ABSTRACT

MUDRICK, NICHOLAS VINCENT. The Influence of Multimedia Discrepancies on Metacomprehension: Evidence from Eye-Movements. (Under the direction of Dr. Roger Azevedo.)

This study investigated the influence of multimedia discrepancies on participants’ metacognitive judgments and eye-movements when learning with multimedia material (text and graphs). A within-subjects design was used to examine the influence of discrepancies on the accuracy of responses to inference questions, study-time allocation, metacognitive judgments, and eye-movements. There were three types of discrepancies: none, text (located between two different sentences of the text), and text and graph (discrepancy located between a sentence in the text and the entire graph). Discrepancies were defined as information within the text or between the text and graphs that directly contradicted each other. Fifty-six (N = 56) participants completed 12 trials where they were asked to provide six metacognitive judgments (ease of learning judgments [EOLs], immediate and delayed text and graph judgments of learning [JOLs], and retrospective confidence judgments [RCJs]) and answers to inference questions. A subset (n = 30) of participants’ fixations, proportion of time spent viewing, and revisits to areas of interest (AOIs; i.e., the text, graph, and inference questions) were collected. Overall, the accuracy of responses to inference questions was significantly influenced by discrepancy type. Participants’ study-time allocation was significantly longer, graph JOLs were significantly lower, the number of revisits was significantly higher, and the proportion of time viewing graphs was significantly higher for content with text and graph discrepancies. Overall, results suggest that participants’ metacognitive judgments and eye-movements are key to understanding metacomprehension with multimedia content.
The Influence of Multimedia Discrepancies on Metacomprehension: Evidence from Eye-Movements

by
Nicholas Vincent Mudrick

A thesis submitted to the Graduate Faculty of North Carolina State University in partial fulfillment of the requirements for the Degree of Master of Science Psychology

Raleigh, North Carolina 2015

APPROVED BY:

Dr. John Nietfeld
Dr. Anne McLaughlin

Dr. Roger Azevedo
Chair of Advisory Committee
DEDICATION

To Leanne.
BIOGRAPHY

Nicholas Vincent Mudrick is a graduate student and research assistant in the Human Factors and Applied Cognition Program at North Carolina State University. Mudrick received his undergraduate degree in Psychology with minors in Behavioral Sciences and Sexual Diversity Studies from McGill University in 2013. During his undergraduate career, Mudrick worked for Dr. Roger Azevedo as an undergraduate research assistant in the Laboratory for the Study of Metacognition and Advanced Learning Technologies (SMART Lab) assisting in data collection. His experience working at the SMART Lab inspired him to follow Dr. Roger Azevedo to pursue his doctorate at North Carolina State University.
ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor, mentor, and friend Dr. Roger Azevedo. Without his guidance, support, tutelage, and feedback, this thesis would never have been possible. His support and encouragement are second to none, and I could not have asked for a better mentor. Under his supervision and with his help, I have far exceeded my own expectations and have been able to accomplish things I never could have dreamed possible. I owe almost all of my accomplishments to Dr. Azevedo’s devotion to my own professional development, and I will never be able to thank him enough. I would also like to thank the members of my advisory committee, Drs. Anne McLaughlin and John Nietfeld, for their comments, input, and incredibly valuable time. I am very grateful to have them be a part of this study.

I would also like to thank my fellow graduate research assistants, who have assisted and supported me throughout this study. More specifically, thank you Michelle Taub for all of your comments, edits, insight, and friendship. Thank you Seth Martin for your assistance in helping to direct data collection, general life advice, and as always, your wonderful stories. I would also like to thank Garrett Millar for his assistance in helping me code the inference question responses and running participants. I would also like to thank Gina Kabat, Jeffery Matthews, Tristan Smith, and Haleigh Meffert for their assistance in running participants—without your help, there would be no data to be analyzed for this thesis.
Thank my parents who have instilled within me the importance of education, the love of learning, and the drive to succeed. I could not have made it here without all of the sacrifices they have made and the support they have given me. I would like to thank my girlfriend, Leanne Jones, for all of her love, support, encouragement, and most of all, patience.

Lastly, I would like to acknowledge financial support and funding for my graduate research assistantship at NCSU throughout my Master’s degree provided by Dr. Azevedo’s grants from the National Science Foundation, Institute of Education Sciences, and the Social Sciences and Humanities Research Council of Canada.
TABLE OF CONTENTS

LIST OF TABLES ................................................................................................................. ix
INTRODUCTION .................................................................................................................. 1
LITERATURE REVIEW ........................................................................................................ 6
  Comprehension Regulation .............................................................................................. 6
  Metacomprehension and Metacognitive Monitoring ....................................................... 7
  Multimedia Learning Research, Metacomprehension, and Eye-Movements ............... 14
  Theoretical Frameworks ................................................................................................. 20
  Purpose of Current Study ............................................................................................... 24
  Research Questions ......................................................................................................... 24
METHOD ............................................................................................................................... 26
  Participants ..................................................................................................................... 26
  Experimental Design ..................................................................................................... 27
  Materials ......................................................................................................................... 27
  Apparatuses .................................................................................................................... 32
  Experimental Procedure ................................................................................................. 33
  Coding and Scoring ........................................................................................................ 37
RESULTS ............................................................................................................................ 43
DISCUSSION ......................................................................................................................... 55
  Research Question 1 ...................................................................................................... 56
  Research Question 2 ...................................................................................................... 56
  Research Question 3 ...................................................................................................... 57
  Research Question 4 ...................................................................................................... 62
  Research Question 5 ...................................................................................................... 63
  Research Question 6 ...................................................................................................... 65
  Limitations of Current Study ......................................................................................... 66
  Contributions to Knowledge and Future Directions ..................................................... 68
REFERENCES .................................................................................................................. 77

APPENDICES .................................................................................................................. 86

