objects than the people who come? Money vs. human connection.

With a code established, the teacher-researcher returned to the stacks of marked-up passages and highlighted each annotation in one of three colors: green for surface level annotations, blue for mid-level annotations, and pink for deepest level annotations. Each note or

underlined segment counted as one annotation. Any patterns that were marked or circled were counted together as one annotation. For example, in Chapter 2, the opening paragraph uses the word “grey” multiple times; instead of counting every separate circling of “grey” as one annotation, the whole set of grey imagery markings was counted as one annotation, as it represented the student marking one idea or concept within the text. Annotations were tallied and totaled by color category to generate a visual model of the quantity and quality of annotations by treatment and comparison condition. What follows is a synopsis of the findings for each level of annotation.

Surface-level annotations. At the baseline assessment, the average number of surface-level annotations for students in the treatment group was 5.20 annotations, whereas the average for students in the comparison group was 2.60 annotations. By the first phase of the instructional intervention, the difference between the two groups narrowed such that the average number of annotations for the treatment group was 3.60 compared to the comparison group’s 3.40. By the second phase of the instructional intervention, the average number of surface-level annotations for both groups diminished: the treatment group’s average was 2.00 and the comparison group’s average was 2.60. When the instructional intervention was complete, at the delayed posttest, the average number of surface-level annotations for the treatment group was 1.80, whereas the average for the comparison group was 2.00. Figure 8 illustrates these group means as a bar chart with the treatment and comparison groups side by side, where quantity of surface-level annotations are represented along the y-axis and the four time intervals are displayed along the x-axis.

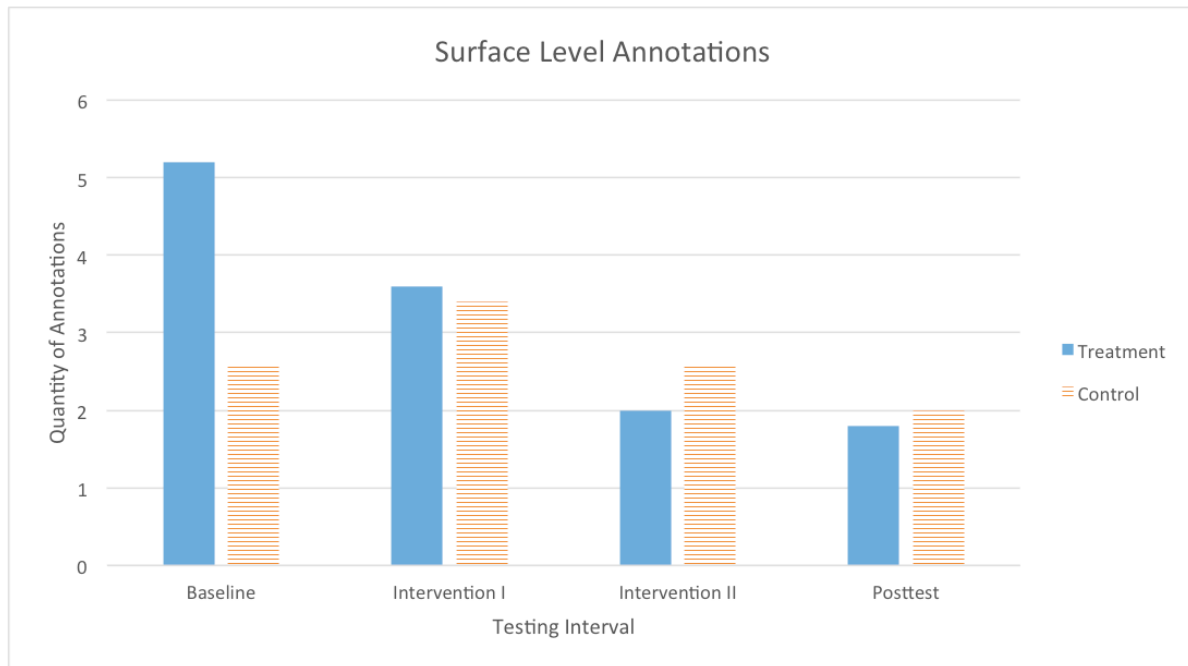


Figure 8. Bar chart of quantity of surface-level annotations at four time intervals for students in treatment and control groups.

In both groups, there was a downward trend in that students marked and underlined more plot points, reacted to more moments, and defined more vocabulary words in the beginning of the novel before they began studying the passages closely. By the fourth testing interval—five of nine chapters into the novel—these surface-level remarks had diminished. Although the treatment group exhibited a greater need to identify and respond to surface-level events in the first chapter, this need dwindled over the course of the study and the treatment group ended up with an average 3.4 fewer surface-level annotations in the fourth testing interval than in the first.

Mid-level annotations. At the baseline assessment, the average number of mid-level annotations for students in the treatment group was 4.80 annotations, whereas the average for students in the comparison group was 8.20 annotations. By the first phase of the instructional intervention, the difference between the two groups had expanded such that the average number of annotations for the treatment group was 15.60 compared to the comparison group's 9.00. By the second phase of the instructional intervention, this difference had widened again; the average number of mid-level annotations for the treatment group was 19.40 and the comparison group's average was 11.00. When the instructional intervention was complete, at the delayed posttest, the average number of mid-level annotations for the treatment group was 9.40, whereas the average for the comparison group was 5.40. Figure 9 illustrates these group means as a bar chart with the treatment and comparison groups side by side, where quantity of mid-level annotations are represented along the y-axis and the four time intervals are displayed along the x-axis.

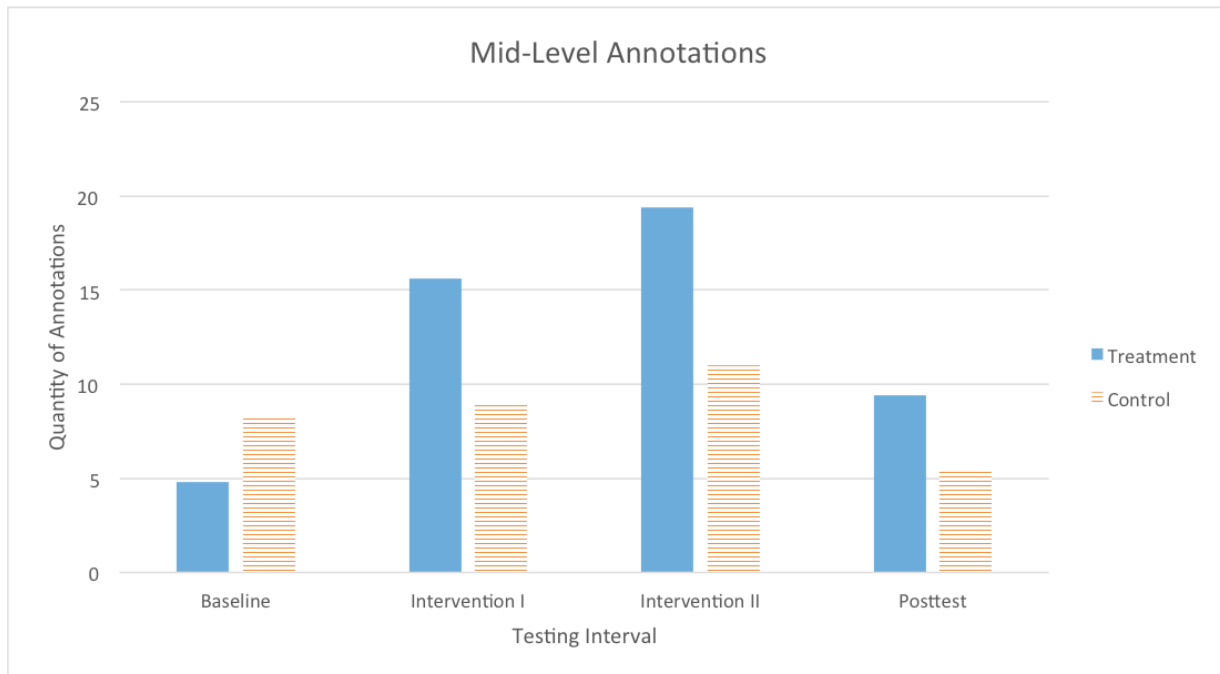


Figure 9. Bar chart of quantity of mid-level annotations at four time intervals for students in treatment and control groups.

The comparison condition follows a relatively predictable pattern of gradually increasing number of annotations from the baseline to the two phases of guided study, with a decline at the posttest. The treatment group follows this similar pattern, but the differences in scores from baseline to intervention to posttest are magnified. Where the comparison group's score shifts by averages of .80 annotations from the baseline to the first intervention and 2.80 annotations from the baseline to the second intervention, the treatment group's average leaps by 10.80 and 14.60 points respectively.

Deepest-level annotations. At the baseline assessment, the average number of deepest-level annotations for students in both the treatment group and the comparison group was 0.40 annotations. By the first phase of the instructional intervention, the difference between the two groups had expanded such that the average number of deepest-level annotations for the treatment group was 2.00 whereas the average for the comparison group was 1.00. By the second phase of the instructional intervention, this difference had narrowed, such that the average number of deepest-level annotations for the treatment group was 1.00 and the average number for the comparison group was .20. When the instructional intervention was complete, at the delayed posttest, the average number of deepest-level annotations for the treatment group was .80, whereas the average for the comparison group was 0. Figure 10 illustrates these group means as a bar chart with the treatment and comparison groups side by side, where quantity of deepest-level annotations are represented along the y-axis and the four time intervals are displayed along the x-axis.

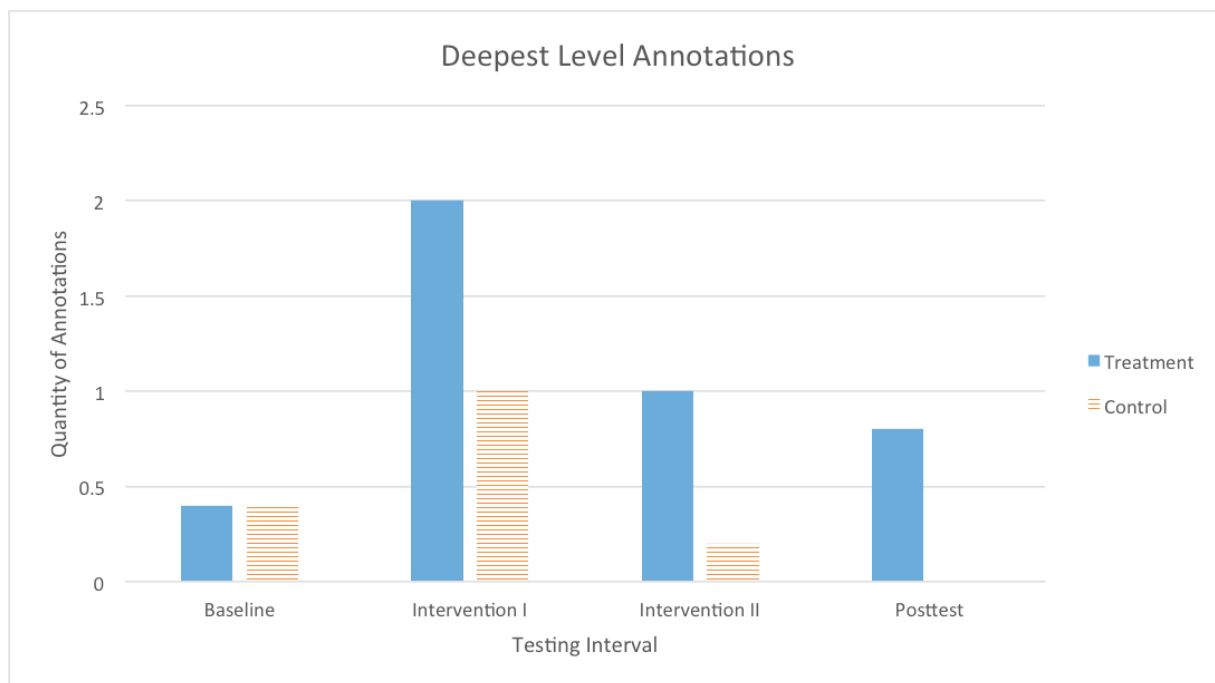


Figure 10. Bar chart of quantity of deepest-level annotations at four time intervals for students in treatment and control groups.

At the deepest level of annotation—discussing the broader implications of the passage on the meaning of the text as a whole—students across both groups struggled equally to convert observations about the subtext into inferences about greater meaning at the baseline assessment. While guided study seems to have helped both groups at the first phase of the intervention, students studying the multimedia think-aloud worked examples were able to make more annotations at the deepest level of analytical thinking than their counterparts answering text-based questions. Neither group was as successful in the second phase of the instructional intervention as they had been in the first. Only the students who studied the multimedia think-aloud worked examples made deepest-level annotations in the posttest.

Beyond the numerical representation of student annotations at each level of critical thinking, one additional pattern emerged for the two matched pairs who are most challenged by

analytical reading comprehension. The “low” and “medium low” pairs, all of whom scored at or below the 50th percentile in reading comprehension on the High School Placement Test (HSPT 2C, 2014), exhibited maximally different annotation patterns during the intervention phases such that these annotations bear further description.

At the baseline assessment, the four students who compose the bottom two pairs each relied heavily upon underlining plot. In fact, most had underlined nearly a third to half of the passage without remark. In addition to underlining, three of the four students boxed or circled colors, such as gold and rosy, and weather, such as warm and windy. The majority of the clarifying comments made at the baseline assessment were meant to help the reader get his or her bearings, such as “a fancy house” next to the text “their house was more elaborate than I expected” or “happy” next to the text “a cheerful red and white Georgian Colonial mansion overlooking the bay.”

By the first phase of the intervention, the relative number of surface-level comments compared to mid-level annotations began to reverse for struggling readers in both groups. Struggling readers in the comparison group continued to underline four or five lines at a stretch, but this time with the addition of circles around color imagery (“ashes”) and the occasional note in the margins, such as “grey everywhere” and “depressing.” Neither student in the comparison group was able to translate these surface-level observations or mid-level annotations beyond one deepest-level inference that had been directly prompted by a question in the video (“How is this setting different from East Egg, which we just visited in the end of Chapter 1? Why do you think Fitzgerald so closely juxtaposed those two settings?”). In contrast, for struggling readers in the treatment group, copious margin notes took the place of aimless underlining, with remarks such as “the whole town is grey, covered in ashes, everyone is very depressed, upset = dissatisfied

with their life” accompanying the numerous circles around “ashes” and “grey.” The two students in the comparison group spotted relatively more instances of grey imagery; where their counterparts in the comparison group identified four or five such examples, and identified only the ones that used the words “grey” and “ash,” students in the treatment group circled and highlighted additional related words that created the grey imagery pattern, such as “smoke,” “powdery air,” “impenetrable cloud,” “dimly,” “invisible,” and “dust.” And while both students were able to connect this passage to the previous chapter to discuss Fitzgerald’s use of juxtaposition, they also both commented on a passage that describes the eyes of Dr. T. J. Eckleburg, asking deepest-level questions about the connection between Eckleburg and the symbolic eyes of God, and the even more nefarious connection between the eyes of God and a forgotten billboard that sits in decay.

By the second phase of the intervention, this pattern persisted: the struggling readers in the treatment group were able to translate their surface and mid-level observations to deeper-level inferences, whereas struggling readers in the comparison condition were not. There continued to be a marked difference in the quality of the annotations between the groups, even within a level band such as “mid-level” annotations. For example, in the opening passage describing the food at Gatsby’s party, a passage that was littered with alliteration, the struggling readers in the comparison group simply underlined or circled these instances and wrote comments such as, “11 Cs and 12 Gs!” or “how much food *is* there?” In contrast, struggling readers in the treatment group also underlined and circled these words, but made remarks such as, “I think Fitzgerald is using illiteration [sic] here to make sure that we the audience understand how important/extraordinary his parties are” and “draws attention to the details to help us

understand how big + grand the party is.” The depth of analysis appears to be greater for struggling readers in the treatment group, in contrast to students in the comparison group.

Finally, by the posttest, when all the instructional scaffolds had been removed, struggling readers in the treatment group continued to make few, if any, surface-level annotations and numerous mid-level annotations throughout the passage (averaging ten), though they made decidedly fewer mid-level annotations than they had during the intervention (averaging 17 and 20 for Phase I and II respectively). All of their markings were supported by thoughtful comments, unlike their baseline assessments in which much of the passage was underlined and much of their underlining was unsupported. Both students wrote annotations that suggested they continued to make use of the close-reading strategies practiced in the multimedia think-aloud worked examples, such as visualizing and asking questions of the text. In the final passage, one of the struggling readers in the treatment group was able to move to the deepest level of analytical thinking, but the other student was not. In contrast, the students in the comparison condition reverted to underlining many lines at a time, making surface-level reactions (“he’s showing off?”) to the text, and highlighting only the color patterns that had been traced throughout the previous passages. Taken together, the annotations from the matched pairs at the bottom of the HSPT reading comprehension spectrum reveal a unique outcome for struggling students participating in the treatment group that was not present for those in the comparison condition, and that was much less dramatic a difference for the other three matched pairs.

Student Interviews

In addition to the visible record of the student interpretation process which was described through passage annotations, one-on-one interviews were conducted with every student who was randomly assigned to the treatment condition to better understand students’ cognitive and

metacognitive experience of engaging in studying literature with multimedia think-aloud worked examples. Prior to coding and analyzing these interview data, recordings were transcribed by a professional transcriptionist. The teacher-researcher then engaged in a preliminary reading of the transcripts to re-acquaint herself with students' remarks. Just as in the steps outlined for processing passage annotation data, she made initial notes about general observations, such the abundance of student comments about the difference between their own reading process and the process demonstrated in the think-aloud. She then conducted a more focused reading of the data, underlining in pencil key words and phrases in the transcripts. As concepts and descriptors were repeated, she began an initial list of possible codes. She re-read the transcripts a third time to review and amend the codes, and to merge codes that described the same phenomenon, such as "my process" and "on my own." In the final reading of the transcripts, she marked and highlighted words and passages that exhibited the revised codes, which are presented in Table 16.

Table 16

Student Interview In Vivo Coding System

Theme	In Vivo Code	Example
Student Process for Analytical Reading	"On my own"	"I'd reread it, try to see what stands out." "Try to break it down." "If I don't understand it, I sometimes just skip it."
Student Perception of Think-Aloud Process	"This Process"	"The videos ... really help like to get like ... like the inside of like your teachers like your mind and like how their ... how their process works." "Well, the think aloud first explains what's going on and then it explains how ... how it's like ... how the author is doing that. And then why the author does so to lead us to meaning ... yeah."
	"Practice"	"It was ... good to see someone else do it as a practice." "Like when I'm reading by myself I can do the same thing and it just seems like practice."

Effect of Think-Aloud on Analytical Reading Comprehension	“Thorough”	<p>“It was definitely a lot more thorough.”</p> <p>“I have a better understanding of what’s happening.”</p> <p>“It was a lot more in-depth than what I think I would have been able to do, at least, at this point.”</p>
	“Attention”	<p>“It was very helpful because there were certain words that I wouldn’t have paid attention to. Once the video is over, then I get to be reflective, but I did like that there were some things that would have otherwise missed.”</p> <p>“I think I definitely saw things that I wouldn’t have been able to figure out by myself.”</p>
Effect of Think-Aloud on Mental Effort	“Brain Power”	<p>“To a certain extent, I think ... once I’m watching [the] video, I kind of stop using my own brain power ... So I’m not ... I’m not really hunting down the words. I’m just thinking.”</p>
	“Easier”	<p>“I think when you did the think aloud, it was like easier for me to like understand.”</p> <p>“Like it just seemed a lot easier to pick up things with that kind of process ... that like step-by-step process ... Well, it was easier, obviously, and it was just kinda ... I just think it was more like laid out and more of like a actual process that you can like go through and like just study it and piece it together and dig a little deeper.”</p> <p>“Well, it was much easier than trying to figure it out on my own.”</p>
Effect of Think-Aloud on Persistence	“You were there”	<p>“It’s like you’re doing it with the teacher. And I thought that was really good.”</p> <p>“Because I would have not done that by myself honestly.”</p> <p>“It’s as if you were at home like helping us ... because we don’t have access to you at home.”</p> <p>“It wasn’t like me sitting at home alone like and kind of in silence. So it was nicer to have it like right there for me to hear and like focus on.”</p>
	“Focus”	<p>“Like instead of sending kids, or like when I go home and just read it alone, I just kind of read it, but if I had this while I was reading it ... it would help me to focus and help me to understand it a lot better.”</p> <p>“It keeps you on track instead of like getting distracted or just stuff like that.”</p>
	“Pause/Rewind”	<p>“There were some points where I rewinded it just to like re-see what you said and then so I could annotate and everything.”</p>

From these data, five central topics emerged:

1. Students' processes for reading and analyzing literature independently
2. Students' perceptions of the think-aloud process as different from their own
3. Students' descriptions of the effect of studying with the think-aloud on analytical reading comprehension
4. Students' descriptions of the effect of studying with the think-aloud on mental effort
5. Students' descriptions of the effect of studying with the think-aloud on persistence.

These topics are discussed in the coming pages, with direct quotations from the students. Note that pseudonyms are used for all students mentioned in the study, in order to protect the identity of the participants and to remain in compliance with IRB expectations around privacy. In parentheses after each student's name is that student's reading comprehension percentile score on the High School Placement Test (HSPT 2C, 2014), which serves to orient the reader in thinking about the relative utility of the multimedia think-aloud worked example based on a student's relative ease or struggle with analytical reading comprehension.

Students' processes for reading and analyzing literature independently. In each of the 18 interviews conducted, the first strategy students listed as core to their process for analytical reading was *re-reading*. One fairly typical sophomore, Jack (63), described his purpose for re-reading as this: "Like I read like a lot around like before and after it to kind of better understand the plot around it." Jack's experience of re-reading to better understand the plot was one that was reiterated time and again by his peers. It appeared that most students relied on re-reading as the key to navigating plot. Jack's classmate, Izzy (46), noted that she "reread[s] it ... until [she] understand[s] it ... Like the main thing, like what's happening." Importantly, the

re-reading strategy was entirely about understanding the surface-level events of the story line and identifying who was involved in that part of the story. It was about plot rather than analysis.

Beyond re-reading, many students discussed their search for literary devices, such as symbols, motifs, and imagery, as another strategy they readily used to analyze literature independently. Eric (71) explained his process: “Then I go through it a second time because I need to find like the key terms, and then I sort of like look at the overall topic and see if like these keys terms relate to the topic ... I usually go over it like more than once ... I try to relate it to like the plot.” Eric was not alone in reflecting on the hunt-and-peck strategy of independent literary analysis. Paige (94), a strong reader, reflected on how she approaches a difficult passage, concluding, “I guess I didn’t really have a process. Just kind of go through and see like if stuff would remind me [of] other things and like pick out kind of devices that I knew already.”

Commonalities aside, there appeared a stark contrast between the independent analytical reading process for strong readers compared with the process for struggling readers. One of the strongest readers in the class, Kelsey (95), described her process for deconstructing a challenging passage in the following way:

Well, first I like to read it all the way through just to make sure that I like know the context of everything that’s happening, and then I look for like the most obvious devices that I can pull out first. And then I like ... there are parts that I know are significant, but I’m not quite sure how they’re significant and so then I like try to use the context of everything else I figured out to put those into like a meaning ... I think about what it could show and ... especially in the context and like how it portrays a deeper meaning for what’s happening.

Kelsey's process is systematic and thorough; it assumes that there is more to the passage beneath the surface, and, importantly, that she will be able to understand that deeper-level meaning with enough thought and effort. On the other hand, when asked what she does when she arrives at a challenging passage in the text, Hailey (42) admitted, "At first it's probably like a mental panic, like I don't know what's even happening. So I probably read it like three times, so I can understand it. And then I try to guess the symbolism." Other struggling readers replied that they often keep moving through the text, despite knowing they need to repair meaning. Naiomi (50) explained, "I usually spend a lot of time on it, but if I don't understand it, I sometimes just skip it." While stronger readers such as Kelsey had a successful schema for breaking down a difficult passage, struggling readers such as Hailey and Naiomi did not; their methods depended on guessing or skipping the challenging passage altogether.

Description of the think-aloud process as different from their own. Student interviews consistently revealed that the analytical reading process modeled in the think-aloud videos was different from the process they naturally used when reading and analyzing independently. Most students were able to articulate the three-step process that they had practiced in the videos in some form: moving from the surface level "what" (What is happening?) to "how" (How is the scene being presented through literary devices and specific diction?) to "why" (Why has Fitzgerald chosen to present the setting in this way to convey some greater meaning about characters and their relationships to one another?). In his own words, Oscar (73) illustrated the three-step process: "The think-aloud first explains what's going on and then it explains how ... how it's like ... how the author is doing that. And then why the author does so to lead us to meaning."

About a quarter of students described the think-aloud as a form of reading “practice” in which they learned by watching how someone more expert than themselves would approach a challenging passage. Many of these students related watching the video to participating in sports practice and theater rehearsal. Hailey (42), an athlete, encouraged her teacher to make more videos because “It was good to see someone else do it as a practice!” Her classmate, Maddie (76), also an athlete, reiterated Hailey’s plea, noting that she liked the think-aloud videos because in the future, “when [she’s] reading by [her]self, [she] can do the same thing and it just seems like practice.” Eric (71) explained how this form of practice functioned for him by elaborating on his experience of the think-aloud process:

It actually pointed out a lot of ideas that I didn’t see like at first, and then when you can look at the ideas it sort of like points out other ideas. So like at the start of the video, I was like wondering ... I’m like how did you find that? And then at the end of the video, I was like looking at this, and then before you even like stated it, I would say like oh, this is gonna happen ... And you focused on key points, but you don’t fully tell ... like tell us at first like what the key points means. That’s really helpful because like then you can make inferences yourself on what the key points mean ... it’s like you’re the director that just like took us behind the scenes.

Eric was not the only one to describe what it was like to go inside the mind of the teacher. One student, Fay (50), offered this explanation: “I feel like this is like really, really useful ... just having a different perspective, like going through someone else’s mind and seeing how they learn and develop their own ideas.”

Mentioned by multiple students, the notion of “going through someone else’s mind” or being taken “behind the scenes” was a common thread in the interviews. Notably, Lori (23), a

student from a monolingual Spanish-speaking home, described her behind-the-scenes experience with the think-aloud in the following way:

I feel like it helps like ... like for me, I ... I'm not big on reading. I feel like it would help ... it would be like an advantage for like those kids that like just either like don't like to read or like have trouble like in the subject of English. I feel like those types of videos will really help like to get like ... like the inside of like your teachers like your mind and like how their ... how their process works and like ... versus like your process and like combining those two. So then like you have like a sense of like, okay, so my teacher's like looking for this. Now that's what I am going to look for. Because, when like I annotate, I just ... I really don't ... like to be honest, I really don't like question anything. At times I do, but not ... not often. So when you were questioning, it was just like, oh, I'm ... I'm missing a step here like when I read by myself. I'm just like, oh, I keep forgetting the question. So then like that kind of helped, especially like when annotating.

Lori highlights how the think-aloud helped her: it allowed her to see into the analytical reading process of someone who *is* “big on reading” and to integrate that thinking process with her own. One facet of the interviews not captured adequately by a transcript was the emotion in each student's voice. Though each and every one of the students interviewed noted that the videos were helpful and that they appreciated the support, the students who were the most effusive in their praise and expressed the most surprise at how the think-aloud process was different from their own reading experience, were the students who struggled most with analytical reading comprehension—students such as Lori, Fay, and Hailey.

The effect of the think-aloud videos on analytical reading comprehension. Students at all levels of reading skill explained that studying the passage with the support of the multimedia

think-aloud worked examples increased their attention to detail and, as a result, the depth of their comprehension. Adam (87) acknowledged that the video was “very helpful, because there were certain words that [he] wouldn’t have paid attention to. Once the video was over, then [he got] to be reflective ... but there were some things that [he] would have otherwise missed.” Adam’s classmate, Fay (50), lends support to his remarks, noting, “I think more light bulbs clicked and more ideas were sparked when you had brought up different topics that I obviously wouldn’t have discovered on my own.” The light bulbs clicked for Brian (92), too, who described the videos in a joking tone, suggesting that the videos were helpful “cuz like when I read I don’t even look for literary devices that often, but like going through the videos it helped me like understand that they’re like *everywhere* and to *look out for them*.”

In addition to drawing their attention to elements of the subtext that they may have glossed over during independent reading, students reported thinking more deeply and critically as a result of watching the think-alouds. Quinn (68) highlights the connection between these two behaviors:

I thought that it was nice to have ... it was a nice guidance. And there are a lot of things that I wouldn’t have found. It’s just ... it was a lot more in-depth than what I think I would have been able to do, at least, at this point. Like looking through—if I had all my papers out like with all the literary things, then I feel like I maybe could have found that stuff. But off the top of my head, I don’t think I could have.

The think-aloud made deeper-level thinking more accessible and achievable for Quinn. As a fairly solid reader, Quinn exhibited confidence that he might have been able to make the types of observations presented in the video, but he knew that it would have required him to use all of his resources and do a great deal of searching. The think-aloud made “in-depth” a more realistic

goal. In addition to having their attention drawn to important details during the instructional intervention, some students noticed a change in their attention to detail after the intervention. Izzy (46) observed changes in her independent reading process after the intervention ended: “I feel like when I was like reading the chapters just like later I felt like I was like annotating more. Like I was actually like looking out for stuff because I knew what to look out for.”

Moreover, students reported that the think-alouds pushed their own analytical thinking and made them more thorough readers. Some students made claims about having “a better understanding of what’s happening” in the passage (Maddie, 76), while others explained that the think-aloud enhanced their existing understanding by making them “more thorough” (Kelsey, 95; Camille, 97). For a few of the stronger readers in the study, the process appeared to be a nice supplement, though not a necessary one. Camille (97) described her experience of studying the passage with the think-aloud in the following way:

It was definitely a lot more thorough. I like tend to not connect the individual passages to entire themes or to other events in the book. Like I just usually look at it as its own thing, and I don’t really like think about it as critically. And the thought process was like more ... well, I guess like ... more thorough and more slow and you would kind of take your time more and try to think, oh, what does this mean like and try to connect it back to the themes more and just relate it to other pass ... or not passages necessarily, but just other things that had happened in the book and like connect the ideas to each other more so it formed like a more cohesive story and it just made more sense in the grand scheme of things.

Another strong reader, Paige (94), similarly explained that studying with the think-aloud was beneficial to her own thinking process:

It kind of expanded like my own thinking like made me think, oh, there could be a lot of stuff that I missed or that I couldn't find. And now I go into other books, I would have new literary devices or just like broader thinking I guess ... It made me a lot more aware of like things going on like underneath the surface instead of just reading it to understand what's happening, kind of me thinking more like, oh, there could be a deeper message or like the author could be specifically highlighting stuff. And I never really thought of that. I just kind of read it to understand what's happening.

Both Camille and Paige expressed being able to think more deeply and more critically about the text because the think-aloud process guided them to move “underneath the surface” (Paige) and to build a “more cohesive story” (Camille).

The effect of the think-aloud videos on mental effort. Student interviews also revealed that studying with the support of a think-aloud video made analytical reading easier. In fact, seven students commented on the relative ease of the process compared with their own. One of these students, Dan (93), suggested that the videos made it “a lot easier to pick up on things,” and he appreciated the “step-by-step process.” In reflecting on the two intervention videos, he explained, “I just think it was more like laid out and more of like a actual process that you can like go through and like just study it and piece it together and dig a little deeper.” Dan's classmate, Adam (87), attributed the ease of studying with the think-aloud to the type of thinking in which he got to engage, compared to the kind to which he is accustomed: “To a certain extent, I think ... I kind of stop using my own brain power and thought, okay, it's been done for me. So I'm not as ... I'm not really hunting down the words. I'm just thinking.” The shift from “hunting” to “thinking” in the think-aloud lightened the “brain power” load for Adam, making it easier for him to focus on deeper-level analysis rather than surface-level searching. Similarly,

Kelsey (95) noted how the think-aloud helped her to bypass the hunting phase of literary analysis and to move directly to deeper-level thinking:

I think I was trying to take what you saw and turn it into like the deeper meaning part. So I was able to just jump like straight to that, instead of like first looking through all the devices and then trying to piece together the meaning ... Because it was, was someone else like laying out what the literary device is, where ... and like kind of how they ... like why they might have been put in that way. I was able to more clearly see the meaning and like why they were used.

Kelsey attributed her ability to jump straight to deeper meaning to the think-aloud, because it helped her to see all the devices without having to find them and lay them out for herself.

According to Kelsey, the surface-level work was done; the only work left to be done was the deepest level of critical thinking, and that was much easier to see.

The effect of the think-aloud videos on persistence. Of particular interest to the teacher-researcher was the students' description of *how* the think-aloud videos helped them achieve a deeper understanding of the text. One pattern that emerged was the common experience of accompaniment. The data from the interviews suggest that a number of students persisted in deconstructing the passage simply because they felt they were not alone. Their language hinted at this pattern early on when one student identified the different phases of the study to his teacher as "when you weren't there" and "when you were there" (Reid, 95). When asked to describe what it was like working with the video as she studied, Naiomi (50) explained:

I think it would have taken longer to analyze the text [on my own]. I don't know. I get confused a lot, like hard to focus when I'm reading. Just because like ... it's as if you were at home like helping us ... because we don't have access to you at home.

Students referred to the think-aloud video and the teacher interchangeably, and they referred to working with the video as similar to working with another person. Lori (23) realized that the video pushed her to think more critically and to spend more time on the passage: “I would have not done that by myself, honestly.” Reid (95) expressed appreciation for having “someone helping you out the whole time.” Camille (97) agreed, adding, “I liked how every step of the way was laid out for you in the video. It’s like you’re doing it with the teacher. And I thought that was really good.” Students underscored that having a companion in the analytical reading process helped them to persist in the face of a challenging text.

Furthermore, students emphasized how studying with the think-aloud pushed them to think more deeply about the text than they otherwise would have. Brian (92) illustrated this difference by describing his process with and without the video, acknowledging the effect of the Internet on his willingness to stick with a challenging passage:

So first I’ll read it, and then if I don’t like understand I’ll read it again, and then if I still don’t get it I’ll probably look up the passage online. Like I’ll just Google like the start of the passage and then see what others have written about it. But [the video] took me through like your thought process as the teacher, and then like in between like you would present an idea, and then like it would stop and then say okay, this is how I got there. Like do [you] think you could do this, too? Like jot down some notes on like this, this, and this and like answer some of these questions that’ll help you get to this point. And I thought that was really helpful because like I can use those on my own instead of like immediately going on Google or looking it up, I can use those questions to like help me better understand the passage.

For Brian, the think-aloud provided an alternative to Google that supported his thinking process and gave him strategies for future passages. For other students, the think-aloud simply helped them to stay focused. Eric (71) noted that the video kept him “on track instead of getting distracted.” Paige (94) confirmed and elaborated on Eric’s experience:

I think it was very like focusing because you just have like your voice and it wasn’t like me sitting at home alone like and kind of in silence. So it was nicer to have it like right there for me to hear and like focus it on ... Like instead of sending kids, or like when I go home and just read it alone, I just kind of read it, but if I had this while I was reading it ... it would help me to focus and help me to understand it a lot better.

It appears that for some students, having the teacher “right there” to guide analytical thinking was about more than companionship; it was about helping them stay tuned in to the assignment. It is perhaps for this reason of distractibility that a full third of participants mentioned liking their ability to pause and rewind the video, and to remain in control of their learning environment. It is worth noting, however, that not every student shared this experience. Two of the strongest readers in the class, Reid (95) and Kelsey (95), expressed that the videos slowed down their natural analytical reading pace, causing one of them to “space out” and the other to “get distracted.”

Ancillary Results

Details from the interviews conducted with students—particularly with students identified as struggling readers—encouraged the teacher-researcher to look more closely at the available data to highlight patterns in analytical reading comprehension scores that did not stand out in the overarching analysis presented for research question one. What follows is a series of ancillary results which (a) deconstruct the analytical reading comprehension score into its three

component parts from the rubric, including textual observations, inferences, and broader implications of the passage; and (b) tease out the results for struggling readers to compare analytical reading comprehension scores between the treatment and comparison conditions.