Appendix A. Recruitment Flyer ...................................................................................... 87
Appendix B. Consent Form ............................................................................................ 88
Appendix C. Demographic Questionnaire ..................................................................... 90
Appendix D. Science Knowledge Pretest ........................................................................ 92
Appendix E. Pretest Related to the Science Content Discrepancies ............................ 96
Appendix F. Graph Comprehension Pretest .................................................................. 99
Appendix G. Perceived Interest Questionnaire (PIQ) .................................................. 104
Appendix H. Example Slide with No Discrepancy ....................................................... 105
Appendix I. Example Slide with Text Discrepancy Emphasized ................................. 106
Appendix J. Example Slide with Text and Graph Discrepancies Emphasized ........... 107
Appendix K. Example of Experimental Setup ................................................................ 108
Appendix L. Illustration of Experimental Procedure .................................................... 109
Appendix M. Example of an Ease of Learning (EOL) Judgment Layout .................... 110
Appendix N. Example Slide as Presented to Students with Attention Tool 5.4 .......... 111
Appendix O. Example of an Immediate and Delayed Text-Based Judgment of Learning (JOL) Layout ................................................................................................. 112
Appendix P. Example of an Immediate and Delayed Graph-Based Judgment of Learning (JOL) Layout ................................................................................................. 113
Appendix Q. Example of a Response Input Layout ....................................................... 114
Appendix R. Example Retrospective Confidence Judgment (RCJ) Layout ................. 115
Appendix S. List of Concept Units ................................................................................ 116
Appendix T. Example of Drawn Areas of Interest (AOIs) for a No Discrepancy Slide ................................................................................................................................ 118
Appendix U. Example of Drawn Areas of Interest (AOIs) for a Text Discrepancy Slide ................................................................................................................................ 119
Appendix V. Example of Drawn Areas of Interest (AOIs) for a Text and Graph Discrepancy Slide ................................................................................................................................ 120
Appendix W. Correlation Matrix with Eye-Movement Variables and Metacognitive Judgments..............................................................121
LIST OF TABLES

Table 1  Demographic Characteristics of Participants in Sample (N = 56)..........................73

Table 2  Participants’ Percentage Scores on the Pretests .................................................73

Table 3  Pearson Product–Moment Correlations among Metacognitive Judgments, Response Scores, and Study-Time Allocation.................................................................74

Table 4  Means, Standard Deviations and ANOVA Results for Six Metacognitive Judgments, by Condition for Univariate Analyses.................................................................75

Table 5  Means, Standard Deviations and ANOVA Results for the Amount of Fixations on, the Proportion of Time Spent Viewing, and the Amount of Revisits to the Text, Graph, and Inference Question, by Condition for Univariate Analyses (n = 30)....76
INTRODUCTION

Self-regulated learning (SRL) involves students actively planning, monitoring, and strategizing as they engage in learning about complex material. A significant component to SRL is metacomprehension, which can be explained as the processes by which students monitor and judge their learning and understanding of the complex material (e.g., assessing how well they understand a particular phrase or science concept). In order to achieve higher learning outcomes and succeed in learning about complex topics, students need to effectively self-regulate their learning (Azevedo, 2014; Winne, 2011).

To engage in successful SRL, students must employ certain cognitive and metacognitive processes relating to planning (e.g., activation of prior knowledge), monitoring (e.g., judgments of learning, time monitoring), and selection and enactment of appropriate learning strategies (e.g., summarization and coordination of information sources; Azevedo & Feyzi-Behnagh, 2010; Winne & Hadwin, 2008; Zimmerman & Schunk, 2011). More specifically, research has suggested that metacognitive monitoring and judgments (i.e., metacomprehension) are important in students’ study-time allocation, especially to effectively control their own learning (Burkett & Azevedo, 2012; Dunlosky, Hertzog, Kennedy, & Thiede, 2005; Dunlosky & Metcalfe, 2009; Metcalfe, 2009; Metcalfe & Finn, 2008; Nietfeld & Schraw, 2002). However, research has shown that students’ learning is seldom self-regulated as they engage with multimedia science content (Azevedo et al., 2013).
The results of the Program for International Student Assessment (PISA) showed that 15-year-old students in the United States are outperformed in science literacy and reading comprehension assessments by students in other countries (OECD, 2013). As of 2012, 22 countries scored higher than the United States in science literacy, and 19 countries scored higher than the United States in reading comprehension. In both areas, U.S. students’ scores matched the international average. These results extend to older students as the National Assessment of Educational Progress (National Center for Educational Statistics, 2014) collected information from 1970 to 2013 and indicated that nationwide, 17-year-olds showed little to no improvement in reading comprehension. Seventy-four percent of 17-year-olds were performing at or above a “basic” reading level, or only having a partial mastery of fundamental reading skills. The Business for Educational Success and Transformation (BEST-NC) report (2015), which provides statistics on educational achievement in North Carolina, noted that only 23% of high school graduates met the science college readiness benchmark, and only 30% met the reading benchmark in 2014. On a smaller scale, results from the Nation’s Report Card (National Center for Education Statistics, 2013) indicated that fourth- and eighth-grade students in North Carolina did not significantly differ from the national average in reading comprehension. The report indicated no changes in scores for fourth graders and only a small improvement for eighth graders from previous years’ assessments (National Center for Educational Statistics, 2013). These results indicate stagnation and a “crisis of mediocrity” (BEST-NC, 2015), emphasizing no educational
performance growth for students in the United States. The question becomes: if U.S. students’ science literacy and reading comprehension scores are at or below average, how will those graduates excel nationally and internationally in science, technology, education, and mathematics (STEM) careers?

Therefore, it is important to consider metacomprehension (e.g., monitoring and judgments of learning), as it is a significant contributor to academic achievement and success when learning about complex science topics with multimedia content and it warrants further study (Dunlosky & Lipko, 2007). Much research in metacomprehension has been devoted to metacognitive judgments, such as judgments of learning. Typically, these studies have participants predict how well they will do on an outcome assessment when presented with laboratory-contrived stimuli (e.g., paired associates; Hertzog, Hines, & Touron, 2013; Koriat, 1997; Koriat, Nussinson, & Ackerman, 2014; Nelson & Dunlosky, 1991; Pyc, Rawson, & Aschenbrenner, 2014). More focused research has involved using educationally relevant stimuli to examine metacomprehension with multimedia science materials, such as content from school textbooks, as well as online science materials (Burkett & Azevedo, 2012; Jaeger & Wiley, 2014; Pilegard & Mayer, 2015). This work has also extended traditional metacomprehension research through the inclusion of discrepancies within the multimedia content (e.g., between text and graph). Most of this research has focused on cognitive and metacognitive processes, and although it has advanced our understanding of key metacomprehension processes (e.g., the negative impact of illustration type on participants’
judgments of learning, cf. Jaeger & Wiley, 2014), it has not addressed whether the presence of discrepancies in multimedia science content, shown to hinder participants’ metacognitive judgments (Burkett & Azevedo, 2012), might also influence their affective responses (e.g., inducing confusion or frustration). These affective responses could in turn be associated with the accuracy of participants’ metacognitive judgments, deployment of cognitive strategies, and study-time allocation.

Furthermore, the processes underlying how these discrepancies influence metacognitive judgments as participants interact with multimedia science content have yet to be investigated. Specifically, is the presence of the discrepancies even recognized, or are participants left largely unaware of the discrepant information, as suggested by Burkett and Azevedo (2012)? The inclusion of eye-tracking measures could provide a more comprehensive description of how the presence of discrepancies affects participants’ metacomprehension as they interact with multimedia science content.