Analytical Reading Comprehension Sub-Scales

In addition to analyzing total analytical reading comprehension scores, a series of independent samples *t* tests were conducted for each of the reading comprehension rubric sub-scales between treatment and comparison groups at each of the four testing periods. One of these sub-scales was textual observations, or the student’s ability to provide relevant textual references from the passage to support their inferences. The rubric used to measure textual observations had five possible points. Scores ranged from one to five. Table 17 presents descriptive statistics, independent-samples *t* test results, and effect sizes (Cohen’s *d*) for the treatment and comparison conditions on the dependent variable of analytical reading comprehension at the level of textual observations along four testing intervals.

Table 17

Textual Observations Sub-Scale Results in Analytical Reading Comprehension: Descriptive Statistics, Independent-Samples t Test Results, and Effect Size at Four Testing Intervals

	Treatment (<i>n</i> = 18)		Comparison (<i>n</i> = 16)		<i>t</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Baseline	2.72	.58	2.69	.79	.15	.05
Intervention I	3.22	.94	2.56	.89	2.09*	.72
Intervention II	3.00	.97	2.56	.81	1.41	.49
Delayed Posttest	3.56	.86	3.06	.85	1.68	.58

* Difference in means is significant at the .05 level (2-tailed).

The treatment group (*M* = 2.72, *SD* = .58) and comparison group (*M* = 2.69, *SD* = .79) began the study with relatively equal means (*d* = .05). After the first phase of the instructional intervention,

the treatment group exhibited, on average, statistically significantly higher scores in textual observations ($M = 3.22$, $SD = .94$) than the comparison group ($M = 2.56$, $SD = .89$), $t(32) = 2.09$, $p < .05$. The effect size ($d = .72$) was moderately large, favoring the multimedia think-aloud worked example treatment condition over the traditional text-based questions comparison condition. After the second phase of the instructional intervention, the treatment group ($M = 3.00$, $SD = .97$) outperformed the comparison group ($M = 2.56$, $SD = .81$) in textual observations, but this difference was not statistically significant. The effect size ($d = .49$) was moderate, favoring the multimedia think-aloud worked example treatment condition over the traditional text-based questions comparison condition. At the delayed posttest, after the conclusion of the instructional intervention, the treatment group ($M = 3.56$, $SD = .86$) again performed higher than the comparison group ($M = 3.06$, $SD = .85$) in textual observations, but this difference was not statistically significant. The effect size ($d = .58$) was moderate, favoring the multimedia think-aloud worked example treatment condition over the traditional text-based questions comparison condition.

Next, an independent-samples t test was conducted to describe the difference in mean scores in inferences, under the larger umbrella of analytical reading comprehension. This subscale measured the extent to which the student's analysis demonstrated an ability to read and interpret a complex text by moving beyond the obvious to consider the implications of subtext. The rubric used to measure inferences also had five possible points. Scores ranged from one to five. Table 18 presents descriptive statistics, independent-samples t test results, and effect sizes (Cohen's d) for the treatment and comparison conditions on the dependent variable of analytical reading comprehension at the level of inferences along four testing intervals.

Table 18

*Inferences Sub-Scale Results in Analytical Reading Comprehension: Descriptive Statistics, Independent-Samples *t* Test Results, and Effect Size at Four Testing Intervals*

	Treatment (<i>n</i> = 18)		Comparison (<i>n</i> = 16)		<i>t</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Baseline	2.83	.71	2.69	.60	.64	.22
Intervention I	3.22	.88	2.69	.95	1.71	.59
Intervention II	3.33	1.03	2.88	.81	1.43	.50
Delayed Posttest	3.22	.88	2.81	.83	1.39	.48

* Difference in means is significant at the .05 level (2-tailed).

The treatment group ($M = 2.83$, $SD = .71$) and comparison group ($M = 2.69$, $SD = .60$) began the study with similar means ($d = .22$). After the first phase of the instructional intervention, the treatment group ($M = 3.22$, $SD = .88$) performed higher than the comparison group ($M = 2.69$, $SD = .95$) in inferences, but this difference was not statistically significant. The effect size ($d = .59$) was moderate, favoring the multimedia think-aloud worked example treatment condition over the traditional text-based questions comparison condition. After the second phase of the instructional intervention, the treatment group ($M = 3.33$, $SD = 1.03$) continued to outperform the comparison group ($M = 2.88$, $SD = .81$) in inferences, but this difference was not statistically significant. The effect size ($d = .50$) was moderate, favoring the multimedia think-aloud worked example treatment condition over the traditional text-based questions comparison condition. At the delayed posttest, after the conclusion of the instructional intervention, the treatment group ($M = 3.22$, $SD = .88$) still scored higher than the comparison group ($M = 2.81$, $SD = .83$) in inferences, but this difference was not statistically significant. The effect size ($d = .48$) was moderate to small, favoring the multimedia think-aloud worked example treatment condition over the traditional text-based questions comparison condition.

Additionally, an independent samples *t* test was conducted to describe the difference in mean scores in students' understanding of the broader implications of the passage studied. This sub-scale measured the student's ability to situate the discussion of the selected passage within the greater thematic context of the text as a whole. The rubric used to measure broader implications also had five possible points. Scores ranged from one to five. Table 19 presents descriptive statistics, independent-samples *t* test results, and effect sizes (Cohen's *d*) for the treatment and comparison conditions on the dependent variable of analytical reading comprehension at the level of broader implications along four testing intervals.

Table 19

*Broader Implications Sub-Scale Results in Analytical Reading Comprehension: Descriptive Statistics, Independent-Samples *t* Test Results, and Effect Size at Four Testing Intervals*

	Treatment (<i>n</i> = 18)		Comparison (<i>n</i> = 16)		<i>t</i>	<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>		
Baseline	2.22	.55	2.13	.62	.49	.17
Intervention I	2.78	1.00	2.00	1.16	2.10*	.72
Intervention II	2.89	1.08	2.50	.63	1.26	.45
Delayed Posttest	2.83	.71	2.31	.60	2.30*	.80

* Difference in means is significant at the .05 level (2-tailed).

The treatment group ($M = 2.22$, $SD = .55$) and comparison group ($M = 2.13$, $SD = .62$) began the study with relatively equal means ($d = .17$). After the first phase of the instructional intervention, the treatment group exhibited statistically significantly higher mean scores in broader implications ($M = 2.78$, $SD = 1.00$) than the comparison group ($M = 2.00$, $SD = 1.16$), $t(32) = 2.10$, $p < .05$. The effect size ($d = .72$) was moderately large, favoring the multimedia think-aloud worked example treatment condition over the traditional text-based questions comparison condition. After the second phase of the instructional intervention, the treatment group ($M =$

2.89, $SD = 1.08$) performed higher than the comparison group ($M = 2.50$, $SD = .63$) in broader implications, but this difference was not statistically significant. The effect size ($d = .45$) was small-to-moderate, favoring the multimedia think-aloud worked example treatment condition over the traditional text-based questions comparison condition. At the delayed posttest, after the conclusion of the instructional intervention, the treatment group exhibited, on average, statistically significantly higher scores in broader implications ($M = 2.83$, $SD = .71$) than the comparison group ($M = 2.31$, $SD = .60$), $t(32) = 2.30$, $p < .05$. The effect size ($d = .80$) was large, favoring the multimedia think-aloud worked example treatment condition over the traditional text-based questions comparison condition.

Struggling Readers

Of particular interest to the teacher-researcher was the effect of multimedia think-aloud worked examples on students who generally struggle with analytical reading comprehension. Prior research indicated that these students should be particularly well served by the intervention, but the sample size available for this study did not lend itself to more robust statistical tests to compare students by reading ability. However, the combination of descriptive statistics and the qualitative data reported in research question three shines a light on the experience of struggling readers. In this case, “struggling readers” are defined as students scoring at or below the 50th percentile in reading comprehension on the national standardized High School Placement Test (HSPT 2C, 2014). The rubric used to measure analytical reading comprehension had fifteen total possible points. Scores for struggling readers ranged from four to eleven. Table 20 presents descriptive statistics for struggling readers in the treatment and comparison conditions on the dependent variable of analytical reading comprehension at four testing intervals.

Table 20

Analytical Reading Comprehension Scores for Struggling Readers: Descriptive Statistics at Four Testing Intervals

	Treatment (<i>n</i> = 5)		Comparison (<i>n</i> = 4)		<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Baseline	6.60	1.67	7.00	.82	.32
Intervention Phase I	7.60	1.67	5.75	1.50	1.17
Intervention Phase II	8.20	2.39	6.50	.58	1.14
Delayed Posttest	8.00	1.58	7.00	1.41	.67

Note. Struggling readers are defined as those scoring at the 50th percentile or lower in reading comprehension on the National Standardized High School Placement Test (HSPT)

There were five struggling readers in the treatment group and four struggling readers in the comparison group. At the baseline assessment, the mean score for struggling readers in the treatment group ($M = 6.60$, $SD = 1.67$) was .40 points below that of the comparison group ($M = 7.00$, $SD = .82$). The effect size for this difference was small ($d = .32$). After the first phase of the instructional intervention, the mean score for struggling readers in the treatment group ($M = 7.60$, $SD = 1.67$) was 1.85 points above that of the comparison group ($M = 5.75$, $SD = .82$), yielding a very large effect size ($d = 1.17$). After the second phase of the instructional intervention, the mean score for struggling readers in the treatment group ($M = 8.20$, $SD = 2.39$) was 1.70 points above that of the comparison group ($M = 6.50$, $SD = .58$), also a very large effect size ($d = 1.14$) favoring the treatment group. When the instructional intervention was complete, at the delayed posttest, the mean score for struggling readers in the treatment group ($M = 8.00$, $SD = 1.58$) was one point above that of the comparison group ($M = 7.00$, $SD = 1.41$). The effect size at the delayed posttest was moderate ($d = .67$). Figure 11 illustrates these group means as a line chart of analytical reading comprehensions scores at the four time intervals for struggling readers in the treatment and the comparison condition.

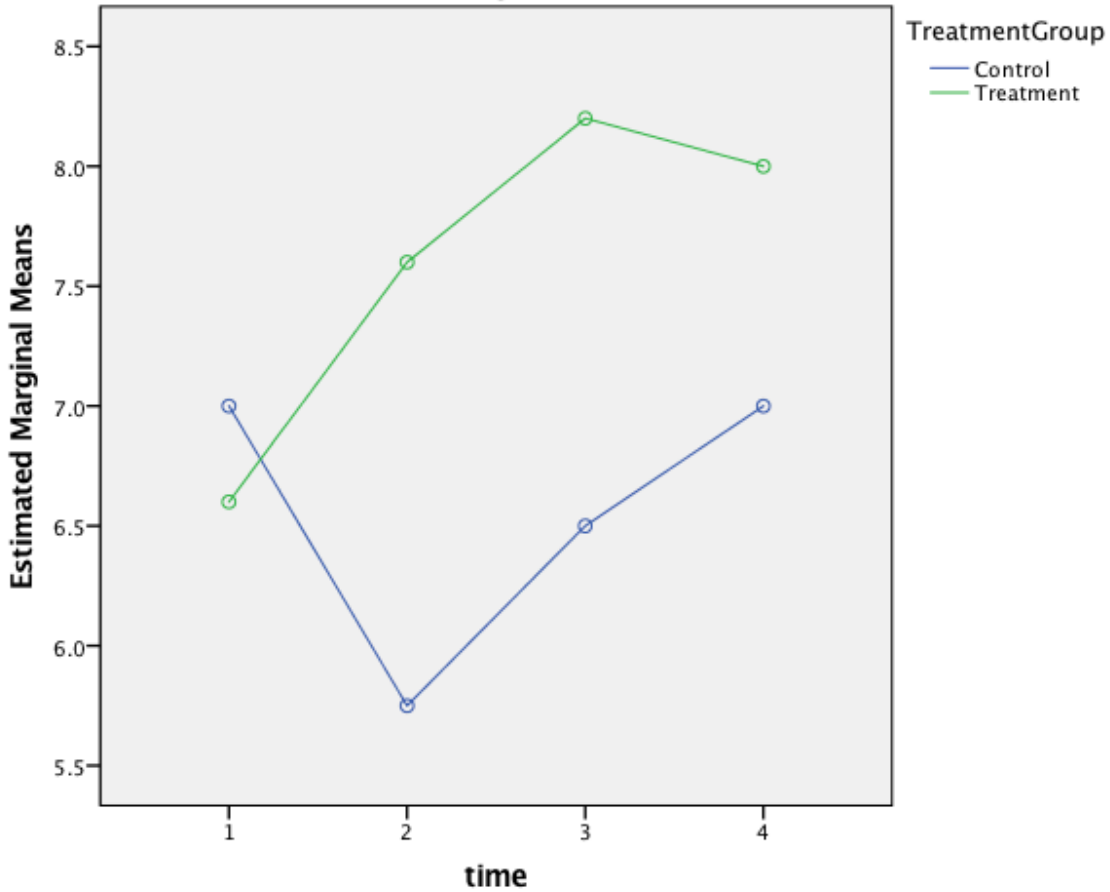


Figure 11. Line graph of analytical reading comprehension scores at four time intervals for students scoring at the 50th percentile or lower on the national standardized High School Placement Test (HSPT).

Figure 11 reveals a difference in mean scores for struggling readers, with the treatment group scoring higher than the comparison group at each interval of testing except for the baseline. The line graph illustrates the pattern of scores for both groups: scores for struggling readers in the treatment group (the top line) increased steadily during the two phases of the instructional intervention and dipped at the delayed posttest. In contrast, scores for struggling readers in the comparison group (the bottom line) dropped between the baseline and the first phase of the intervention, and then rose in the second phase of the intervention and the posttest, though they remained below the scores of the treatment group at the end of the study.

In order to further describe the scores in the treatment and comparison conditions, gain scores were calculated by subtracting each struggling reader’s baseline score from his or her subsequent scores. Table 21 presents descriptive statistics of the gain scores for struggling readers in the treatment and comparison conditions on the dependent variable of analytical reading comprehension at three testing intervals.

Table 21

Analytical Reading Comprehension Gain Scores for Struggling Readers: Descriptive Statistics at Three Testing Intervals

	Treatment (<i>n</i> = 5)		Comparison (<i>n</i> = 4)		<i>d</i>
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	
Gain: Intv. I - Baseline	1.00	1.41	-1.25	2.21	1.24
Gain: Intv. II - Baseline	1.60	1.82	-.50	1.00	1.49
Gain: Posttest - Baseline	1.40	1.34	.00	1.41	1.02

Note. Struggling readers are defined as those at the 50th percentile or lower in reading comprehension on the National Standardized High School Placement Test (HSPT).

The average gain from the baseline to the first phase of the intervention for struggling readers in the treatment group was one point ($M = 1$, $SD = 1.41$), whereas on average, struggling readers in the comparison group lost 1.25 points ($M = -1.25$, $SD = 2.21$). The effect size was very large ($d = 1.24$), favoring the multimedia worked example think-aloud treatment condition over the traditional text-based questions comparison condition. During the second round of the intervention, the average gain from the baseline for struggling readers in the treatment group was 1.60 points ($M = 1.60$, $SD = 1.82$), whereas on average, struggling readers in the comparison group lost half a point ($M = -.50$, $SD = 1.00$). Here the effect size was also very large ($d = 1.49$). At the delayed posttest, the average gain from the baseline for struggling readers in the treatment group was 1.40 points ($M = 1.40$, $SD = 1.34$), whereas on average, struggling readers in the

comparison group experienced no gain from the baseline ($M = .00$, $SD = 1.41$). The effect size for this final measurement was also very large ($d = 1.02$).

Summary of Results

The purpose of this study was to test the worked examples principle on the ill-defined problem of analytical reading comprehension in the naturalistic setting of a high school English language arts classroom using the practitioner model of the think-aloud as guidance. The study considered the effect of multimedia think-aloud worked examples on analytical reading comprehension and mental effort, as well as on the student experience of studying complex passages from literary texts.

In the quantitative strand of the study—research questions one and two—statistically significant differences were found between the treatment and comparison conditions for the analytical reading comprehension dependent variable at the first phase of the intervention and at the delayed posttest. These results were confirmed by statistically significant gain score results at the first gain (intervention I minus baseline) and at the third gain (delayed posttest minus baseline). No statistically significant results were found for the mental effort dependent variable.

In the qualitative strand of the study—research question three—coded passage annotations revealed that students in the treatment condition included, on average, more mid-level and deepest-level annotations on their passages for the two intervention phases than students in the comparison condition. Furthermore, students in the treatment condition included a higher number of high-quality comments than students in the comparison condition. One-on-one interviews highlighted five central themes: (a) students' processes for reading and analyzing literature independently hinge on re-reading, identifying plot, and searching for literary devices; (b) students described the think-aloud process as being different from their own because it

allowed them to see inside the mind of a more skilled reader; (c) students expressed that think-aloud videos increased their attention to detail and the depth of their analytical reading comprehension; (d) students reported that studying with the support of a think-aloud made analytical reading easier; and (e) the think-aloud videos increased their willingness to persist because of their experience of accompaniment, increased focus, and increased control.

CHAPTER V

DISCUSSION OF RESULTS

This study investigated how multimedia worked examples that explicitly model the reading habits of successful readers through teacher think-alouds could facilitate effective differentiated analytical reading instruction for high school English language arts students who have access to 1:1 technology. This study extended previous think-aloud research by (a) taking the elementary classroom practice of the teacher-facilitated think-aloud for early literacy instruction and placing it in the high school classroom setting as a scaffold for analytical reading comprehension, and (b) applying empirical, quantitative methodology to a classroom practice that has been studied and reported predominantly through professional development texts and practitioner journals, or as a measurement protocol rather than an instructional method (Kucan & Beck, 1997). Moreover, it extended previous worked examples research by (a) situating the quasi-experiment in the naturalistic setting of a high school English language arts classroom instead of the traditional controlled laboratory setting, and (b) applying the model of worked examples to the ill-defined task of analytical reading to determine if worked examples of analytical reading effectively support differentiated instruction in English language arts. This closing chapter summarizes the study, identifies its limitations, discusses key findings, draws conclusions, and presents implications for research and practice.

Summary of the Study

Independent analytical reading comprehension is a core competency required for success inside the classroom and out, and is one of the best predictors of college readiness, not to mention long-term success in college or career (Biancarosa & Snow, 2006; Gallagher, 2009;

Lattimer, 2014). Beyond basic comprehension of vocabulary and plot, analytical reading comprehension is the process of thinking by which students extract and construct meaning from a text (RAND, 2002), actively building an understanding of the subtextual elements of literature to include theme and the development of thematic ideas; character development, motivation, and conflict; denotative, connotative, and figurative meaning of language; structure, pacing, and organization of a story; and an author's use of language in relation to tone and a sense of time and place (Common Core State Standards Initiative, 2012).

Despite its critical importance, many secondary students struggle with this skill, and the national statistics around proficiency in analytical reading comprehension are grim. When placed in heterogeneous classes and a one-size-fits-all curriculum, struggling readers lose access to valuable instruction and invariably fall behind (Tomlinson, 2014; Tovani, 2000). Fortunately, as classrooms have become wired and connected spaces, educators have gained a transformative tool for reaching struggling readers. A technology-rich environment offers teachers a diversity of options for how to meet the needs of all learners and the potential to present a lesson in a manner that is both timely and targeted (Neebe & Roberts, 2015).

One such possibility for leveraging technology to support meaningful learning is the multimedia think-aloud worked example. This study drew upon research in each of these domains—multimedia, think-alouds, and worked examples—as the underpinning for its design and methodology. The think-aloud is an instructional strategy for supporting developing readers in which the teacher models the strategies of a successful analytical reader, such as activating prior knowledge, decoding text at multiple levels, making predictions, visualizing, summarizing, asking questions, making connections, and monitoring and clarifying understanding (Appleman, 2010; Wilhelm, 2001). The multimedia worked example (Renkl & Atkinson, 2010; Renkl, 2104)

is a well-researched instructional strategy, albeit only in other disciplines, that is statistically effective, multimedia-based, and parallels the practitioner process of the analytical reading think-aloud. A multimedia worked example presents an expert's step-by-step thought process to solving a complex problem for students to study and emulate (Atkinson et al., 2000).

The theoretical framework for this study had two pillars. The first was Sweller's (1988, 2010a) cognitive load theory, a psychological theory and set of instructional design principles based on an understanding of human cognition. Nested within that theory is the worked examples principle (Sweller & Cooper, 1985; Sweller, 2006; Renkl & Atkinson, 2010), which was the specific instructional method used in this study to reduce cognitive overload for students engaging in the complex task of analytical reading. The second pillar for this study was Mishra and Koehler's (2006) TPACK model, a conceptual model created to advance discussions around educational technology that describes a teacher's technological pedagogical content knowledge—his or her understanding of how and when to use a variety of technological tools to enhance teaching and improve learning within his or her content area.

This study addressed the essential quandary that most English language arts teachers face: how to provide an appropriate level of analytical reading instruction for all learners, how to support novice readers without inhibiting advanced readers, and how to challenge advanced readers without ignoring novice readers. Therefore the study aimed to apply a parallel solution to the problem at hand: to reimagine the face-to-face, teacher-facilitated think-aloud as a multimedia worked example that could be leveraged for differentiated, blended instruction.

In this experiment, an explanatory sequential mixed-methodology study, 34 sophomore English language arts students were randomly assigned to either the worked examples treatment condition or the traditional instruction comparison condition. During the two iterations of the

instructional intervention, students in the treatment condition watched a worked example video that presented a think-aloud deconstruction of a passage, whereas students in the comparison condition watched a video that presented a series of close reading text-based questions for the passage that led students to the same content and depth of analysis as the think-aloud. Using a classic treatment-comparison repeat measures pretest-posttest design, students' analytical reading comprehension and perceived mental effort was assessed. Later, in the qualitative phase of the study, the participant experience was described through interviews and annotations in order to more deeply understand the quantitative data collected. The study lasted three weeks during a six-week unit on *The Great Gatsby* (Fitzgerald, 1925). The sample came from an independent, college-preparatory high school in California's Silicon Valley, a hub of innovation in educational technology. The study considered three research questions:

1. Analytical Reading Comprehension. To what extent is there an effect of multimedia think-aloud worked examples on analytical reading comprehension in a secondary English language arts class, as measured by open-ended written responses to literature?
2. Mental Effort. To what extent is there an effect of multimedia think-aloud worked examples on cognitive load, as measured by self-reported mental effort (Paas, 1992)?
3. Student Experience. What do student annotations and interviews reveal about the participant experience of studying multimedia think-aloud worked examples?

Summary of Findings

The quantitative strand of data collection assessed students' analytical reading comprehension and perceived mental effort. The first research question investigated the difference in mean scores between the treatment and comparison condition on an assessment of analytical reading comprehension at four testing periods: the baseline assessment before any

intervention had occurred, the first phase of the intervention, the second phase of the intervention, and the delayed posttest after the intervention was completed. Statistically significant differences were found between the treatment and comparison conditions for the analytical reading comprehension dependent variable at the first phase of the intervention with a large effect size ($d = .82$) and at the delayed posttest with a moderately large effect size ($d = .73$). These results were confirmed by statistically significant gain score results at the first gain from the baseline to intervention one ($d = .82$) and at the third gain from the baseline to the delayed posttest ($d = .90$).

The second research question investigated the difference in mean scores between the treatment and comparison condition on a scale of perceived mental effort at the same four testing periods. No statistically significant results were found for the mental effort dependent variable at any of the testing intervals or for any of the mental effort sub-scales, such as task-relevant behavior, task-irrelevant behavior, or element interactivity. All effect sizes were .19 or lower, considered as having no effect.

The qualitative strand of data collection engaged students in a series of written annotations and open-ended interviews to supplement the findings from the first, quantitative strand of data collection. The third research question investigated what student annotations and interviews revealed about the participant experience of studying multimedia think-aloud worked examples. Coded passage annotations revealed that students in the treatment condition included, on average, more mid-level and deepest-level annotations on their passages for the two intervention phases than students in the comparison condition. Furthermore, students in the treatment condition included a higher number of high-quality comments than students in the comparison condition. One-on-one interviews highlighted five central themes: (a) students'

processes for reading and analyzing literature independently hinged on re-reading, identifying plot, and searching for literary devices; (b) students described the think-aloud process as being different from their own by allowing them to see inside the mind of a more skilled reader; (c) students expressed that think-aloud videos increased their attention to detail and the depth of their analytical reading comprehension; (d) students reported that studying with the support of a think-aloud made analytical reading easier; and (e) the think-aloud videos increased students' willingness to persist because of their experiences of accompaniment, increased focus, and increased control.

Limitations

This study breaks from the traditional participant demographics and settings used in previous think-aloud and multimedia worked examples research. Although the naturalistic high school English language arts setting is in many ways a unique attribute of this study, it does present a number of limitations. This section will examine the five central limitations of the present study, including sample size, the mental effort rating instrument, the analytical reading comprehension instrument, qualitative coding reliability, and participant bias.

First, the sample size for this study was small ($n = 34$), which limits the reliability of the data and the generalizability of the study. The treatment group ($n = 18$) and the comparison group ($n = 16$) were both well below the minimum ideal group size of 30, which suggests that the statistical tests conducted were not robust with respect to the normal distribution assumption because the sample size was not large enough for the central limit theorem to apply. It is for this reason that the teacher-researcher pursued a mixed-methodology design to augment the quantitative findings with qualitative data. Moreover, the small sample size restricted the types of statistical tests that could be conducted. Previous studies (Cooper & Sweller, 1987; Kalyuga,

2007; Kalyuga & Sweller, 2004; Nievelstein et al., 2013; Reisslein, Atkinson, Seeling, & Reisslein, 2006; Tuovinen & Sweller, 1999) made the case for analyzing the data based on students' levels of prior knowledge, which, in this study, would have meant breaking the treatment and comparison groups into sub-groups based on previous reading achievement. Ideally, a two-way repeated measures ANOVA would have been conducted to compare mean differences for a novice reader group and an advanced reader group in both the treatment and comparison conditions. However, the sample size simply was too small to justify such a test.

Second, the data from the mental effort rating scale were collected using self-reporting, which has inherent flaws. A careful analysis of the means and standard deviations for each of the sub-scales and testing intervals reveals a tight distribution of scores clustered around the middle of the seven-point scale. At all four testing intervals, standard deviations for both the treatment and the comparison groups were less than 1.00, ranging from .62 to .90. Even though students were reminded of the anonymity of their responses, it is conceivable that social desirability may have led students to misrepresent their mental effort or to deliberately select a neutral middle number. Furthermore, it is possible that young teenage students, not yet capable of the metacognitive move of evaluating their own mental effort, were led to select the middle score on the scale. It is also possible that the mental effort rating scale (Paas, 1992), which has been used successfully with well-defined problems, is not suited for measuring mental effort for multi-step, ill-defined tasks.

Third, it is a standard practice to assess high school students' analytical reading comprehension through their written responses to literature, since student compositions can convey more effectively the complexity of literature than over-simplified multiple-choice questions. This practice, however, is imperfect because it takes student writing ability into

consideration as part of the assessment of reading. Every effort was made to remove language from the rubric that judged student achievement based on quality of writing, and to limit rubric categories to those that specifically discussed quality of textual observations and depth of inferences. However, the fact remains that students' analytical reading comprehension was assessed in written form.

Fourth, as with many studies that are more naturalistic, there may have been unintended bias on the part of the teacher-researcher. This potential limitation is why (a) the teacher-researcher partnered with a co-researcher, (b) a learning management system was used to keep students' identities hidden during the study, and (c) the teacher-researcher enlisted the help of a second reader for the quantitative strand of the study, to score analytical reading comprehension responses. Had this been a completely qualitative study, a second reader would have coded, organized, and categorized themes for the annotation and interview data, and possibly could have identified additional patterns and offered insights that the teacher-researcher did not reach independently. Because the qualitative strand of the study served purely as a supplement to the quantitative strand of the study, multiple raters were not utilized.

Fifth, the intervention took place over a series of hour-long class periods rather than in a highly controlled research environment. Though it is atypical for students to discuss a text or class activities outside of the class period when those items have no bearing on their course grade, and though students were specifically asked not to talk about the study until the conclusion of data collection, it is feasible that students spoke to one another between test intervals to ask about each others' video experiences.

Discussion of Findings

The findings from this study of multimedia think-aloud worked examples enrich and extend previous research in literacy instruction and multimedia learning. This section situates the findings of the present study within the broader conversation around think-alouds and multimedia worked examples, and discusses those findings in light of established principles and acknowledged challenges within each domain: schema acquisition for adolescent literacy, worked examples with ill-defined problems, differentiation and the expertise reversal effect, literacy learning in the zone of proximal development, and measuring mental effort.

Schema Acquisition for Adolescent Literacy

Literature on teacher-facilitated think-alouds suggests that this instructional strategy works because it makes visible the thinking process that happens invisibly during reading, providing developing readers with a clear schematic model to imitate and assimilate (Beers, 2003; Davey, 1983; Olshavsky, 1977; Wilhelm, 2001). Researchers posit that think-alouds afford developing readers the opportunity to access a complex text without being overburdened by the cognitive load required to understand it independently, because the teacher assumes most of the cognitive load (Frey & Fisher, 2013). Despite the popularity of think-alouds in classroom instruction and the pervasive recommendation toward think-alouds in practitioner texts (Appleman, 2010; Beers, 2003; Burke, 2000; Keene & Zimmermann, 2007; Milner et al., 2011; Wilhelm, 2001), studies into think-aloud methodology and efficacy have centered on early literacy instruction during the “learning to read” years, rather than on adolescent literacy during the “reading to learn” years (Baumann et al., 1993; Beers, 2003; Davey, 1983; Olshavsky, 1977). Little is known in the research literature on think-alouds about the statistical significance of this strategy as it pertains to student cognition or reading comprehension in adolescent readers.

This study yielded statistically significant results at two testing intervals on the dependent variable of analytical reading comprehension for adolescent readers who were using multimedia think-aloud worked examples to study complex literary content. The effect size—a measure of practical importance—was large at the first phase of the instructional intervention and moderately large at the delayed posttest, suggesting that multimedia think-aloud worked examples are effective in supporting analytical reading comprehension, a finding that is consistent with the think-aloud literature. What is surprising in the results is the increase in mean score at the delayed posttest, when students are relatively unsupported in their analytical process. One possible explanation for this final increase is that students had acquired the schema needed to effectively deconstruct the selected passage and could apply what they had learned in the previous phases of the intervention. This is, perhaps, too charitable an explanation. More realistic is the notion that, having read five of the novel’s nine chapters, most students were no longer “novices” with the text at hand, and thus they no longer needed the scaffolding that would be provided for a developing reader with limited familiarity with a text. Naturally, the passage studied for the delayed posttest featured many of the characters and motifs students had considered throughout the previous four chapters, so they likely benefited from their own experience with the text. However, if the prime consideration of this study is the difference between the treatment and the comparison groups at the two phases of the instructional intervention, in which the effect size of the treatment was large and moderate respectively, then the results suggest that the multimedia think-aloud worked example is effective in supporting adolescent analytical reading comprehension, especially earlier in a text.

The qualitative data collected shine a light on why the intervention may have been effective by illustrating the interpretive moves that students were able to make as a result of

studying the multimedia think-aloud worked examples. Passage annotations provide a visible record of these interpretive moves, illustrating how students' reliance on surface-level annotations diminished during the two phases of the instructional intervention, whereas their emphasis on mid-level annotations increased. The quality of their observations and inferences similarly increased, as did their ability to translate these observations into deepest-level analytical commentary. Moreover, students reported in one-on-one interviews that the think-alouds helped them to “practice” the moves an expert reader makes by going “inside the mind” of the teacher. Think-aloud videos made them more attuned to textual details and literary devices that they may have otherwise missed, and guided them in drawing connections between these observations to better understand the “so what” of the passage. Together, annotation and interview data suggest that studying with the support of multimedia think-aloud worked examples helped push students to deeper levels of thinking than they were able to reach on their own or with traditional instruction.

This study extends previous think-aloud research by successfully taking the elementary classroom practice of the teacher-facilitated think-aloud for early literacy instruction and placing it in the high school classroom setting as a scaffold for analytical reading comprehension. The findings indicate that the think-aloud is of greatest utility early in the analytical reading process, or in early sections of a text, when prior knowledge is lowest and complexity remains high.

Worked Examples with Ill-Defined Problems

The majority of research on worked examples has focused predominantly on well-defined problems in fields such as mathematics, science, and mechanics (Atkinson et al., 2000; Renkl & Atkinson, 2010; Renkl, 2014; Sweller & Cooper, 1985). More recently, however, researchers have expanded their focus to include worked examples of ill-defined problems such as teaching

argumentation (Schworm & Renkl, 2007), identifying designers' styles (Rourke & Sweller, 2009), writing effective learning journals (Hübner et al., 2010), reasoning about legal cases (Nivelstein et al., 2013), and composing essays on English literature (Kyun et al., 2013). Each of these studies found a significant main effect for worked examples, highlighting how worked examples may be effective both in concrete disciplines such as math and science, as well as in more abstract disciplines such as design, literature, and the law. Despite these gains, no known studies apply multimedia worked examples to analytical reading comprehension or to triple-content worked examples (Renkl et al., 2009) in which students must contend with three domains: the learning domain (analytical reading), the exemplifying domain (the text being read), and the cognitive strategy domain (heuristic strategies and schema for approaching analytical reading).

This study found a statistically significant and practically important difference between the treatment and comparison conditions for the analytical reading comprehension dependent variable at the first phase of the intervention and at the delayed posttest, favoring the multimedia think-aloud worked example treatment condition over the traditional text-based questions comparison condition. In particular, students in the treatment group performed significantly better on measures of textual observations and broader implications than their counterparts in the comparison condition. Qualitative data support these quantitative results. Passage annotations for students in the treatment condition were of a higher caliber than for students in the comparison condition, and students in the treatment condition expressed that the think-aloud videos increased their attention to detail and the depth of their analytical reading comprehension, concepts that map onto the constructs of textual observations and broader implications. Collectively, these findings are consistent with previous studies of ill-defined problems and echo the claim that

worked examples, which traditionally identify and model discrete solution steps, are still effective in ill-defined domains where the problem is less clear-cut, the path to solution less linear, the rules less established, and the end result less definitive (Nievelein et al., 2013). In fact, in this study, it appears that the multimedia think-aloud worked examples equipped students to make a more complex analytical thinking move—jumping from textual observations to broader implications, or from “what” to “why”—than they were able to do independently.

This study offers a unique contribution to the small but growing field of ill-defined worked examples research in that it situates the quasi-experiment in a naturalistic setting—a high school English language arts classroom—instead of the traditional controlled laboratory setting, and applies the model of worked examples to a type of ill-defined task that had not been previously studied.

Differentiation and the Expertise Reversal Effect

Experts in multimedia worked examples with self-explanation assert that worked examples are most effective for novice students in the early stages of new skill development (Sweller, 1994; Renkl & Atkinson, 2010). Studies of well-defined problems illustrate that, whereas students with low levels of prior knowledge certainly benefit from teacher modeling, students with higher levels of prior knowledge may be hindered by the additional support (Plass et al., 2010), and often, the worked example effect disappears entirely for these advanced students (Kalyuga et al., 2003; Renkl & Atkinson, 2010). On the other hand, some studies of ill-defined problems (Nievelein et al., 2013) challenge the influence of this so-called expertise reversal effect and hint at the possibility that it may only affect those studying well-structured tasks (Schmidt et al., 2007). Despite this area of friction in the literature, research into cognitive

load theory and worked examples unequivocally advocates the importance of appropriately matching instructional scaffolds to individual student learning needs.