As such, the area of metacomprehension has largely focused on cognitive and metacognitive processes without considering the role of affective states (e.g., confusion, frustration) that may be induced by cognitive factors (e.g., detecting discrepancies within and across multiple representations of multimedia science content) and that can differentially affect metacognitive judgments (e.g., ease of learning judgments [EOLs], judgments of learning [JOLs], and retrospective confidence judgments [RCJs]).
In this study, we aimed to address some of these issues through the inclusion of affect measurements (e.g., facial expressions of affective states and electrodermal activity [EDA]) and eye-tracking. We also used more complex, educationally relevant content such as longer texts and diagrams. In addition to using the traditional indices of metacognitive judgments, such as EOLs, JOLs, and RCJs, we incorporated trace data by collecting online measures of metacognition, as well as the affective states associated with these metacognitive judgments. In doing so, we adopted several novel interdisciplinary methods, such as eye-tracking, facial expressions of emotions classification, and EDA measures, to assess the role of affective states and eye-movements during metacomprehension as participants learn from multimedia science content. As such, this study augments current empirical research on metacomprehension and multimedia by using tools that give us access to the cognitive, affective, and metacognitive processes at multiple levels.¹

¹ Although facial expressions of emotions and physiological data were collected, they were not analyzed for this study.
LITERATURE REVIEW

The first part of the literature review briefly describes traditional comprehension regulation research, followed by a more detailed discussion of metacomprehension and metacognitive judgments such as EOLs, JOLs, and RCJs. The second part of the literature review covers metacomprehension and multimedia research. This will be followed by a description of methodology used in multimedia learning research and a discussion of research findings on eye-movements and multimedia learning.

Comprehension Regulation

Traditionally, comprehension regulation research has focused on short texts (e.g., a six-sentence paragraph, cf. Otero & Companario, 1990), containing contradictions or discrepancies and has shown that learners seldom detect or locate these contradictions in text (Otero & Companario, 1990; Otero & Kintsch, 1992; Otero, León, & Graesser, 2002). However, this line of research has involved placing the contradictory sentences in the same location of the stimuli (e.g., the last sentence; Otero & Companario, 1990). Additionally, this research has rarely specified the amount of time given to participants to detect these discrepancies (Otero & Companario, 1990; Otero & Kintsch, 1992; Otero et al., 2002). As such, results from this research might not be indicative of how participants might detect discrepant information in a more natural setting (Burkett & Azevedo, 2012). An example of such a setting would be a Wikipedia page about a particular science concept.
Using a modified comprehension experimental paradigm, Burkett and Azevedo (2012) administered multimedia science content to participants that contained either no discrepancies or different types of discrepancies (i.e., two sentences in different paragraphs, and between a sentence in the text and the entire graph). They allowed participants as much time as needed to study the content. Higher JOLs indicated that participants perceived they better understood content that contained no discrepancies than content that contained discrepancies within the text or between the text and corresponding graph (Burkett & Azevedo, 2012). However, participants did not spend more time on the multimedia science content with discrepancies in comparison to the nondiscrepant content. These results suggest that participants realize their understanding is lower for information containing discrepancies, but are unaware of what specifically has hindered their comprehension (Burkett & Azevedo, 2012). The results from these studies might have been augmented by the inclusion of eye-tracking measures, which could indicate whether participants detected and processed the discrepant information presented within the multimedia content. Participants might have indeed recognized the discrepancy but were not able to resolve it (e.g., García-Rodicio & Sánchez, 2014).

**Metacomprehension and Metacognitive Monitoring**

Accurate judgment of one’s comprehension of learned material is fundamental to learning (Dunlosky & Lipko, 2007). Specifically, an accurate judgment of what has or has not been learned is important in directing attention and allocating time to study content that is
poorly understood (Dunlosky & Metcalfe, 2009; Metcalfe, 2008). Therefore, it is imperative to understand the processes by which students understand and judge their own understanding, as they underlie critical components of successful learning. Research on metacognition and metacomprehension has typically focused on three types of judgments: EOLs, JOLs, and RCJs (Dunlosky & Lipko, 2007; Metcalfe, 2009; Nelson & Narens, 1990).

Nelson and Narens’ (1990) framework of metamemory indicates that EOLs, JOLs, and RCJs occur at specific time points throughout a learning session. EOLs normally take place before learning and ask learners to judge how easy or difficult the content will be to learn. JOLs occur alongside learning, but usually at specific points (e.g., after the target material has been read), and ask the participants to assess how well they understand the presented content. Additionally, research suggests that delaying JOLs increases their accuracy in comparison to those made immediately or alongside learning (Metcalfe & Finn, 2008; Nelson & Dunlosky, 1991). Lastly, RCJs occur after the learning session (or trial) has been completed and ask participants to gauge their confidence in the answers they provided. As such, for this study, EOLs were made prior to the beginning of each trial, immediate and delayed (by 30 seconds) JOLs were made for both the text and diagram during each trial, and RCJs were made for the inference questions presented with the multimedia science content following each trial. The following sections will discuss the research regarding each of these metacognitive judgments in more detail.
Ease of learning judgments (EOLs). EOLs usually occur before learning and call for participants to evaluate the ease or difficulty of the content they are about to learn. EOLs are an important metacognitive judgment; if the expected material is regarded as difficult, the participant should spend more time studying that content. Nelson and Leonesio (1988) found that in self-paced study trials, participants studied longer on items where their EOLs indicated they perceived the content as more difficult. Thiede, Anderson, and Therriault (2003) extended Nelson and Leonesio’s (1988) findings, when they determined EOLs impact participants’ study-time allocation for more difficult content. Participants were shown the titles of six expository texts covering various topics, asked to judge how easy the information would be to learn, and then asked to read the six texts and answer questions related to the texts. After being given general feedback on their performance, participants were given the opportunity to restudy selected texts. Results indicated that participants who made more accurate EOLs chose to restudy the content they scored poorly on. Although highlighting that accurate EOLs are inherent to appropriate learning strategy deployment (i.e., study-time allocation), this study did not include multimedia materials and solely used expository texts as their stimuli. To address this, Burkett and Azevedo (2012) asked participants to judge the ease of learning with multimedia science content. Participants were asked to make EOLs after reading inference questions before being presented with multimedia science content (e.g., text and graph) containing discrepancies. As predicted, due to the nature of EOLs being made before participants interacted with the discrepant multimedia content, the discrepancies
had no effect on their EOL judgments. As the researchers did not provide the opportunity to restudy the discrepant multimedia content, it is difficult to infer the impact that EOLs would have had on the participants’ propensity to restudy the discrepant information.

As such, the current study attempts to augment previous literature by associating eye-movements and metacognitive judgments. As shown by Thiede et al. (2003), EOLs are significantly associated with the deployment of appropriate learning strategies such as study-time allocation. Examining the subsequent eye-movement behavior following an EOL could indicate the type of participants’ study choice strategies (e.g., integrating information between the text and graph). Participants might fixate on and have multiple revisits to different components on the multimedia content based on the accuracy of their EOL judgments.

**Judgments of learning (JOLs).** JOLs can occur at any point during the learning session, but are usually made at specific points (e.g., after reading a specified amount of material such as multiple paragraphs) and assess the likelihood that participants will be able to recall the recently learned information. Research in the area of metacognitive judgments has predominantly focused on JOLs. More specifically, it has attempted to understand the underlying processes contributing to these judgments in addition to studying factors in relation to improved judgment accuracy (Dunlosky & Lipko, 2007; Koriat, 1997; Koriat, Ackerman, Adiv, Lockl, & Scheider, 2014; Pyc et al., 2014).
Research has indicated that participants utilize multiple contributing cues in making JOLs. Koriat (1997) outlined a cue-utilization framework, which states that three types of cues can influence JOLs: intrinsic, mnemonic, and extrinsic. Intrinsic cues are characteristics representative of how easy or difficult an item is to learn (e.g., familiarity vs. unfamiliarity of the item). Mnemonic cues are based on the participant’s subjective experience during the time of item encoding. Extrinsic cues are situated in the encoding task itself (i.e., paired-associates learning, sentence generation) or the learning environment (e.g., number of trials to study the items). All three cues have been determined to be used during a JOL and impact the accuracy of a participant’s JOLs throughout learning paradigms (e.g., paired associates, criterion learning; Koriat, 1997; Koriat et al., 2014; Pyc et al., 2014). The Koriat (1997) framework suggests that intrinsic cues are utilized early in the learning process, and that participants shift toward utilizing mnemonic cues with more practice (i.e., studying paired associates). Though there exists substantial support for the cue-utilization framework (cf. Koriat & Ackerman, 2010; Serra & Ariel, 2014; Soderstrom, Clark, Halamish, & Bjork, 2015), research has shown that participants’ JOLs are consistently inaccurate (Dunlosky & Lipko, 2007; for a review, see Maki, 1998).

**Delayed judgments of learning (JOLs).** Nelson and Dunlosky (1991) introduced the concept of delaying JOLs in an attempt to increase judgment accuracy. This concept rests on the monitoring-dual-memories (MDM) principle, which suggests that delaying JOLs increases their accuracy, such that the to-be-judged item in short-term memory is no longer...
interfering with participants’ ability to monitor whether the information is stored in their long-term memory (Nelson & Dunlosky, 1991). In their experiment, participants were asked to study word pairs (e.g., chateau-castle), and were then asked to make a judgment about how likely they would recall the word pair on a later test. After immediately making their JOL, participants were also prompted to make a JOL following a delay of 30 seconds. Goodman-Kruskal gamma correlations indicated that accuracies of immediate JOLs were fairly low (Goodman-Kruskal gamma = +0.38) and that accuracy significantly increased following the 30-second delay (Goodman-Kruskal gamma = +0.90). The results of this study have been consistently replicated across groups and experimental paradigms, and have become well known within the metacognitive monitoring literature as the delayed JOL effect (Dunlosky & Metcalfe, 2009; Koriat et al., 2014; Metcalfe & Finn, 2008; Pyc et al., 2014; Son & Metcalfe, 2005). These findings are significant for the current study as they highlight the difference between immediate and delayed JOL accuracies, with delayed judgments being more accurate.

**Retrospective confidence judgments (RCJs).** RCJs usually occur after participants have provided a response to an item on a learning outcome measure (e.g., a multiple choice or free response question) and ask the participants how confident they are in their response (Dunlosky & Bjork, 2008; Dunlosky & Metcalfe, 2009; Nietfeld, Cao, & Osborne, 2005). When compared to actual performance, research has indicated that these comprehension judgments are largely inaccurate (Dunlosky & Metcalfe, 2009). Koriat and Goldsmith (1996)
investigated the relationship among standard or misleading question type, question response, and confidence judgments. Results indicated that participants consistently reported higher confidence scores for their responses to the misleading questions in comparison to the standard questions. Results suggested that participants were also unaware that their answers were incorrect when responding to both standard and misleading questions. In addition to being consistently inaccurate, RCJs have been found to be stable across learning time and generalizable across assessments. For example, Mengelkamp and Bannert (2010) investigated the stability and predictive validity of RCJs throughout a learning session. Participants were asked to navigate a hypertext environment and read content related to operant conditioning. Participants made confidence judgments after taking a pretest (before the session began), an intermediate test (10 minutes after examining the content), a comprehension test, and final transfer tests (after the session). Results indicated that RCJs were consistently inaccurate throughout the learning session and that only the RCJ at the pretest significantly predicted learning outcomes. Taken together, these findings emphasize that participants were often unaware whether their answers were correct or incorrect (i.e., reporting consistent confidence judgments regardless of their actual understanding of the content) and continued to make inaccurate judgments throughout the learning session.

Though past research has provided insight on some of the processes underlying how participants make metacognitive judgments, it is still lacking in its reliance on inadequate experimental designs. More specifically, the majority of the research on metacognitive
judgments has used laboratory-contrived stimuli, with little attention on both the occurrence and measurement (e.g., eye-tracking) of these processes as participants interact with more educationally relevant material (e.g., multimedia content containing text and diagrams on a particular science concept). The following section will discuss the research examining the effects of multimedia content on metacognitive processes, in addition to research on eye-movements and multimedia learning.

**Multimedia Learning Research, Metacomprehension, and Eye-Movements**

**Metacomprehension and multimedia learning.** Research suggests multimedia learning is more effective than learning with text alone (Mayer, 2005, 2009, 2014a, b; Schnotz, 2014; Schnotz & Bannert, 2003) and also leads to better learning outcomes, such as higher retention (Schwepp, Eitel, & Rummer, 2015). With the literature predominantly examining the cognitive processes underlying multimedia learning (e.g., Eitel, Kühl, Scheiter, & Gerjets, 2014; Eitel, Scheiter, & Schüler, 2013; Eitel, Scheiter, Schüler, Nyström, & Holmqvist, 2013; Mayer, 2005, 2014a; Schnotz, 2014; Schnotz & Bannert, 2003), little attention has been paid to examining participants’ metacomprehension as they interact with multimedia content.

Serra and Dunlosky (2007) began examining the relationship between metacomprehension and multimedia learning by investigating the impact of content presentation (i.e., text only, text and graph) on the accuracy of participants’ EOLs and JOLs. Results indicated that for both EOLs and JOLs, participants in the text and graph condition
made higher and more accurate judgments compared to those in the text-only condition. Serra and Dunlosky (2010) replicated these findings and also found participants’ beliefs about the format of multimedia impacted their metacognitive judgments, such that participants’ judgments were higher for texts that included diagrams. These studies demonstrate the relationship between metacomprehension and multimedia learning, yet they do little to explicate the influence of specific multimedia components, such as graphs or diagrams, on participants’ metacomprehension.