The findings in this study both confirm and conflict with prior research. Qualitative data from student interviews and annotations, coupled with quantitative data reported in the ancillary results section, confirm that multimedia think-aloud worked examples may be disproportionately beneficial for struggling students—in this case, students with lower prior analytical reading knowledge. Students from the two lowest matched pairs examined for the qualitative analysis of passage annotations exhibited substantially higher quality and quantity annotations at the mid- and deepest levels than their counterparts in the comparison condition—a difference that was less apparent at higher reading levels. Moreover, in one-on-one interviews, these struggling readers expressed that without the support of a think-aloud, their independent analytical reading methods depended on guessing or skipping challenging passages altogether. Descriptive statistics presented in the ancillary results section reveal a fairly dramatic difference in mean scores for students from the treatment group versus students from the comparison group, especially at the first round of the instructional intervention (Chapter 2 of the novel). These data describing the experience of struggling readers confirm prior research and indicate that students who are novices with the task at hand, or those who struggle with analytical reading comprehension, are particularly well served by multimedia worked examples, especially at the earliest stages of skill development or the initial chapters of the novel.

At the other end of the spectrum, the findings from this study offer a subtly mixed review of the utility of multimedia think-aloud worked examples on analytical reading comprehension for advanced readers. In keeping with the findings from other studies of worked examples with ill-defined problems (Nievelein et al., 2013), advanced students appear to have benefited from

the treatment, as evidenced by quantitative scores, reported in one-on-one interviews, and illustrated through matched-pair annotations. Students at the top of the HSPT spectrum described experiencing greater understanding of the text, greater awareness of its details and literary elements, and a clearer sense of the step-by-step process that leads the reader from the surface to the subtext. However, a small selection of student interviews also point to the possibility of the expertise reversal effect, which stands in stark relief to the Nievelstein et al. (2013) findings. Two of the strongest readers in the class—those with the strongest existing schema for analytical reading comprehension, or highest prior knowledge—described feeling slowed down and distracted by the think-aloud videos. One possible explanation is that they were experiencing the expertise reversal effect (Kalyuga et al., 2003; Kyun et al., 2013; Plass et al., 2010); that is, learning a new analytical reading schema, as presented in the video, may have interfered with their own existing schemas, forcing them to attend to unnecessary cognitive demands (Sweller, 2010a). In a sense, the qualitative data are at odds with themselves and do not paint a completely clear picture of the relative advantage of the intervention for stronger readers. Taken together, these early findings indicate that multimedia think-aloud worked examples bolster the analytical reading comprehension for all readers, and that they may—or may not—trigger the expertise reversal effect in advanced readers.

Literacy Learning in the Zone of Proximal Development

A wealth of practitioner literature suggests that think-alouds are one of the most effective instructional strategies for supporting developing readers (Keene & Zimmermann, 2007) and that without proper scaffolding and explicit guidance, these readers continue to have difficulty following basic fiction structures and identifying plot elements in novels and short stories (Griffin et al., 2008; Lewis & Ferretti, 2009; Rupley et al., 2009; Unsworth & McMillan, 2013).

Often the students in this category give up and quit reading altogether (Schoenbach et al., 2012). Moreover, reading research suggests that literacy learning happens in partnership with a more knowledgeable other. The literature on think-alouds reiterates this notion (Keene & Zimmermann, 2007), highlighting the connection to Vygotsky's (1930/1978) theory of cognitive modeling, wherein students learn by imitating the actions of the adults around them. Learning how to read analytically through a think-aloud is an example of a socially mediated process in which students who practice a skill that is beyond their actual and independent development level are able to move, with teacher guidance and instructional scaffolds, to their potential development level (Schoenbach et al., 2012). The think-aloud deliberately targets this "zone of proximal development" (Vygotsky, 1930/1978), and it is effective in large part because of the sense of partnership that students share with their teachers.

One of the concerns in moving from a live, in-class think-aloud to a recorded multimedia think-aloud worked example is that the feeling of a personal relationship would be lost. Despite the possible disadvantages of a live model for more advanced readers experiencing extraneous cognitive load, one of the advantages of a live model is that students can interact with the teacher and respond to his or her prompting and questions during the think-aloud. Among early critiques of the present study was the suggestion that students would be too passive while watching a recorded multimedia think-aloud worked example, and that the teacher-student relationship would not translate to the screen.

The qualitative data from this study suggest otherwise. The data are consistent with the literature on adolescent reading instruction, showing that the multimedia think-aloud worked example promotes literacy learning through a socially mediated process in which the student and teacher interact, albeit in different ways. Throughout the one-on-one interviews, students

expressed a sense of partnership with the teacher, feeling like the teacher was present to their experience of studying a challenging passage. They discussed what it was like to learn by listening to their teacher's process, going "behind the scenes" and following the lead of someone with more expertise whom they trusted. They noted that they were able to make leaps in their thinking that they would not have made independently, and that they valued the support of the teacher in video form. Moreover, multiple students described the experience of accompaniment—persisting with a complex passage simply because they felt they were not alone. Students explained that they did not give up, even in the face of a challenging text, because they had their teacher "right there" with them for "every step of the way." Together, these descriptions illustrate a process of reaching a level of potential development that is rooted in the student's connection to and relationship with a more knowledgeable other.

Furthermore, qualitative data reveal that even though students were studying with a pre-recorded video, they actively engaged with the think-aloud. It is possible that part of what made students feel like they could interact with the learning material was their ability to slow down or revisit the teacher's cognitive modeling by pausing or rewinding the video. Previous research into cognitive load theory and instructional design for complex learning asserts that students benefit from retaining control over the pace of the instruction they are processing (van Merriënboer & Kester, 2014). In one-on-one interviews, a third of the students mentioned appreciating that they could pause and rewind the video, thus remaining in control of their learning environment. Passage annotations for participants in the treatment group reveal that students were diligent in responding to self-explanation prompts and think-aloud guidance, suggesting that despite the pre-recorded nature of the multimedia think-aloud worked example, students were active learners rather than passive observers.

These findings indicate that multimedia think-aloud worked examples, in which the classroom teacher is the creator of the learning material, still may offer students the interactive relationship they need to learn in the zone of proximal development. It is unclear if the videos would have been as effective if the narrator had been a stranger to the students rather than their classroom teacher or some other trusted, more knowledgeable person.

Measuring Mental Effort

Worked examples research has yet to reach a consensus as to the most accurate way to measure mental effort. In the early stages of worked examples research, studies measured mental load through proxy measures such as time on task, or through outcome measures such as number of errors made (Sweller & Cooper, 1985; Cooper & Sweller, 1987). As worked examples research continued to develop, researchers sought a way to measure the efficacy of the intervention on reducing extraneous load through other means (Paas, 1992). Paas' scale has become the standard-bearer for measuring mental effort and has been employed in numerous experimental studies (Paas & van Merriënboer, 1993, 1994; Paas et al., 2003). However, Paas himself suggests that although the scale is highly reliable and serves as an effective starting point for mental effort research in well-defined domains, it is insensitive to the varying types of cognitive load and cannot yet distinguish between intrinsic load, extraneous load, and germane load (Brünken et al., 2010). Very few studies of mental effort in ill-defined domains have successfully employed the Paas scale (Nivelstein et al., 2013).

This study used an extended version of the Paas (1992) mental effort scale and embedded the Paas single-item scale within it as the first item. Quantitative findings from this study suggest difficulty in using a mental effort rating scale—single-item or multi-item—with adolescents studying worked examples in the ill-defined domain of analytical reading. As previously

mentioned in the limitations of the study, the means and standard deviations for each of the subscales and testing intervals were tightly distributed and clustered around the middle of the scale. Standard deviations for both the treatment and comparison groups at all four testing intervals were less than 1.00. One possible explanation for this phenomenon is that adolescent students are too young and not yet capable of the metacognition required to evaluate their own mental effort. However, there are examples of studies with adolescent students in which the Paas mental effort rating scale is successfully employed to evaluate mental effort in studying well-defined problems (Paas, 1992). Thus, another possible explanation is that it is more challenging to measure mental effort for ill-defined problems than for well-defined problems, and that a more appropriate scale for ill-defined problems has not yet been created.

However, the quantitative and qualitative findings from this study are at odds with each other. Where quantitative data reveal no advantage for the treatment group, qualitative data around mental effort suggest that studying multimedia think-aloud worked examples makes analytical reading “easier” and demands less “brain power.” These inconclusive mental effort findings are consistent with previous literature on measuring mental effort in that they reveal the real challenge of estimating mental processes and reiterate the call for further research to develop a more accurate instrument to measure students’ perception of mental effort after completing a multi-step task in an ill-defined domain.

Conclusions

This study set out to reimagine the face-to-face, teacher-facilitated think-aloud as a multimedia worked example that could be leveraged for differentiated, blended instruction. The findings suggest that there is great promise in applying the legacy of empirical learning around multimedia worked examples to practitioner’s think-alouds. This small study offers an early

indication that multimedia worked examples that explicitly model an expert reader's cognitive processes might be a high-leverage strategy for educators interested in differentiated analytical reading instruction for adolescents. Multimedia worked examples have a proven record of supporting students in processing complex information by reducing extraneous cognitive load, thus decreasing the total mental effort necessary to comprehend challenging content. Though the mental effort ratings for this study were inconclusive, qualitative data indicate that it is still possible that the reason the worked examples *worked* is because they optimized cognitive load by reducing the strain on the working memory and returning student focus to schema construction or germane load processing.

This study further suggests that the ill-defined task of analytical reading can be supported through multimedia worked examples, which adds to the limited research base around worked examples research of ill-defined problems. In keeping with the majority of worked examples studies, the findings from this study offer a slight indication that the scaffolding offered in worked examples should be distributed only to those students with lower prior knowledge who demonstrate a need for support rather than to all students in the class. Finally, the qualitative findings from this study suggest that multimedia think-aloud worked examples can be an effective means of targeting individualized instruction to the zone of proximal development without the cost of whole-class time.

Implications for Research

For this first known study of multimedia worked examples with adolescent analytical reading, the results uncovered more questions than answers. The findings from this nascent study represent an initial foray into the field of multimedia worked examples research using the practitioner's think-aloud model to guide literacy instruction, and hopefully it will lead to future

studies that confirm and challenge the results presented in this dissertation. There are four central areas for further research, which are discussed in this section.

First, the most logical next step is a replication of this study with a larger, more generalizable sample. Early worked examples research did just this: they replicated studies to confirm and validate findings, made minor adjustments to compensate for areas of methodological weakness, and retested their hypotheses under slightly different conditions (Sweller & Cooper, 1985; Cooper & Sweller, 1987). This study is limited as it draws from a convenience sample of 34 sophomore students, represents the work of one researcher with inherent biases, and spans the duration of studying but one novel in the American literary canon. Future studies could modify the existing study by testing the multimedia worked example effect with analytical reading think-alouds on a larger sample, a slightly older or younger sample, a different novel, or even in a different field (such as analytical reading for social studies or science). Moreover, they could refine the step-by-step structure for the think-aloud process itself, which has been understudied in empirical research. Ideally, these future studies would utilize a second coder and reader, and they would establish inter-rater reliability before reporting conclusive qualitative findings.

Second, this study highlights the need for further research on worked examples of ill-defined problems that include a cognitive strategy domain, such as the heuristic strategies and schema for approaching analytical reading that were presented in the multimedia think-aloud worked examples in this study. Heretofore, the scope of worked examples research of ill-defined problems has included only double-content examples (Schworm & Renkl, 2007), and researchers have called for further investigation into the methods for approaching learning with triple-content worked examples with self-explanation (Renkl et al., 2009). Although this study makes a

first attempt at suggesting that such triple-content worked examples are effective, it is only one small study with a very narrow sample of students. Further research in this field is merited.

Third, this study makes the case for a more careful look at how researchers measure mental effort with ill-defined problems. As mentioned in the limitations and discussion, worked examples research has yet to reach a consensus as to the most accurate way to measure mental effort. While some studies reveal that mental effort can be self-reported by participants accurately and reliably (Paas, 1992; Paas & van Merriënboer, 1993, 1994; Paas et al., 2003) and applied to ill-defined domains (Nivelstein et al., 2013), other literature suggests that the extant instrument is not sensitive enough and should be improved (Brünken et al., 2010). The tightly clustered nature of the mental effort results in this study, for both the treatment and comparison conditions, illustrates the challenges the cognitive load community faces in effectively measuring an invisible mental construct when relying on self-report data. Further research is needed in this area.

Fourth, future studies of multimedia think-aloud worked examples might consider the role of the narrator and his or her relationship to students as a factor influencing the efficacy of the intervention. Dubbed the “personalization principle” (Mayer, 2014c), multimedia learning research has examined the effect of conversational style compared with a more formal speaking style, a standard-accented human voice compared with a computerized voice, and on-screen narrators compared with no on-screen narrators. The literature indicates that students benefit from a human voice that speaks in a conversational tone, but less is known about whether that human voice must belong to someone the student knows. Recent experimental studies have suggested that students may benefit over time from learning with multimedia that features a voice different from the teacher’s (Roche, 2016). However, the qualitative findings from this

study—coupled with the research on literacy learning as situated within a socially mediated process that happens in relationship with a more knowledgeable other—suggest that perhaps further research is needed to determine if adolescent readers would benefit more from studying multimedia that features either the teacher’s voice or a voice other than the teacher’s.

Implications for Practice

This research project began from a place of deep respect for the wisdom of practice; in these closing pages, it returns to that place to humbly offer insights gained during this study. The quantitative and qualitative findings suggest that thoughtful application of multimedia worked examples to the classroom practice of the think-aloud benefits students’ reading comprehension when compared with a traditional model. Multimedia think-aloud worked examples offer practitioners opportunities to provide just-in-time support in a differentiated manner that teaches skills in the context of content and leads students to persist with a challenging analytical reading task.

This study suggests that teachers well-versed in the practice of whole-class think-alouds should feel emboldened to leverage the technology in their classrooms and at students’ fingertips to record their think-alouds for students to watch independently, so long as those videos remain in compliance with the design principles for multimedia learning. Qualitative interview data affirm that one of the advantages of the multimedia think-aloud worked example is that students can take it home with them and experience the just-in-time support they need while they are reading analytically for homework and when they typically do not have access to the teacher. The data suggest that multimedia think-aloud worked examples may be particularly effective early on in the novel, when students are still tuning their ear to new language and becoming familiar with new characters and settings.

Moreover, this study highlights the importance of thoughtfully scaffolding instruction and indicates that teachers may need to be judicious in determining when, in the course of study and skill development, students need the additional support of a multimedia think-aloud worked example—as well as, perhaps, which students do and do not need that level of support. The quantitative results suggest that all students benefited in their early study of the novel. By contrasting the qualitative and quantitative results at the delayed posttest, in which students in the treatment condition scored well on the analytical reading comprehension rubric but had fewer annotations than they did during the intervention, this study suggests that students need less support to achieve at higher levels. Additionally, students who historically struggled the most with analytical reading comprehension seemed to benefit disproportionately from the intervention, as compared to their average or top-performing peers in annotations and analytical reading comprehension. The implication of these data for classroom teachers is a reiteration of the call to thoughtfully differentiate instruction by taking into consideration not only *who* receives scaffolding, but *when* those scaffolds naturally come down.

Additionally, although the think-aloud has traditionally been a live, whole-class instructional strategy for early literacy development, this study suggests that it should not be dismissed for use with older students who are “reading to learn” rather than “learning to read.” Literature teachers who feel the crunch of content, and who perceive that they must prioritize the coverage of disciplinary knowledge over reading strategy instruction, may find relief with multimedia think-aloud worked examples. The current study suggests a possible solution for supporting adolescent readers’ cognitive processing of complex texts through a differentiated approach that does not compete for live class time. Of particular benefit, this solution integrates the skill-based approach and the content-based approach by re-contextualizing the discrete

literacy skills students need within the process of reading interesting and challenging materials, thus bridging the existing pedagogical divide in the literacy instruction community (Appleman, 2010; Learned et al., 2011).

Furthermore, in an age that lauds grit and persistence as paramount to success, this study offers an avenue for supporting emerging analytical readers in persisting with the challenging task of deconstructing an unfamiliar text. Indeed, as English language arts teachers know from practice, a student's resolve to stick with a difficult passage or a complex text is crucial for becoming a more skilled reader. This study suggests that the multimedia think-aloud worked example fosters a sense of partnership between the student and the teacher, and supports a willingness on the part of the student to attempt a task that he or she might not have tried alone. Thus, early results indicate that the multimedia think-aloud worked example may be an effective means of targeting literacy learning in the zone of proximal development in which the student is an active participant in his or her own digital learning environment.

Finally, the multimedia think-aloud worked example may be a promising practice for promoting equity in the classroom. Decade after decade, achievement test data make visible the yawning gap in reading comprehension between students of color and their white peers (Education Trust, 2013, 2015). In schools similar to the one in which this research study was conducted, the divide between students is much deeper than reading scores alone; there is a concomitant divide in students' access to excellent educational resources, such as private tutors and enrichment classes, which causes the achievement gap to widen evermore. Students with access to these resources have the support to become stronger readers. These students read more, becoming exponentially stronger readers; in contrast, struggling readers read less, falling further and further behind their counterparts. Reading researchers have dubbed this pattern of

“cumulative advantage” the Matthew Effect (Stanovich, 1986, p. 374). Reading begets reading. It is a self-extending system (Clay, 1991). The multimedia think-aloud worked example offers a particularly exciting path forward for literacy education, as teachers gain the flexibility to provide additional support outside of the confines of the school day to students with the greatest need through 1:1 technology and well-crafted digital lessons. The multimedia think-aloud worked example equips teachers to engage all students, and particularly our most vulnerable students, in reading deeply and analytically with the support of the teacher as virtual tutor.

Teachers, it seems, always stand at the precipice of change—and navigating that change deftly requires contributions from professionals of all backgrounds, be it literacy instruction, multimedia learning, or cognitive load research. The emergence of the multimedia think-aloud worked example illustrates the natural symmetry that exists across disciplines, as well as the perennial importance and tremendous benefit of fusing theory with practice. As classrooms continue to evolve and access to technology becomes even more pervasive, educators have the opportunity—and the obligation—to transform teaching and learning, guided by research and inspired by possibility.