More recently, two studies have addressed this by manipulating components of multimedia content (e.g., diagrams, graphs) to investigate the impact on participants’ metacognitive judgments. Burkett and Azevedo (2012) investigated the impact of multimedia content containing discrepancies (i.e., none, within text, between text and graph) on participants’ EOLs, JOLs, delayed JOLs, and RCJs metacognitive judgments, finding significant relationships between the metacognitive judgments and type of discrepancy. In addition, they found the type of discrepancy impacted each of these metacognitive judgments separately. Furthermore, Jaeger and Wiley (2014) manipulated the type of illustration presented with text to be either conceptual (i.e., informatively equivalent and redundant with the written text) or decorative (i.e., related to the text-based content, aesthetically pleasing but offering no relevant information to augment or explain the written content). Consistent with the aforementioned findings (e.g., Serra & Dunlosky, 2007, 2010), the presence of these illustrations impacted the accuracy of the participants’ metacognitive judgments. Results
indicated that the presence of conceptual illustrations did not always benefit the participants’ metacognitive judgments, as no significant differences were found between the conceptual and no-image conditions. However, the participants in the decorative illustration condition showed significantly lower metacomprehension accuracy in comparison to the conceptual and no-image conditions.

These studies suggest that the design of multimedia content significantly affects participants’ metacognitive judgments. However, examining participants’ eye-movements as they interacted with the multimedia content could explain what components of the content were most attended to in order to make increased metacognitive judgments, such as in the Serra and Dunlosky (2007, 2010) studies. Additionally, measuring participants’ eye-movements could indicate if participants detected the presence of discrepancies in Burkett and Azevedo (2012). Lastly, Jaeger and Wiley (2014) suggested the decorative images seduced the participants into relying too much on the images as a source for their comprehension judgments. Although this seems plausible, eye-movement data could have specified what elements of the illustrations were being attended to and processed by participants.

**Eye-movement and multimedia learning research.** Eye-tracking can enable inferences about the cognitive and metacognitive processes that occur when learning with multimedia (Mayer, 2010; van Gog, Jarodzka, Scheiter, Gerjets, & Paas, 2009). More specifically, analyzing eye-movements can specify what participants are attending to, and for
how long (cf. Mayer, 2010; Scheiter & van Gog, 2009; van Gog & Jarodzka, 2013; van Gog et al., 2009). This methodology has been applied to studying multimedia content and can provide information on variables such as fixations on areas of interest (AOIs; e.g., a word, paragraph, picture, graph), the amount of time spent on AOIs, and the proportion of time spent on AOIs in comparison to total study time.

The use of eye-tracking in multimedia is adapted from reading research (Rayner, 1998, 2009). As with research on reading processes, eye-tracking is seen to explicate the moment-to-moment attentional and cognitive processes that occur during comprehension of text-based stimuli such as expository texts (Rayner, Chase, Slattery, & Ashby, 2006). Specifically, the immediacy assumption suggests that the link between attention and cognitive processing occurs immediately, such that once the stimulus is perceived it is also cognitively processed (Just & Carpenter, 1980). This assumption has been applied to most research on reading processes and is now being extended to research on multimedia learning.

Traditionally, reading research has used different measures to distinguish among different types of cognitive processing (i.e., basic attentional vs. comprehension processes; Hyönä & Nurminen, 2006). For example, first-pass indicators are used to describe processes that occur during the first encounter of an AOI (e.g., initial fixations on the graph AOI) and are indicative of differences in visual salience or basic reading processes. Second-pass indicators are associated with more intentional cognitive processes, such as those related to comprehension or the integration of information, and are reflected by revisits to AOIs.
(Scheiter & Eitel, 2015). It is important to note that a majority of research on multimedia learning and eye-movements fails to distinguish between first- and second-pass indicators (Scheiter & Eitel, 2015). As such, in this study we differentiated between these indicators and addressed this gap in the current literature by examining both revisits and fixations as participants interacted with multimedia science content.

Though some studies have examined the overall effect of multimedia design on participants’ eye-movements (cf. Hegarty & Just, 1993; Mason, Pluchino, Tornatora, & Ariasi, 2013; Schmidt-Weigand, Kohnert, & Glowalla, 2010), and others have examined the impact of specific multimedia design principles in isolation (Bauhoff, Huff, & Schwann, 2012; Johnson & Mayer, 2012), most of the literature has examined the relationship between attention guidance and learning outcomes during learning with multimedia (Hegarty & Just, 1993; Jamet, 2014; Ozcelik, Arslan-Ari, & Cagiltay, 2010; Scheiter & Eitel, 2015). For example, Ozcelik et al. (2010) examined the impact of signaling, by changing the color of the label on the illustration, on participants’ transfer and matching tests as they learned about a turbofan jet engine. Results indicated that participants in the signaling group outperformed participants in the nonsignaling group on both transfer and matching tasks; however, both groups performed equally on the retention test. Eye-movement analyses revealed that participants attended to more relevant information when visually signaled, and the signaling also facilitated more efficient and effective visual searches for relevant information (Ozcelik et al., 2010). This finding has been extended by Scheiter and Eitel (2015), who highlighted
elements of text and diagrams and found that this signaling fostered better integration of the information presented in both the text and diagram. Results indicated that participants in the signaling condition outperformed those not signaled to specific components of the text and diagram on text–diagram integration questions.

Other research on eye-movements and multimedia learning has approached signaling participants’ attention through the use of eye-movement modeling examples (EMMEs; Jarodzka et al., 2012; Jarodzka, van Gog, Dorr, Scheiter, & Gerjets, 2013; Mason, Pluchino, Tornatora, & Ariasi, 2013; Mason, Pluchino, & Tornatora, 2013). This line of research entails showing participants an expert’s eye-movements as they learn with multimedia content, and results suggest participants shown these EMMEs indicate better integration of text and illustrations and higher learning outcomes (Mason, Pluchino, & Tornatora, 2013, 2015; Mason, Pluchino, Tornatora, & Ariasi, 2013). In sum, research on eye-movements and multimedia learning has predominantly focused on examining the relationship between attention guidance (e.g., signaling) and learning outcomes as participants interact with multimedia content.

Although the findings of these studies are important, research on eye-movements and multimedia learning is still limited in certain ways. Specifically, this research has mainly been situated within the cognitive processes (i.e., attention on text or diagram) with little focus on the metacognitive processes that occur as participants learn with multimedia. Additionally, much of this research has neglected to explicate the eye-movements that occur
during multimedia learning in the absence of attentional signaling. Furthermore, this literature has traditionally failed to distinguish between first- and second-pass indicators of eye-movements. As such, the goal of this study is to address the lack in the literature regarding the relationships among metacomprehension, eye-movements, and multimedia learning by examining the associations among these processes as participants interact with multimedia science content containing discrepancies.