References

- ACT. (2006). *Reading between the lines: What the ACT reveals about college readiness and reading*. Iowa City, IA: Author.
- Afflerbach, P. P. (1990). The influence of prior knowledge on expert readers' main idea construction strategies. *Reading Research Quarterly*, 25(1), 31–46.
- Afflerbach, P., & Johnston, P. (1984). On the use of verbal reports in reading research. *Journal of Reading Behavior*, 16(4), 307–322.
- Alfassi, M. (2004). Reading to learn: Effects of combined strategy instruction on high school students. *The Journal of Educational Research*, 97(4), 171–185.
- Allington, R. L. (2002). You can't learn much from books you can't read. *Educational Leadership*, 60(3), 16–19.
- Allington, R., & McGill-Franzen, A. (2009). Comprehension difficulties among struggling readers. In S. Israel & G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 551–568). New York, NY: Routledge.
- American Institutes for Research. (2013). *Three decades of education reform: Are we still "A Nation at Risk"?*? Retrieved from <http://www.air.org/resource/three-decades-education-reform-are-we-still-nation-risk>
- Anderson, J. R., & Bower, G. (1983). *Human associative memory*. Washington, DC: Winston.
- Anderson, R. C., & Pearson, P. D. (1984). A schema-theoretic view of basic processes in reading. In P. D. Pearson (Ed.), *Handbook of reading research* (pp. 255-291). New York, NY: Longman.
- Appleman, D. (2010). *Adolescent literacy and the teaching of reading: Lessons for teachers of*

- literature*. Urbana, IL: National Council of Teachers of English.
- Atkinson, R. K., Derry, S. J., Renkl, A., & Wortham, D. (2000). Learning from examples: Instructional principles from the worked examples research. *Review of Educational Research, 70*(2), 181–214.
- Atkinson, R. K., Renkl, A., & Merrill, M. M. (2003). Transitioning from studying examples to solving problems: Combining fading with prompting fosters learning. *Journal of Educational Psychology, 95*, 774–783.
- Ausubel, D. P. (1968). *Educational psychology: A cognitive view*. New York, NY: Holt, Rinehart and Winston.
- Baker, L., & Brown, A. L. (1984). Cognitive monitoring in reading. In J. Flood (Ed.), *Understanding reading comprehension* (pp. 21–44). Newark, DE: International Reading Association.
- Barnhouse, D., & Vinton, V. (2012). *What readers really do: Teaching the process of meaning making*. Portsmouth, NH: Heinemann.
- Baumann, J. F., Jones, L. A., & Seifert-Kessell, N. (1993). Using think alouds to enhance children's comprehension monitoring abilities. *The Reading Teacher, 47*(3), 184–193.
- Beers, K. (2003). *When kids can't read, what teachers can do*. Portsmouth, NH: Heinemann.
- Bereiter, C., & Bird, M. (1985). Use of thinking aloud in identification and teaching of reading comprehension strategies. *Cognition and Instruction, 2*(2), 131–156.
- Biancarosa, C., & Snow, C. E. (2006). *Reading next—A vision for action and research in middle and high school literacy: A report to Carnegie Corporation of New York* (2nd ed.). Washington, DC: Alliance for Excellent Education.
- Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind,*

- experience, and school*. Washington, DC: National Academy Press.
- Bransford, J. D., & Johnson, M. K. (1972). Contextual requisites for understanding: Some investigations of comprehension and recall. *Journal of Verbal Learning and Verbal Behavior*, *11*, 717–726.
- Bratfisch, O., Borg, G., & Dornic, S. (1972). Perceived item-difficulty in three tests of intellectual performance capacity. *Reports from the Institute of Applied Psychology*, *29*.
- Brown, A. L., & Smiley, S. S. (1977). Rating the importance of structural units of prose passages: A problem of metacognitive development. *Child Development*, *48*, 1–8.
- Bruner, J. S. (1978). The role of dialogue in language acquisition. In A. Sinclair, R. J. Jarvella, & W. J. M. Levelt (Eds.), *The child's concept of language* (pp. 241-256). New York, NY: Springer-Verlag.
- Brünken, R., Seufert, T., & Paas, F. (2010). Measuring cognitive load. In J. Plass, R. Moreno, & R. Brünken (Eds.), *Cognitive load theory* (pp. 181–202). New York, NY: Cambridge University Press.
- Burke, J. (2000). *Reading reminders: Tools, tips, and techniques*. Portsmouth, NH: Boynton/Cook.
- Burke, J. (2012). *The English teacher's companion* (4th ed.). Portsmouth, NH: Heinemann.
- California Department of Education. (2015). *California high school exit exam results*. Retrieved from <http://cahsee.cde.ca.gov>
- Caron, T. A. (1989). Strategies for reading expository prose. In S. McCormick & J. Zutell (Eds.), *Cognitive and social perspectives for literacy research and instruction (Thirty-eighth yearbook of the National Reading Conference)* (pp. 293–299). Chicago, IL: National Reading Conference.

- Cartwright, K. B. (2009). The role of cognitive flexibility in reading comprehension: Past, present, and future. In S. Israel & G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 115–139). New York, NY: Routledge.
- Casap, J. (2015). Foreword: How technology is transforming learning. In D. Neebe & J. Roberts (Eds.), *Power up: Making the shift to 1:1 teaching and learning* (pp. ix–xi). Portland, ME: Stenhouse.
- Casner-Lotto, J., & Benner, M. W. (2006). *Are they really ready to work? Employers' perspectives on the basic knowledge and applied skills of new entrants to the 21st-century U.S. workforce*. NY: The Conference Board; Washington, DC: Corporate Voices for Working Families; Washington, DC: Partnership for 21st-Century Skills; and Alexandria, VA: Society for Human Resource Management.
- Chi, M. T. H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161–238). Mahwah, NJ: Lawrence Erlbaum Associates.
- Chi, M. T. H., Bassok, M., Lewis, M. W., Reimann, P., & Glaser, R. (1989). Self-explanations: How students study and use examples in learning to solve problems. *Cognitive Science*, *13*, 145–182.
- Chi, M., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science*, *5*, 121–152.
- Chi, M., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. Sternberg (Ed.), *Advances in the psychology of human intelligence* (pp. 7–75). Hillsdale, NJ: Erlbaum.
- Christenbury, L. (2006). *Making the journey: Being and becoming a teacher of English language arts* (3rd ed.). Portsmouth, NH: Heinemann.

- Clay, M. (1991). *Becoming literate: The construction of inner control*. Portsmouth, NH: Heinemann.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.
- The College Board. (2016). *AP English literature and composition 2016 scoring guidelines*. Retrieved from https://secure-media.collegeboard.org/digitalServices/pdf/ap/ap16_english_literature_sg.pdf
- Collins, A., & Smith, E. E. (1982). Teaching the process of reading comprehension. In D. K. Detterman & R. J. Sternberg (Eds.), *How and how much can intelligence be increased* (pp. 173–185). Norwood, NJ: Ablex.
- Common Core State Standards Initiative. (2012). *English language arts standards*. Retrieved from <http://www.corestandards.org/ELA-Literacy>
- Conley, M. (2009). Improving adolescent comprehension strategies: Developing comprehension strategies in the content areas. In S. Israel & G. Duffy (Eds.), *Handbook of research on reading comprehension* (p. 531–550). New York, NY: Routledge.
- Cooper, G., & Sweller, J. (1987). Effects of schema acquisition and rule automation on mathematical problem-solving transfer. *Journal of Educational Psychology*, 79(4), 347–362.
- Crain-Thoreson, C., Lippman, M. Z., & McClendon-Magnuson, D. (1997). Windows on comprehension: Reading comprehension processes as revealed by two think-aloud procedures. *Journal of Educational Psychology*, 89(4), 579–591.
- Creswell, J. W. (2015). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). Thousand Oaks, CA: Sage.

- Crippen, K. J., & Earl, B. L. (2007). The impact of web-based worked examples and self-explanation on performance, problem solving, and self efficacy. *Computers and Education, 49*, 809–821.
- Cuevas, J. A., Russell, R. L., & Irving, M. A. (2012). An examination of the effect of customized reading modules on diverse secondary students' reading comprehension and motivation. *Educational Technology Research and Development, 60*(3), 445–467.
- Davey, B. (1983). Think aloud: Modeling the cognitive process for reading comprehension. *Journal of Reading, 27*, 44–47.
- De Groot, A. (1966). Perception and memory versus thought: Some old ideas and recent findings. In B. Kleinmuntz (Ed.), *Problem solving: Research, method, and theory* (pp. 19–50). New York, NY: Wiley.
- Dole, J., Nokes, J., & Dritis, D. (2009). Cognitive strategy instruction. In S. Israel & G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 347–372). New York, NY: Routledge.
- Education Trust. (2013). *Gaining traction, gaining ground: How some high schools accelerate learning for struggling students*. Retrieved from <http://edtrust.org/wp-content/uploads/2013/10/GainingTractionGainingGround.pdf>
- Education Trust. (2015). *Black minds matter: Supporting the educational success of black children in California*. Retrieved from <https://west.edtrust.org/wp-content/uploads/sites/3/2015/10/Ed-Trust-West-Black-Minds-Matter-FINAL-PDF.pdf>
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data (Rev. ed.)*.

- Cambridge, MA: MIT Press.
- Fisher, D., Frey, N., & Hattie, J. (2016). *Visible learning for literacy: Implementing the practices that work best to accelerate student learning*. Thousand Oaks, CA: Corwin.
- Fitzgerald, F. (1925). *The great Gatsby*. New York, NY: Scribner.
- Flavell, J. H. (1973). Metacognitive aspects of problem-solving. In L. B. Resnick (Ed.), *The nature of intelligence* (pp. 231-235). Hillsdale, NJ: Erlbaum.
- Fowler, F. (2014). *Survey research methods* (5th ed.). Los Angeles, CA: Sage.
- Freebody, P., & Luke, A. (1990). Literacy programs: Debates and demands in cultural context. *Prospect: Australian Journal of TESOL*, 5(7), 7–16.
- Frey, N., & Fisher, D. (2013). *Rigorous reading: 5 access points for comprehending complex texts*. Thousand Oaks, CA: Corwin Press.
- Freytag, G. (1863/2016). Freytag's pyramid. In D. Chandler & R. Munday (Eds.), *A dictionary of media communication* (2nd ed.). Available from <http://www.oxfordreference.com/view/10.1093/acref/9780199568758.001.0001/acref-9780199568758>.
- Gallagher, K. (2004). *Deeper reading: Comprehending challenging texts, 4-12*. Portland, ME: Stenhouse.
- Gallagher, K. (2009). *Readicide: How schools are killing reading and what you can do about it*. Portland, ME: Stenhouse.
- Gerjets, P., Scheiter, K., & Schuh, J. (2008). Information comparisons in example-based hypermedia environments: Supporting learners with processing prompts an interactive comparison tool. *Educational Technology Research and Development*, 56, 73–92.
- Glogger, I., Holzapfel, L., Schwonke, R., Nückles, M., & Renkl, A. (2009). Activation of

- learning strategies in writing learning journals: The specificity of prompts matters. *German Journal of Educational Psychology*, 23, 95–104.
- Gough, P. B. (1972). One second of reading. In J. F. Kavanagh & I. G. Mattingly (Eds.), *Language by ear and eye* (pp. 331–358). Cambridge, MA: MIT Press.
- Gray, D. (2014). *Doing research in the real world* (3rd ed.). London: Sage.
- Greeno, J. G., Collins, A. M., & Resnick, L. (1996). Cognition and learning. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 15–46). New York, NY: Macmillan.
- Griffin, T. D., Wiley, J., & Thiede, K. W. (2008). Individual differences, rereading, and self-explanation: Concurrent processing and cue validity as constraints on meta-comprehension accuracy. *Memory & Cognition*, 36(1), 93–103.
- Hare, V. C. (1981). Readers' problem identification and problem solving strategies for high- and low-knowledge articles. *Journal of Reading Behavior*, 13(4), 359–365.
- Hartman, D. K. (1995). Eight readers reading: The intertextual links of proficient readers reading multiple passages. *Reading Research Quarterly*, 30(3), 520–561.
- Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. New York, NY: Routledge.
- Hemingway, E. (1932). *Death in the afternoon*. New York, NY: Scribner.
- Herber, H. L. (1978). *Teaching reading in content areas* (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.
- Herbert, M., Bohaty, J., Nelson, J., & Brown, J. (2016). The effects of text structure instruction on expository reading comprehension: A meta-analysis. *Journal of Educational Psychology*, 108(5), 609–629.

- High School Placement Test: Form 2C. (2014). Bensenville, IL: Scholastic Testing Service.
- Hollenbeck, A., & Saternus, K. (2013). Mind the comprehension iceberg: Avoiding titanic mistakes with the CCSS. *Reading Teacher, 66*(7), 558–568. doi:10.1002/TRTR.1160
- Hübner, S., Nückles, M., & Renkl, A. (2010). Writing learning journals: Instructional support to overcome learning-strategy deficits. *Learning and Instruction, 20*, 18–29.
- Huey, E. B. (1908/1968). *The psychology and pedagogy of reading: With a review of the history of reading and writing and of methods, texts, and hygiene in reading*. Cambridge, MA: MIT Press.
- International Society for Technology in Education. (2008). *ISTE standards for teachers*. Retrieved from <http://www.iste.org/standards/standards/standards-for-teachers>
- Jung, H., & Kim, D. (2006). The effect of the worked-out example types on knowledge sharing process in computer-supported collaborative learning. *Journal of Educational Technology in Korea, 22*(3), 23–56.
- Kalyuga, S. (2006). Rapid cognitive assessment of learners' knowledge structures. *Learning and Instruction, 16*, 1–11.
- Kalyuga, S. (2007). Expertise reversal effect and its implications for learner-tailored instruction. *Educational Psychology Review, 19*, 509–539.
- Kalyuga, S. (2010). Schema acquisition and sources of cognitive load. In J. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive load theory* (pp. 48–64). New York, NY: Cambridge University Press.
- Kalyuga, S., Ayres, P., Chandler, P., & Sweller, J. (2003). The expertise reversal effect. *Educational Psychologist, 28*, 23–32.
- Kalyuga, S., Chandler, P., & Sweller, J. (1998). Levels of expertise and instructional design.

- Human Factors*, 40(1), 1–17.
- Kalyuga, S., & Sweller, J. (2004). Measuring knowledge to optimize cognitive load factors during instruction. *Journal of Educational Psychology*, 96, 558–568.
- Keene, E. O. (2010). New horizons in comprehension. *Educational Leadership*, 67(6), 69–73.
- Keene, E. O., & Zimmermann, S. (2007). *Mosaic of thought: The power of comprehension strategy instruction* (2nd ed.). Portsmouth, NH: Heinemann.
- Khan, S. (Producer). (2011). Liberating the classroom for creativity. *Edutopia*. Video available from <http://www.edutopia.org/salman-khan-academy-flipped-classroom-video>
- Kim, K. (2005). Effect of practice problem on cognitive skill acquisition in multimedia-based instruction. *The Journal of Educational Information and Media*, 21(3), 33–62.
- Kincaid, J. (1990). *Lucy: A novel*. New York, NY: Farrar, Straus, and Giroux.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review*, 98, 163–182.
- Kintsch, W., & van Dijk, T. A. (1978). Toward a model of text comprehension and production. *Psychological Review*, 85, 363–394.
- Koehler, M. J., & Mishra, P. (2009). What is technological pedagogical content knowledge? *Contemporary Issues in Technology and Teacher Education*, 9(1), 60–70.
- Kucan, L., & Beck, I. (1997). Thinking aloud and reading comprehension research: Inquiry, instruction, and social interaction. *Review of Educational Research*, 67, 271–299.
- Kyun, S., Kalyuga, S., & Sweller, J. (2013). The effect of worked examples when learning to write essays in English literature. *Journal of Experimental Education*, 81(3), 385–408.
- Kyun, S. A., & Lee, H. (2009). The effects of worked examples in computer-based instruction: Focus on the presentation format of worked examples and prior knowledge of learners.

- Asia Pacific Education Review*, 10(4), 495–503.
- Langer, J. A. (1990). The process of understanding: Reading for literary and informative purposes. *Research in the Teaching of English*, 24(3), 229–260.
- Langer, J. A. (2002). *Effective literacy instruction: Building successful reading and writing programs*. Urbana, IL: National Council of Teachers of English.
- Lattimer, H. (2010). *Reading for learning: Using discipline-based texts to build content knowledge*. Urbana, IL: National Council of Teachers of English.
- Lattimer, H. (2014). *Real-world literacies: Disciplinary teaching in the high school classroom*. Urbana, IL: National Council of Teachers of English.
- Learned, J. E., Stockdill, D., & Moje, E. B. (2011). Integrating reading strategies and knowledge building in adolescent literacy instruction. In S. J. Samuels & A. E. Farstrup (Eds.), *What research has to say about reading instruction* (4th ed., pp. 159–185). Newark, DE: International Reading Association.
- Lewis, W. E., & Ferretti, R. P. (2009). Defending interpretations of literary texts: The effects of TOPOI instruction on the literary arguments of high school students. *Reading & Writing Quarterly*, 25(4), 250–270.
- Lipson, M. Y. (1983). The influence of religious affiliation on children's memory for text information. *Reading Research Quarterly*, 18, 448–457.
- Lowry, L. (1989). *Number the stars*. New York, NY: Laurel Leaf.
- Mayer, R. E. (2005). *The Cambridge handbook of multimedia learning*. New York, NY: Cambridge University Press.
- Mayer, R. E. (2009). *Multimedia learning* (2nd ed.). New York, NY: Cambridge University Press.

- Mayer, R. E. (2014a). Introduction to multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 1–24). New York, NY: Cambridge University Press.
- Mayer, R. E. (2014b). Cognitive theory of multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 43–71). New York, NY: Cambridge University Press.
- Mayer, R. E. (2014c). Principles based on social cues in multimedia learning: Personalization, voice, image, and embodiment principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 345–368). New York, NY: Cambridge University Press.
- Mayer, R. E., & Fiorella, L. (2014). Principles for reducing extraneous processing in multimedia learning: Coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 279–315). New York, NY: Cambridge University Press.
- Mayer, R. E., & Moreno, R. (2010). Techniques that reduce extraneous cognitive load and manage intrinsic cognitive load during multimedia learning. In J. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive load theory* (pp. 131–152). New York, NY: Cambridge University Press.
- McKeown, R. G., & Gentilucci, J. L. (2007). Think-aloud strategy: Metacognitive development and monitoring comprehension in the middle school second-language classroom. *Journal of Adolescent & Adult Literacy*, *51*(2), 136–147.
- McNamara, D. S., & Kintsch, W. (1996). Learning from texts: Effects of prior knowledge and text coherence. *Discourse Processes*, *22*, 247–282.

- Miller, D. (2009). *The book whisperer: Awakening the inner reader in every child*. San Francisco, CA: Jossey-Bass.
- Miller, G. A. (1956). The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological Review*, 63(2), 81–97.
- Milner, J., Milner, L., & Mitchell, J. (2011). *Bridging English* (5th ed.). Boston, MA: Pearson.
- Mishra, P., & Koehler, M. J. (2006). Technological pedagogical content knowledge: A framework for teacher knowledge. *Teachers College Record*, 108(6), 1017–1054.
- Mitchell, M. (2015). *Effect size tutorial*. Retrieved from <http://mathewmitchell.net/tutorials/effectsiz>
- Moreno, R., & Mayer, R. (2010). Techniques that increase generative processing in multimedia learning: Open questions for cognitive load research. In J. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive load theory* (pp. 153–177). New York, NY: Cambridge University Press.
- Moreno, R., & Park, B. (2010). Cognitive load theory: Historical development and relation to other theories. In J. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive load theory* (pp. 9–28). New York, NY: Cambridge University Press.
- Mousavi, S., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology*, 87, 319–334.
- National Center for Education Statistics. (2015a). *National assessment of educational progress: 2015 reading assessment report card*. Retrieved from http://www.nationsreportcard.gov/reading_math_g12_2015/#reading
- National Center for Education Statistics. (2015b). *The condition of education 2015*. Retrieved from <http://nces.ed.gov/pubs2015/2015144.pdf>

- National Endowment for the Arts. (2007). *To read or not to read: A question of national consequence. Research report #47*. Washington, DC: National Endowment for the Arts. Retrieved from <https://www.arts.gov/sites/default/files/ToRead.pdf>
- National Reading Panel. (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction. Reports of the subgroups*. Retrieved from <https://www.nichd.nih.gov/publications/pubs/nrp/Documents/report.pdf>
- Neebe, D., & Roberts, J. (2015). *Power up: Making the shift to 1:1 teaching and learning*. Portland, ME: Stenhouse.
- Nievelstein, F., van Gog, T., Van Dijck, G., & Boshuizen, H. P. A. (2013). The worked example and expertise reversal effect in less-structured tasks. *Contemporary Educational Psychology, 38*, 118–125.
- Nückles, M., Hübner, S., & Renkl, A. (2009). Enhancing self-regulated learning by writing learning protocols. *Learning and Instruction, 19*(3), 259–271.
- Oakhill, J., & Yuill, N. (1996). On the nature of the difference between skilled and less skilled comprehenders. *Journal of Research in Reading, 9*, 80–91.
- Olshavsky, J. E. (1977). Reading as problem solving: An investigation of strategies. *Reading Research Quarterly, 12*(4), 654–674.
- Olson, C. B., & Land, R. (2007). A cognitive strategies approach to reading and writing instruction for English language learners in secondary school. *Research in the Teaching of English, 41*(3), 269–303.
- Olson, G. M., Duffy, S. A., & Mack, R. L. (1984). Thinking-out-loud as a method for studying real-time comprehension processes. In D. E. Kieras & M. A. Just (Eds.), *New methods in*

- reading comprehension research* (pp. 253–286). Hillsdale, NJ: Erlbaum.
- Organisation for Economic Co-operation and Development. (2012). *Programme for international student assessment (PISA): Results from the PISA 2012*. Retrieved from <https://www.oecd.org/unitedstates/PISA-2012-results-US.pdf>
- Oster, L. (2001). Using the think-aloud for reading instruction. *The Reading Teacher*, 55, 64–69.
- Paas, F. (1992). Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84(4), 429–434.
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, 38(1), 1–4.
- Paas, F., & Sweller, J. (2014). Implications of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 27–42). New York, NY: Cambridge University Press.
- Paas, F., Tuovinen, J. E., Tabbers, H., & Van Gerven, P. W. M. (2003). Cognitive load measurement as a means to advance cognitive load theory. *Educational Psychologist*, 38, 63–71.
- Paas, F., & van Merriënboer, J. (1993). The efficiency of instructional conditions: An approach to combine mental-effort and performance measures. *Human Factors*, 35, 737–743.
- Paas, F., & van Merriënboer, J. (1994). Variability of worked examples and transfer of geometrical problem-solving skills: A cognitive-load approach. *Journal of Educational Psychology*, 86, 122–133.
- Paivio, A. (1986). *Mental representations: A dual coding approach*. England: Oxford University Press.
- Palinscar, A. S., & Brown, A. L. (1984). Reciprocal teaching of comprehension-fostering and

- comprehension monitoring activities. *Cognition and Instruction*, 1(2), 117–175.
- Palincsar, A. S., & Brown, A. L. (1988). Teaching and practicing thinking skills promote comprehension in the context of group problem solving. *Remedial and Special Education*, 9(1), 53–59.
- Palincsar, A. S., & Klenk, L. (1992). Fostering literacy learning in supportive contexts. *Journal of Learning Disabilities*, 25(4), 211–225.
- Paris, S. G., & Lindauer, B. K. (1976). The role of inference in children's comprehension and memory for sentences. *Cognitive Psychology*, 8, 217–227.
- Pearson, P. D., Hansen, J., & Gordon, C. (1979). The effect of background knowledge on young children's comprehension of explicit and implicit information. *Journal of Reading Behavior*, 11, 201–209.
- Pichert, J., & Anderson, R. C. (1977). Taking different perspectives on a story. *Journal of Educational Psychology*, 69, 309–315.
- Piekarz, J. A. (1956). Getting meaning from reading. *The Elementary School Journal*, 56(7), 303–309.
- Pirolli, P., & Recker, M. (1994). Learning strategies and transfer in the domain of programming. *Cognition and Instruction*, 12, 235–275.
- Plass, J., Kalyuga, S., & Leutner, D. (2010). Individual differences and cognitive load theory. In J. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive load theory* (pp. 65–87). New York, NY: Cambridge University Press.
- Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction*, 12(1), 61–86.
- Pressley, M., & Afflerbach, P. (1995). *Verbal protocols of reading: The nature of constructively*

- responsive reading*. Hillsdale, NJ: Erlbaum.
- Pressley, M., Graham, S., & Harris, K. (2006). The state of educational intervention research as viewed through the lens of literacy intervention. *British Journal of Educational Psychology, 76*, 1–19.
- RAND Reading Study Group. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Santa Monica, CA: RAND.
- Reed, S. K., & Bolstad, C. A. (1991). Use of examples and procedures in problem solving. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 17*, 753–766.
- Reiser, B., & Tabak, I. (2014). Scaffolding. In R. K. Sawyer (Ed.), *The Cambridge handbook of learning sciences* (2nd ed., pp. 44–62). New York, NY: Cambridge University Press.
- Reisslein, J., Atkinson, R. K., Seeling, P., & Reisslein, M. (2006). Encountering the expertise reversal effect with a computer-based environment on electrical circuit analysis. *Learning and Instruction, 16*, 92–103.
- Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science, 12*, 1–29.
- Renkl, A. (2014). The worked examples principle in multimedia learning. In R. E. Mayer (Ed.) *The Cambridge handbook of multimedia learning* (2nd ed., pp. 391–412). New York, NY: Cambridge University Press.
- Renkl, A., & Atkinson, R. K. (2010). Learning from worked-out examples and problem solving. In J. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive load theory* (pp. 91–108). New York, NY: Cambridge University Press.
- Renkl, A., Hilbert, T., & Schworm, S. (2009). Example-based learning in heuristic domains: A cognitive load theory account. *Educational Psychology Review, 21*(1), 67–78.

- Roche, C. (2016). *The effect of more and less relevant details and teacher voice on student retention and problem-solving transfer in teacher-created multimedia* (Doctoral dissertation). Retrieved from ProQuest. (10147157).
- Rogoff, B. (1995). Observing sociocultural activity on three planes: Participatory appropriation, guided participation, and apprenticeship. In J. V. Wertsch, P. del Rio, & A. Alvarez (Eds.), *Sociocultural studies of mind* (pp. 139–164). Cambridge, England: Cambridge University.
- Rosenblatt, L. (1938/1983). *Literature as exploration* (4th ed.). New York, NY: Modern Language Association of America.
- Rourke, A., & Sweller, J. (2009). The worked-example effect using ill-defined problems: Learning to recognize designers' styles. *Learning and Instruction, 19*, 185–199.
- Rummel, N., Spada, H., & Hauser, S. (2009). Learning to collaborate while being scripted or by observing a model. *International Journal of Computer-Supported Collaborative Learning, 4*(1), 69–92.
- Rupley, W. H., Blair, T. R., & Nichols, W. D. (2009). Effective reading instruction for struggling readers: The role of direct/explicit teaching. *Reading & Writing Quarterly, 25*(2–3), 125–138.
- Sams, A., & Bergmann, J. (2013). Flip your students' learning. *Educational Leadership, 70*(6), 16–20.
- Schellings, G., Aarnoutse, C., & Leeuwe, J. Van. (2006). Third-graders' think-aloud protocols: Types of reading activities in reading an expository text. *Learning and Instruction, 16*, 594–568.
- Schmidt, H., Loyens, S., van Gog, T., & Paas, F. (2007). Problem-based learning is compatible

- with human cognitive architecture: Commentary on Kirschner, Sweller, and Clark (2006). *Educational Psychologist*, 42, 91–97.
- Schoenbach, R., Greenleaf, C., & Murphy, L. (2012). *Reading for understanding: How reading apprenticeship improves disciplinary learning in secondary and college classrooms*. San Francisco, CA: WestEd.
- Schworm, S., & Renkl, A. (2007). Learning argumentation skills through the use of prompts for self-explaining examples. *Journal of Educational Psychology*, 99, 285–296.
- Shanahan, C. (2009). Disciplinary comprehension. In S. Israel & G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 240–260). New York, NY: Routledge.
- Shulman, L. S. (1986). Those who understand: Knowledge growth in teaching. *Educational Researcher*, 15(2), 4–14.
- Shulman, L. S. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Educational Review*, 57(1), 1–22.
- Silven, M., & Vauras, M. (1992). Improving reading through thinking aloud. *Learning and Instruction*, 2(2), 69–88.
- Simons, H. D. (1971). Reading comprehension: The need for a new perspective. *Reading Research Quarterly*, 6(3), 338–363.
- Smagorinsky, P. (2001). If meaning is constructed, what's it made from? Toward a cultural theory of reading. *Review of Educational Research*, 71, 133–169.
- Smagorinsky, P., & Mayer, R. E. (2014). Learning to be literate. In R. K. Sawyer (Ed.), *The Cambridge handbook of learning sciences* (2nd ed., pp. 605–625). New York, NY: Cambridge University Press.
- Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual

- differences in the acquisition of literacy. *Reading Research Quarterly*, 360–407.
- Swanson, H. (1993). Individual differences in working memory: A model testing and subgroup analysis of learning-disabled and skilled readers. *Intelligence*, 17(3), 285–332.
- Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2), 257–285.
- Sweller, J. (1994). Cognitive load theory, learning difficulty, and instructional design. *Learning and Instruction*, 4, 295–312.
- Sweller, J. (2006). The worked example effect and human cognition. *Learning and Instruction*, 16, 165–169.
- Sweller, J. (2010a). Cognitive load theory: Recent theoretical advances. In J. Plass, R. Moreno, & R. Brunken (Eds.), *Cognitive load theory* (pp. 29–47). New York, NY: Cambridge University Press.
- Sweller, J. (2010b). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, 22(2), 123–138.
- Sweller, J. (2012). Human cognitive architecture. In K. R. Harris, S. Graham, & T. Urdan (Eds.), *APA Educational Psychology Handbook* (pp. 295–326). Washington, DC: American Psychological Association.
- Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. *Cognition and Instruction*, 12, 185–233.
- Sweller, J., & Cooper, G. A. (1985). The use of worked examples as a substitute for problem solving in learning algebra. *Cognition and Instruction*, 2(1), 59–89.
- Sweller, J., van Merriënboer, J., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3), 251–296.

- Tankersley, K. (2005). *Literacy strategies for grades 4-12: Reinforcing the threads of reading*. Alexandria, VA: ASCD.
- Taylor, B. (1980). Children's memory for expository text after reading. *Reading Research Quarterly, 19*, 134–146.
- Thorndike, E. L. (1917). Reading as reasoning: A study of mistakes in paragraph reading. *Journal of Educational Psychology, 8*(6), 323–332.
- Tomlinson, C. (2014). *The differentiated classroom: Responding to the needs of all learners* (2nd ed.). Alexandria, VA: Association for Supervision and Curriculum Development.
- Tovani, C. (2000). *I read it, but I don't get it: Comprehension strategies for adolescent readers*. Portland, ME: Stenhouse.
- Tucker, C. (2012). *Blended learning in grades 4-12: Leveraging the power of technology to create student-centered classrooms*. Thousand Oaks, CA: Corwin.
- Tuovinen, J., & Sweller, J. (1999). A comparison of cognitive load associated with discovery learning and worked examples. *Journal of Educational Psychology, 91*, 334–341.
- Unsworth, N., & McMillan, B. D. (2013). Mind wandering and reading comprehension: Examining the roles of working memory capacity, interest, motivation, and topic experience. *Journal of Experimental Psychology, 39*(3), 832–842.
- US Department of Education. (2015). *Race to the top: Program overview*. Retrieved from <http://www2.ed.gov/programs/racetothetop/index.html>
- Vaca, R. T., & Vaca, J. L. (2008). *Content area reading: Literacy and learning across the curriculum* (9th ed.). Boston, MA: Pearson.
- Van Dijk, T. A., & Kintsch, W. (1983). *Strategies for discourse comprehension*. Orlando, FL: Academic Press.

- Van Gog, T., Kirschner, F., Kester, L., & Paas, F. (2012). Timing and frequency of mental effort measurement: Evidence in favor of repeated measures. *Applied Cognitive Psychology*, 26(6), 833–839.
- Van Merriënboer, J., & Kester, L. (2014). The four-component instructional design model: Multimedia principles in environments for complex learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (2nd ed., pp. 104–148). New York, NY: Cambridge University Press.
- Van Merriënboer, J. J. G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent developments and future directions. *Educational Psychology Review*, 17, 147–177.
- Vygotsky, L. S. (1930/1978). *Mind in society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.
- Vygotsky, L. S. (1934/1962). *Thought and language*. Cambridge, MA: MIT Press.
- Wade, S. E., Buxton, W. M., & Kelly, M. (1999). Using think-alouds to examine reader-text interest. *Reading Research Quarterly*, 34(2), 194–216.
- Webb, R. (1856). *Webb's normal readers, number 4: Designed to teach correct reading, to improve and expand the mind, and to purify and elevate the character*. New York: Sheldon, Lamport & Blakerman.
- Wigfield, A., Guthrie, J. T., Perencevich, K. C., Taboada, A., Klauda, S. L., McRae, A., & Barbosa, P. (2008). The role of reading engagement in mediating effects of reading comprehension instruction on reading outcomes. *Psychology in the Schools*, 45, 432–445.
- Wilhelm, J. D. (2001). *Improving comprehension with think-aloud strategies: Modeling what good readers do*. New York, NY: Scholastic.

- Winne, P., & Azevedo, R. (2014). Metacognition. In R. K. Sawyer (Ed.), *The Cambridge handbook of learning sciences* (2nd ed., pp. 63–87). New York, NY: Cambridge University Press.
- Wood, D., Bruner, J. S., & Ross, G. (1976). The role of tutoring in problem solving. *Journal of Child Psychology and Psychiatry, 17*, 89–100.
- Yeo, G. B., & Neal, A. (2004). A multilevel analysis of effort, practice and performance: effects of ability, conscientiousness, and goal orientation. *Journal of Applied Psychology, 89*, 231–247.
- Yeo, G., & Neal, A. (2008). Subjective cognitive effort: A model of states, traits, and time. *Journal of Applied Psychology, 93*, 617–631.

Appendix A
Principal Letter of Support



SACRED HEART SCHOOLS
PREPARATORY

November 7, 2016

Dear Members of the IRB Committee:

I write to affirm that I fully endorse the research study proposed by Diana Neebe to evaluate the efficacy of multimedia worked examples in her US Literature courses. The study seeks to address a significant practical, pedagogical problem in providing differentiated support to adolescent readers who are developing their analytical reading comprehension processes. Participation in the intervention, as well as in subsequent interviews, is likely to be educative for the students.

I have reviewed the IRB protocol for this proposed study. Given my direct experience both conducting educational research and guiding doctoral candidates (in my previous position I was Associate Dean of Teacher Education at CU-Boulder), I can affirm that the study is well-designed, that the assessment of risk is accurate, and that all appropriate measures for clear parental communication and data anonymity are being taken, given that minors are participants in the study.

Once the study is approved, I will send a parent letter home with the consent form in which I acknowledge my awareness of the study, endorse our student's participation, and also express how I value the opportunity for our school to be an active site for educational research. I believe the students will likely benefit from their participation, and the instructional materials developed will likely be of pedagogical value within our school and with those who desire to strengthen adolescent literacy.

Please contact me if you have any questions.

Sincerely,

Jennie Whitcomb

Jennie Whitcomb, PhD
Principal



SACRED HEART SCHOOLS
PREPARATORY

November 14, 2016

Dear Parents:

I write to affirm that I fully endorse the research study proposed by Diana Neebe and encourage you to consent to have your son or daughter participate.

Ms. Neebe is working to assess the effectiveness of a specific teaching method around analytical reading that will be of practical value and help teachers enhance individualized reading instruction. The study she will be engaging your son or daughter in is one that will likely help them gain insight into their reading comprehension processes, particularly when tackling challenging literary texts. Thus, the research process is one that is both instructionally sound and likely a powerful learning experience for your child.

I value the opportunity for our school to be an active site for educational research. Both our students and others elsewhere stand to benefit from well-conceived studies such as this.

Please contact me if you have any questions.