**Theoretical Frameworks**

Given that the existing paradigm of metacomprehension research has traditionally used Nelson and Narens’ (1990) theoretical framework to examine learners’ metacognitive judgments, it is important to discuss other theoretical frameworks relevant to the current study, as its focus is to augment and contribute to the current metacomprehension literature by examining eye-movements as learners interact with multimedia content. Though Nelson and Narens’ framework will be used in this study, its scope is limited in its sole focus on metacognitive processes, with little discussion of the cognitive processes that occur alongside and as a result of metacognition. This is best exemplified by providing an example of a particular stimulus used in the current study.

Participants were asked to read an inference question regarding a specific science subject (e.g., “Why is the melting point of a substance considered to be a physical property rather than a chemical property?”) and then asked to provide an EOL before being presented with a multimedia science content slide containing text and graph (with no discrepancy, a
Following their interaction with the multimedia content, they were asked to make immediate JOLs for both the text and graph. They were then presented with a stop sign image, asked to wait 30 seconds and think about the content they just viewed, and then make delayed JOLs for the text and graph. Following this, participants provided an answer to the original inference question and then made an RCJ before moving on to the next trial.

Though no theoretical frameworks exist for eye-movements during multimedia learning, research on eye-movements during reading comprehension and multimedia learning can provide a basis on which to explain the aforementioned example (Hegarty & Just, 1993; Rayner, 2009; Scheiter & Eitel, 2015). When presented with the multimedia science content slide containing the inference question, text, and graph (with no discrepancy, a text discrepancy, or a text and graph discrepancy), participants would have first inspected the inference question, then read the text and inspected the graph. Specifically, the number of fixations and revisits and the amount of time spent viewing specific content have been implicated in the processing of information, where more fixations and revisits and longer amounts of time spent viewing specific content are indicative of increased processing demands (Rayner, 1998, 2009). As such, for slides with a text discrepancy, or a text and graph discrepancy, more time and more fixations should have been spent on the text, or the text and graph, respectively. However, results from research on eye-movements and multimedia learning also suggest that participants would be biased to spend more time
inspecting the text as opposed to a graph, in general (Hegarty & Just, 1993; Schmidt-Weigand et al., 2010). Additionally, saccades between and revisits to the text and graph would be made as participants attempted to integrate the information presented in the text and the graph (Hegarty & Just, 1993; Mason, Pluchino, & Tornatora, 2013; Mason, Pluchino, Tornatora, & Ariasi, 2013; Scheiter & Eitel, 2015; Schmidt-Weigand et al., 2010).

However, in order to interpret these eye-movements, they need to be aligned with other data channels (e.g., log-files, screen recordings) and theoretical frameworks (e.g., metamemory, cognitive theory of multimedia learning). In the next sections, Nelson and Narens’ (1990) theoretical framework of metamemory and Mayer’s (2005, 2014b) cognitive theory of multimedia learning will be discussed to articulate the assumptions related to the metacognitive and cognitive processes that are active as participants interact with multimedia science content containing discrepancies.

**Metamemory.** Nelson and Narens’ (1990) framework articulates the metacognitive judgments and control strategies, their temporal placement during learning, and their relationship to knowledge acquisition. More specifically, this framework illustrates the associations between metacognitive judgments (e.g., EOLs, JOLs, and RCJs) that have been traditionally studied in isolation (Burkett & Azevedo, 2012). In accordance with the framework’s temporal placement of metacognitive judgments during learning, participants’ EOLs were made prior to learning and were expected to influence not only the selection of content to study, but also the allocation of study time and effort (Metcalfe, 2009). The JOLs
that occurred alongside, and 30 seconds after, the learning task could have also impacted the participants’ subsequent selection of learning strategies, such as study-time allocation (Dunlosky & Metcalfe, 2009). Lastly, the RCJs that occurred after the information had been retrieved could have impacted the subsequent metacognitive judgments (e.g., EOLs and JOLs). Altogether, this framework emphasizes that these metacognitive processes are related and can influence each other (Dunlosky & Bjork, 2008; Dunlosky & Metcalfe, 2009). However, this framework does not articulate the cognitive processes that occur during learning with multimedia science content. As such, examining these processes will have to be situated within the cognitive theory of multimedia learning (Mayer, 2005, 2014b).

**Cognitive theory of multimedia learning.** Mayer’s (2005, 2014b) cognitive theory of multimedia learning (CTML) outlines three main assumptions: (1) different channels are used to process content in text versus in a diagram; (2) each of these channels is limited in the amount of information that can be processed at a given time; and (3) learners have an active role in multimedia learning. The CTML specifies two channels (visual and auditory), three types of memory (sensory, working, and long-term memory), and five cognitive processes (selecting words, selecting images, organizing words, organizing images, and integrating words and images). Mayer argues that learning from multimedia involves viewing content in both text and diagram form, the processing of which begins within sensory memory. Then the learner selects words and images to be processed and integrated in working memory, ultimately leading to a coherent cognitive representation. Prior knowledge
of the content can then be retrieved from long-term memory to help process the new information in working memory. Once processing is finished in the working memory, the newly consolidated knowledge can be transferred to the long-term memory for permanent storage. The model does not acknowledge the temporal nature of this processing, nor does it account for variation in the sequence. However, we expected that participants engaged in processing and integrating the multimedia science content (i.e., text and graph) at the sensory and working memory levels, respectively, and then integrated it with prior knowledge before permanent storage in their long-term memory.

**Purpose of Current Study**

This study intends to provide a better understanding of the relationship among accuracy of responses to inference questions, study-time allocation, metacognitive judgments, and eye-movements when learning with multimedia science content. The current study focused on multimedia content that contained an inference question, text, and graph. A within-subjects design was used to examine how discrepancy type (none, text discrepancy, and text and graph discrepancy) influenced the accuracy of responses to inference questions, study-time allocation, metacognitive judgments, and eye-movements.

**Research Questions**

To better understand the relationship among metacognitive judgments, and the cognitive processes (inferred from eye-movements) associated with learning multimedia content, the proposed study will seek to answer the following questions:
1. Does discrepancy type influence the accuracy of inference question responses during multimedia learning?

2. Does discrepancy type influence study-time allocation during multimedia learning?

3. Does discrepancy type influence metacognitive judgments during multimedia learning?

4. Does discrepancy type influence the number of fixations on the text, graphs, and inference questions presented during multimedia learning?

5. Does discrepancy type influence the proportion of time spent viewing the text, graphs, and inference questions presented during multimedia learning?

6. Does discrepancy type influence the number of revisits to the text, graphs, and inference questions during multimedia learning?
METHOD

Participants

Fifty-six (N = 56) undergraduate students enrolled at North Carolina State University participated in this study. Participants’ ages ranged from 18 to 27 (M = 20.39, SD = 1.88; see Table 1). Participants were recruited through posted advertisements (see Appendix A) and paid $20.00 for their participation in the study. The majority of participants were psychology majors (39.3%) and reported junior as their classification (37.5%). The majority (n = 52) of participants reported taking previous science courses (M = 3.30, SD = 2.74, range = 0 to 16) at the university. Additionally, the majority (n = 54) of participants reported taking previous university mathematics courses (M = 2.03, SD = 1.06, range = 0 to 5). Only 14.28% (n = 8) of the participants reported previous work experience in the fields related to chemistry, physics, biology, physical science, or math.

Scores from the science pretest and the discrepancy-related science pretest indicated participants had low prior knowledge about the science domains tested (see Table 2). Percentage scores on the 4-foil, 20-item multiple-choice science pretest ranged from 35% to 100% (M = 69.64%, SD = 14.23%). Additionally, results indicated that participants did not have prior knowledge about the science content presented in the slides with discrepancies. Percentage scores on the 4-foil, 12-item multiple-choice discrepancy-related pretest ranged from 17% to 92%, and at least half of the questions were answered incorrectly (M = 55.65%, SD = 16.21%), whereas percentage scores on the 4-foil, 12-item multiple-choice graph
comprehension pretest indicated participants were able to read and understand basic graph components ($M = 78.72\%, SD = 10.65\%, \text{range } = 25\% \text{ to } 92\%$).

**Experimental Design**

The study used a within-subjects design in which each participant experienced each of three discrepancy types: none, text (discrepancy located between two different sentences of the text), and text and graph (discrepancy located between a sentence in the text and the entire graph). A discrepancy was an instance where the information within the text or between the text and graph directly contradicted each other.

**Materials**

The materials used in this study included the following: an informed consent form (Appendix B), a demographic questionnaire (Appendix C), a test of basic science knowledge (Appendix D) (Burkett & Azevedo, 2012), a researcher-developed science knowledge pretest related to the discrepancies presented in the slides (Appendix E), as well as a test of basic graph knowledge (Appendix F) (Burkett & Azevedo, 2012). Additionally, the Perceived Interest Questionnaire (PIQ; Schraw, Bruning, & Svoboda, 1995) was administered to participants after they completed the study (see Appendix G). The informed consent form provided information about the experiment and informed participants that they had the option to terminate their participation at any point during the session. The demographic questionnaire asked participants for the following information: age, gender, university year, major, and any previous university-level courses and relevant work experience related to the
science domains of chemistry, physics, biology, physical science, and math. Additionally, the computer-based materials for this study consisted of 12 researcher-developed (Burkett & Azevedo, 2012) multimedia science content slides presented on iMotions Attention Tool 5.4 (iMotions, 2014), described in greater detail below.

**Basic science knowledge pretest.** Participants were given a researcher-developed test of basic science knowledge in order to assess individual differences in prior knowledge of chemistry, physics, biology, and physical science (from Burkett & Azevedo, 2012; see Appendix D). The test consisted of 20 4-foil items with 5 questions each for the above-mentioned science domains. The questions were taken from practice exams for standardized tests for college admissions, such as the SAT, ACT, and Trends in International Mathematics and Science Study (TIMSS) 2003, which released a set of questions for eighth-grade science programs (L. Callihan & D. Callihan, 2009; International Association for the Evaluation of Educational Achievement, 2003; SparkNotes, 2004, 2007). Difficulty levels for the questions taken from the TIMSS test were divided into three categories: “easy,” because more than 50% of eighth-grade students from the United States who completed the assessment got the answer correct; “moderate,” because approximately 50% of eighth-grade students got the answer correct; and “difficult,” because less than 50% got the answer correct. In summary, out of the 20 questions on the science pretest, 8 were from practice exams for standardized tests for college admission and 12 were from the TIMSS assessment (4 of which were easy, 4 of which were moderate, and 4 of which were difficult). A sample item from the pretest is
“Oxygen, hydrogen and water are substances. Which of these substances are elements?” with the correct answer being “oxygen and hydrogen only.” This questionnaire was administered using the survey hosting website, Qualtrics.com (Qualtrics, 2015).

**Discrepancy-related science knowledge pretest.** Participants were also given a researcher-developed test of science content related to the slides with discrepancies in order to determine if participants had any prior knowledge of the content presented in the 12 slides of the experiment (see Appendix E). The test consisted of 12 4-foil items. The questions were developed based on the inference questions presented on the slides during the experimental session. For example, if the inference question in the slide asked, “Why would you expect to find a higher level of relative humidity during the spring than during the summer?,” the pretest question related to that content asked, “Why would you expect to find a lower level of relative humidity during the summer than during the spring?” This questionnaire was administered using the survey hosting website, Qualtrics.com (Qualtrics, 2015).

**Graph comprehension pretest.** A researcher-developed (Burkett & Azevedo, 2012) test of graph comprehension was administered to confirm participants were familiar with basic components of graphs (e.g., axes, units of analysis, and trends; see Appendix F). The graph comprehension test consisted of 10 4-foil items and 2 questions that asked participants to indicate where plotted points were located. Five of the twelve items were included to assess participants’ familiarity with basic components of graphs (e.g., x-axis and y-axis). The other 7 questions were included to ascertain if participants were able to read and extract basic
conclusions from the presented graph (e.g., during what weeks did coffee sales increase). This questionnaire was administered using the survey hosting website, Qualtrics.com (Qualtrics, 2015).

**Perceived Interest Questionnaire (PIQ).** The PIQ was administered to students following the experimental session in order to ascertain participants’ perceived levels of interest in the science content they had just read. The PIQ was adapted for the purposes of this study to fit the content (see Appendix G). The PIQ is a 10-item questionnaire designed to assess participants’ level of situational interest in texts immediately after reading. Participants responded on a 5-point Likert scale by indicating how much they agreed or disagreed with each statement. The internal consistency of the PIQ is $\alpha = 0.91$. A sample item of the adapted PIQ is “I thought the science content was very interesting.” This questionnaire was administered following the completion of the study using the survey hosting website, Qualtrics (Qualtrics, 2015).

**Science content slides.** The computer-based materials for this study included 12 researcher-developed multimedia science content slides (from Burkett & Azevedo, 2012). The slides contained one science topic per slide across four science domains (chemistry, physics, biology, and physical science). Each slide contained an inference question for which the participant needed to provide an answer at a later stage in the experiment. Each slide contained text that ranged between 240–250 words ($M = 246.58$) and a corresponding graph that illustrated the particular concept that was described in the text. The content for each slide...
was drawn from four college textbooks from the four science domains of chemistry, physics, biology, and physical science (A. Getis, J. Getis, & Fellmann, 2009; Halliday, Resnick, & Walker, 2008; Hoefnagels, 2009; Masterton & Hurley, 2006; Tillery, 2007; for example slides, see Appendices H–J and Appendix N).