Sincerely,

Jennie Whitcomb

Jennie Whitcomb, PhD
Principal

Appendix B

Informed Consent Letter



14 November 2016

Dear Parents and Guardians:

In addition to my work at Sacred Heart as your child's English teacher, I am also a doctoral student in the School of Education at the University of San Francisco. I am sending this letter to explain why I am requesting that your child participate in my research study. I am in the final stages of my dissertation and will be conducting a study of an instructional intervention around analytical reading comprehension. My area of research rests at the intersection of adolescent literacy and technology. With the introduction of technology into the classroom, many teachers, myself included, have relied on "trial and error" to determine best practices in education now that all of our students have access to iPads and laptops. My goal is to take a step back and study what is working, what isn't, and where we have opportunities to better leverage the technology at our fingertips to improve the quality of instruction we deliver.

Last year, I conducted a pilot study with students at Sacred Heart. The results were promising. In this phase of the project, I am seeking to better understand students' meaning-making processes and strategies for analytical reading comprehension, and am testing the effectiveness of a particular technology-rich instructional strategy for supporting analytical reading. The study will take place during our unit on *The Great Gatsby* in January. I have developed the study under the guidance of our principal, Dr. Jennie Whitcomb, to fit seamlessly within the existing unit and to minimize the impact on student learning time. Students will be randomly assigned to one of two groups: the new instructional strategy being tested, or the traditional method that my colleagues and I have used at Sacred Heart for the past few years. Regardless of assigned group, all students will have access to all of the instructional materials once the intervention phase is complete.

With your permission, I will ask your child to take a brief survey on three occasions during class that should take no more than five minutes each time. I will collect your child's annotations and his or her in-class analytical writing exercises for the first five chapters of the novel. In analyzing the data I collect, I will match up students' responses with the HSPT reading comprehension scores we have on file, allowing me to identify patterns in students' reading experiences, and ultimately, equipping me to more effectively target instructional techniques to students' areas of need. At the end of the study, I will ask your child to spend approximately fifteen minutes participating in an interview or in a focus group with his or her peers, discussing the analytical reading process.

Your child's participation in this study is voluntary and will not affect his or her grade in any way. Your child may quit this study at any time by simply saying "I do not wish to participate." It may seem odd that I am asking for permission to talk with your child about his or her experience with reading, or to analyze his or her writing about a text we are reading together. It is true that we engage in these activities regularly *in my capacity as his/her English teacher*. However, legally, I have to take my "teacher hat" off and put my "researcher hat" on for this project, which means I need your permission to study your child's reading experience.

The study will be conducted during the first three weeks of the new semester. Your child will have the opportunity to sign up for an interview or focus group session that works with his or her schedule. There are no known risks involved in this study and your child will not receive any compensation for his or her participation. To protect your child's confidentiality, your child's name will not appear on any record sheets. The information obtained will not be shared with anyone, unless required by law. The records will be maintained by me, and by my dissertation chair, Dr. Mathew Mitchell. If you have any questions, please contact me at 650-454-8370 or via email at dneebe@shschools.org.

This letter will serve as a consent form for your child's participation and will be kept by both Dr. Jennie Whitcomb, principal at Sacred Heart Preparatory (jwhitcomb@shschools.org), and by Dr. Mathew Mitchell, faculty sponsor at the University of San Francisco School of Education (mitchellm@usfca.edu). If you have any questions about your child's rights as a participant, you may contact the University of San Francisco Institutional Review Board (IRB) for tests of human subjects at IRBPHS@usfca.edu. Please have your child return the signed form (next page) to me via Schoology by Monday, November 21st.

With Gratitude,

Diana Neebe

Please Sign One of the Statements Below

Statement of Consent

I read the above consent form for the project entitled *Adolescent Reading Comprehension Study*, conducted by Diana Neebe, a doctoral student at the University of San Francisco. I give my consent for my child to participate in this study. The nature, demands, risk, and benefits of the project have been explained to me. I am aware that I have the opportunity to ask questions about this research. I understand that I may withdraw my consent and discontinue my child's participation at any time without penalty.

Child's Name (print clearly)

Signature of Legal Guardian

Date

- OR -

Refusal of Consent

I read the above consent form for the project entitled *Adolescent Reading Comprehension Study*, conducted by Diana Neebe, a doctoral student at the University of San Francisco. I do not give my consent for my child to participate in this study.

Child's Name (print clearly)

Signature of Legal Guardian

Date

Appendix C

Multimedia Think-Aloud Worked Example Video

URL and Screenshots

Chapter 2 multimedia think-aloud worked example video is available online via:
<http://tinyurl.com/neebe-worked-example>

Close Reading

Great Gatsby Chapter 2

What? How? Why?

About half way between West Egg and New York the motor-road hastily joins the railroad and runs beside it for a quarter of a mile, so as to shrink away from a certain desolate area of land. This is a valley of ashes—a fantastic farm where ashes grow like wheat into ridges and hills and grotesque gardens where ashes take the forms of houses and chimneys and rising with a transcendent effort, of men who move dimly and already crumbling through the powdery air. Occasionally a line of grey cars crawls along an invisible track, gives out a ghastly creak and comes to rest, and immediately the ash-grey men swarm up with leaden spades and stir up an impenetrable cloud which screens their obscure operations from your sight.



About half way between West Egg and New York the motor-road hastily joins the railroad and runs beside it for a quarter of a mile, so as to shrink away from a certain desolate area of land. This is a valley of **ashes** a fantastic farm where **ashes** grow like wheat into ridges and hills and grotesque gardens where **ashes** take the forms of houses and chimneys and **rising smoke** and finally, with a transcendent effort, of men who move **dimly** and already crumbling through the **powdery air**. Occasionally a line of **grey** cars crawls along an **invisible** track, gives out a ghastly creak and comes to rest, and immediately the **ash-grey** men swarm up with leaden spades and stir up an **impenetrable cloud** which screens their obscure operations from your sight.



How have I used the analytical reading strategy of **visualizing** in the opening paragraph?

east
egg

valley
of ashes

Appendix D

Traditional Text-Based Questions Video

URL and Screenshots

Traditional text-based questions video for Chapter 2 is available online via:
<http://tinyurl.com/neeбетraditional>

Close Reading

Great Gatsby Chapter 2



What? How? Why?

3

Now, go back through the first paragraph. What **colors** stand out to you? Mark all the instances you see color imagery. What does that color make you feel about this setting?

4

The setting is called the Valley of Ashes. What do you associate with **ashes**? What does this association make you think about the people living here?

8

Finally, how is this setting different from East Egg, which we just visited in the end of Chapter 1? Why do you think Fitzgerald so closely **juxtaposed those two settings**?

Appendix E

Passage Annotation Handouts

Chapter 1, Pages 6-8

Please annotate the passage below, marking the specific details in the text that illustrate the setting and that help you make inferences about the characters associated with this setting.

And so it happened that on a warm windy evening I drove over to East Egg to see two old friends whom I scarcely knew at all. Their house was even more elaborate than I expected, a cheerful red and white Georgian Colonial mansion overlooking the bay. The lawn started at the beach and ran toward the front door for a quarter of a mile, jumping over sun-dials and brick walks and burning gardens--finally when it reached the house drifting up the side in bright vines as though from the momentum of its run. The front was broken by a line of French windows, glowing now with reflected gold, and wide open to the warm windy afternoon...

We walked through a high hallway into a bright rosy-colored space, fragilely bound into the house by French windows at either end. The windows were ajar and gleaming white against the fresh grass outside that seemed to grow a little way into the house. A breeze blew through the room, blew curtains in at one end and out the other like pale flags, twisting them up toward the frosted wedding cake of the ceiling--and then rippled over the wine-colored rug, making a shadow on it as wind does on the sea.

PROMPT: What are the specific details that you notice about the setting in this passage? Based on these details, what inferences can you draw about the character(s) associated with that setting? *Your response will have two parts: what I see, and what I infer. Start the first part of your response with, "In this passage, I see..." Start the second part of your response with, "Therefore, I infer that..." Try to include as many observations and inferences as you can, and make sure that your observations and inferences are clearly connected. Type your response in Schoology in the special course for Mrs. Neebe's study.*

Chapter 2, Page 23

Please annotate the passage below, marking the specific details in the text that illustrate the setting and that help you make inferences about the characters associated with this setting.

About half way between West Egg and New York the motor-road hastily joins the railroad and runs beside it for a quarter of a mile, so as to shrink away from a certain desolate area of land. This is a valley of ashes--a fantastic farm where ashes grow like wheat into ridges and hills and grotesque gardens where ashes take the forms of houses and chimneys and rising smoke and finally, with a transcendent effort, of men who move dimly and already crumbling through the powdery air. Occasionally a line of grey cars crawls along an invisible track, gives out a ghastly creak and comes to rest, and immediately the ash-grey men swarm up with leaden spades and stir up an impenetrable cloud which screens their obscure operations from your sight.

But above the grey land and the spasms of bleak dust which drift endlessly over it, you perceive, after a moment, the eyes of Doctor T. J. Eckleburg. The eyes of Doctor T. J. Eckleburg are blue and gigantic--their retinas are one yard high. They look out of no face but, instead, from a pair of enormous yellow spectacles which pass over a nonexistent nose. Evidently some wild wag of an oculist set them there to fatten his practice in the borough of Queens, and then sank down himself into eternal blindness or forgot them and moved away. But his eyes, dimmed a little by many countless days under sun and rain, brood on over the solemn dumping ground.

The valley of ashes is bounded on one side by a small foul river, and when the drawbridge is

Research ID #: _____

up to let barges through, the passengers on waiting trains can stare at the dismal scene for as long as half an hour. There is always a halt there of at least a minute and it was because of this that I first met Tom Buchanan's mistress.

PROMPT: What are the specific details that you notice about the setting in this passage? Based on these details, what inferences can you draw about the character(s) associated with that setting? *Your response will have two parts: what I see, and what I infer. Start the first part of your response with, "In this passage, I see..." Start the second part of your response with, "Therefore, I infer that..." Try to include as many observations and inferences as you can, and make sure that your observations and inferences are clearly connected. Type your response in Schoology in the special course for Mrs. Neebe's study.*

Chapter 3, Page 40

Please annotate the passage below, marking the specific details in the text that illustrate the setting and that help you make inferences about the characters associated with this setting.

At least once a fortnight a corps of caterers came down with several hundred feet of canvas and enough colored lights to make a Christmas tree of Gatsby's enormous garden. On buffet tables, garnished with glistening hors-d'oeuvre, spiced baked hams crowded against salads of harlequin designs and pastry pigs and turkeys bewitched to a dark gold. In the main hall a bar with a real brass rail was set up, and stocked with gins and liquors and with cordials so long forgotten that most of his female guests were too young to know one from another.

By seven o'clock the orchestra has arrived--no thin five-piece affair but a whole pitful of oboes and trombones and saxophones and viols and cornets and piccolos and low and high drums. The last swimmers have come in from the beach now and are dressing upstairs; the cars from New York are parked five deep in the drive, and already the halls and salons and verandas are gaudy with primary colors and hair shorn in strange new ways and shawls beyond the dreams of Castile. The bar is in full swing and floating rounds of cocktails permeate the garden outside until the air is alive with chatter and laughter and casual innuendo and introductions forgotten on the spot and enthusiastic meetings between women who never knew each other's names.

The lights grow brighter as the earth lurches away from the sun and now the orchestra is playing yellow cocktail music and the opera of voices pitches a key higher. Laughter is easier, minute

Research ID #: _____

by minute, spilled with prodigality, tipped out at a cheerful word. The groups change more swiftly, swell with new arrivals, dissolve and form in the same breath--already there are wanderers, confident girls who weave here and there among the stouter and more stable, become for a sharp, joyous moment the center of a group and then excited with triumph glide on through the sea-change of faces and voices and color under the constantly changing light.

PROMPT: What are the specific details that you notice about the setting in this passage? Based on these details, what inferences can you draw about the character(s) associated with that setting? *Your response will have two parts: what I see, and what I infer. Start the first part of your response with, "In this passage, I see..." Start the second part of your response with, "Therefore, I infer that..." Try to include as many observations and inferences as you can, and make sure that your observations and inferences are clearly connected. Type your response in Schoology in the special course for Mrs. Neebe's study*

Chapter 5, Page 92

Please annotate the passage below, marking the specific details in the text that illustrate the setting and that help you make inferences about the characters associated with this setting.

His bedroom was the simplest room of all--except where the dresser was garnished with a toilet set of pure dull gold. Daisy took the brush with delight and smoothed her hair, whereupon Gatsby sat down and shaded his eyes and began to laugh.

"It's the funniest thing, old sport," he said hilariously. "I can't--when I try to----"

He had passed visibly through two states and was entering upon a third. After his embarrassment and his unreasoning joy he was consumed with wonder at her presence. He had been full of the idea so long, dreamed it right through to the end, waited with his teeth set, so to speak, at an inconceivable pitch of intensity. Now, in the reaction, he was running down like an overwound clock.

Recovering himself in a minute he opened for us two hulking patent cabinets which held his massed suits and dressing-gowns and ties, and his shirts, piled like bricks in stacks a dozen high.

"I've got a man in England who buys me clothes. He sends over a selection of things at the beginning of each season, spring and fall."

He took out a pile of shirts and began throwing them, one by one before us, shirts of sheer linen and thick silk and fine flannel which lost their folds as they fell and covered the table in many-colored disarray. While we admired he brought more and the soft rich heap mounted

Research ID #: _____

higher--shirts with stripes and scrolls and plaids in coral and apple-green and lavender and faint orange with monograms of Indian blue. Suddenly with a strained sound, Daisy bent her head into the shirts and began to cry stormily.

PROMPT: What are the specific details that you notice about the setting in this passage? Based on these details, what inferences can you draw about the character(s) associated with that setting? *Your response will have two parts: what I see, and what I infer. Start the first part of your response with, "In this passage, I see..." Start the second part of your response with, "Therefore, I infer that..." Try to include as many observations and inferences as you can, and make sure that your observations and inferences are clearly connected. Type your response in Schoology in the special course for Mrs. Neebe's study*

Appendix F

Analytical Reading Comprehension Rubric

	5 <i>A Model Response</i>	4 <i>Solid Response</i>	3 <i>Meets Expectations</i>	2 <i>Basic Response</i>	1 <i>Misses the Mark</i>
Textual References (What I See)	Response provides a diversity of high-quality relevant textual references which support the inferences by showing rather than telling.	Response provides relevant textual references which support the inferences by showing rather than telling.	Response provides textual references which support most of the inferences; some evidence may be too surface-level, or tell rather than show.	Textual references are present, but rely entirely on telling rather than showing, simply summarizing the text.	Textual references are unclear, irrelevant, or missing.
Analytical Thinking (What I Infer)	Response presents an original, insightful analysis of the passage which demonstrates the student's ability to read and interpret a complex text deeply and perceptively. Specifically, the analysis shows a sophisticated understanding of the author's use of setting, what the setting reveals about character, and how the author communicates that message.	Response presents a clear and thorough analysis of the passage which demonstrates the student's ability to move beyond the obvious by considering the implications of subtext. Specifically, the analysis presents a thoughtful discussion of what setting suggests about character, without developing precisely how that message is conveyed.	Response may not demonstrate a full understanding or awareness of the passage and its implications for character development, but key details in the passage are present and analyzed. Specifically, the analysis may make leaps in thinking from setting to character that are not sufficiently explored.	Response tends to scratch the surface of the passage by being too general or simplistic. Specifically, the analysis may simply summarize the text without conveying its message.	Response indicates that the writer does not understand the text. Specifically, the analysis reveal gaps or inaccuracies in the writer's understanding of the setting and its greater significance in developing character.
Broader Implications (So What?)	Response demonstrates writer's ability to accurately situate the discussion of the selected passage within the greater thematic context of the novel as a whole.	Response links to another passage or part of the novel; the writer mentions a thematic idea, but does not fully develop this line of thinking.	Response draws connections to character development or other parts of the novel, but these connections do not lead to a discussion of thematic ideas or the broader implications of the novel.	Response attempts to make connections between the passage and character development, but these connections may be a bit obvious or forced.	Response does not discuss anything outside the passage itself.
TOTAL COMP SCORE					

Appendix G

Mental Effort Rating Scale for Adolescent Analytical Reading

THE GREAT GATSBY

Gatsby Chapter 1 Survey

Please focus your responses on the passage you just analyzed...

* Required

Email address *

Your email

Overall...

How much brain power did it take for you to understand the passage you analyzed? *

1 2 3 4 5 6 7
Very Little Very Much

Barriers to Engagement...

How much did you find yourself zoning out while you were supposed to be analyzing the passage? *

1 2 3 4 5 6 7
Very Little Very Much

How much did you find yourself re-reading the passage because you zoned out? *

1 2 3 4 5 6 7
Very Little Very Much

How much were you distracted by the classroom environment (climate, noise, classmates) ? *

1 2 3 4 5 6 7
Very Little Very Much

How much were you distracted by personal factors (hunger, exhaustion, stress, drama)? *

1 2 3 4 5 6 7
Very Little Very Much

How much did you find yourself worrying about not understanding the passage? *

1 2 3 4 5 6 7
Very Little Very Much

Once you're engaged in the passage...

How much thinking did you have to do to make sense of the vocabulary in the passage? *

1 2 3 4 5 6 7

Very Little Very Much

How much thinking did you have to do to make sense of the complex sentences in the passage? *

1 2 3 4 5 6 7

Very Little Very Much

How much thinking did you have to do to follow what happened in the plot of the passage? *

1 2 3 4 5 6 7

Very Little Very Much

How much thinking did you have to do to keep track of character relationships in the passage? *

1 2 3 4 5 6 7

Very Little Very Much

How much thinking did you have to do to understand the significance of the literary elements (symbols, motifs, theme, etc) in the passage? *

1 2 3 4 5 6 7

Very Little Very Much

How much did the combination of all the factors above (vocabulary, complex sentences, plot, character relationships, literary elements) challenge your understanding of the passage? *

1 2 3 4 5 6 7

Very Little Very Much

How much thinking did you have to do to understand the significance of the passage in connection to the novel as a whole? *

1 2 3 4 5 6 7

Very Little Very Much

How much did you re-read to deepen your understanding of the passage (not because you zoned out)? *

1 2 3 4 5 6 7

Very Little Very Much

Please enter your identifying NUMBER for the research study *

Your answer _____

SUBMIT

Fin.