The slides were divided into three categories: no discrepancy, text discrepancy, and text and graph discrepancy. The content for the text in the no discrepancy slides was taken directly from the textbook (for an example slide with no discrepancy, see Appendix H). In the text discrepancy slides, the text was taken from the textbook and then altered to contain two sentences that directly contradicted one another. For example, in the slide about the components of atoms, one sentence of the text said that electrons are negatively charged particles, while another sentence indicated that electrons create a positive charge (for an example text discrepancy slide, see Appendix I). For the text and graph discrepancy slides, the text was taken from the textbook and a graph was created in Microsoft Office Excel 2013. The graphs in the text and graph discrepancy slides were created to illustrate a direct contradiction to a sentence in the text. For example, in the slide about the relationship between force and acceleration, a sentence in the text stated that there is a positive correlation between force and acceleration, while the graph illustrated a negative correlation between force and acceleration (see Appendix J). Despite the presence of the discrepancies, each slide contained the correct information needed to answer the specific inference question. The 12 slides were created using Microsoft Office PowerPoint 2013 to ensure that each slide used in
the study had consistent features (e.g., font, size, layout, etc.). The slides were saved as a PDF and uploaded into iMotions Attention Tool 5.4 (iMotions, 2014). The slides were counterbalanced and equally divided among the three discrepancy types, such that there were four slides with no discrepancy, four slides with a discrepancy between two sentences within the text, and four slides with a discrepancy between a sentence in the text and the entire graph.

**Apparatuses**

The content was presented using the iMotions Attention Tool 5.4 software developed by iMotions (2014). In addition to presenting the 12 slides of the 12 trials, Attention Tool 5.4 was used throughout the experimental session to collect the following: time spent in session, time spent on each slide, participants’ metacognitive judgments (i.e., EOLs, immediate and delayed text and graph JOLs, and RCJs), and time to make each metacognitive judgment.

**Eye-tracking.** While examining the 12 slides, participants’ fixations, saccades, and eye-gaze behavior data were recorded using the SMI RED 250 eye-tracker with a sampling rate of 60Hz. The gaze position accuracy (i.e., the gaze intersection with a target) of the SMI RED 250 is 0.4°, meaning the level of expected error of the eye-tracker is 0.4 visual degrees to the target with a normal (i.e., 5- or 9-point) calibration. Additionally, the spatial resolution (i.e., the smallest eye-movement visible to the eye-tracker) of the SMI RED 250 is 0.03°. This means that any eye-movement smaller in amplitude than 0.03° is undetectable to the
eye-tracker. The eye-tracker was integrated with iMotions Attention Tool 5.4 eye-tracking module (iMotions, 2014; SensoMotoric Instruments, 2014).

**Electrodermal activity (EDA).** Participants wore on their wrist a Shimmer3 (Shimmer™, 2014) wireless bracelet designed to record EDA. The Shimmer3 logged participants’ phasic and tonic EDA fluctuations at a sampling rate of 102.8Hz as participants progressed through the 12 trials. These data were collected but not analyzed for this study.

**Experimental Procedure**

The experimental procedure for this study included three phases: (1) the presession, where participants took the science content pretest, discrepancy-related science content pretest, and graph comprehension pretest online, before coming to the lab to complete the study; (2) setup, where participants filled out an informed consent form and provided demographic information, the eye-tracker was calibrated, and participants’ baseline EDA was collected; and (3) the experimental session where participants completed 12 trials, which entailed providing several metacognitive judgments (i.e., EOLs, immediate and delayed text and graph JOLs, RCJs), reading inference questions, reading text, inspecting graphs, and answering the inference questions for each of the 12 slides. For examples of the experimental setup and a full illustration of the experimental procedure, see Appendices K and L.

**Pre-session.** Participants were asked to complete the science content and graph comprehension pretests online, prior to arriving to the lab for the experimental session. The website address was provided to participants when they scheduled their session to participate,
and their experimental session was scheduled within two days following the completion of the online pretests. This was done to ensure that participants were not primed to the science content presented during the experimental session.

**Setup.** After arriving at the lab, participants were asked to complete the informed consent form and to wear the EDA bracelet until the completion of the experiment. Participants were then asked to position themselves at approximately 60cm from the computer monitor for accurate eye-tracking data to be captured (see Appendix K). After verifying the distance between the participant and the eye-tracker, the calibration was started by the researcher. A 9-point calibration was performed, which required participants to look at specific points in different regions of the screen (i.e., dots on nine areas of the screen). The calibration accuracy of the eye-tracker was verified by the researcher; if the calibration values were inaccurate (i.e., the offset of eye-movements was greater than 0.05 mm), the calibration procedure was repeated until the offset of the eye-movements was less than or equal to 0.05 mm. Lastly, participants were presented with a blank grey screen for 6 minutes where they were asked to sit still and face the screen to collect baseline EDA data. The entire phase lasted approximately 8 minutes.

**Experimental session.** The experimental session began once the eye-tracker was calibrated and the baseline EDA data were collected. The experiment was conducted using iMotions Attention Tool 5.4 (2014). First, the experimenter instructed participants how to interact with the environment (e.g., how to progress through the trials by pressing the arrow
button on the bottom of the screen, along with an example of one complete trial; see Appendix L). This entailed participants being informed that they would view science content and make judgments about the content. Participants were informed they would see a stop sign that would last 30 seconds, during which they should think about the content they just viewed. Then participants were informed that after the stop sign, they would make more judgments about the content, answer a question related to the content, and judge their confidence in their answer.

The 12 trials (including the four no discrepancy trials, the four text discrepancy trials, and the four text and graph discrepancy trials) had the identical structure. In each trial, participants first viewed an inference question and were subsequently asked to make an EOL by answering, “How easy do you think it will be to learn the information needed to answer this question?” (see Appendix M). Participants made their judgment on a 0% to 100% scale, increasing in increments of 20% (i.e., 0%, 20%, 40%, 60%, 80%, 100%). A judgment of 0% indicated participants found the question to be very difficult to answer, whereas a judgment of 100% indicated that participants found the question to be very easy to answer.

Participants were then presented with the slide containing three paragraphs of content, a corresponding graph, and the inference question that was presented previously. For each slide (e.g., atoms, transpiration, etc.) the three paragraphs, graph, and inference question were presented simultaneously. Once the participants read the paragraphs, inspected the graph, and finished viewing the content, they clicked the arrow button on the bottom center of the screen
to progress to the next step of the trial (see Appendix N for an example slide as shown to participants in Attention Tool 5.4).

Upon their clicking of the arrow button, participants were then asked to make an immediate JOL by answering, “How well do you understand the text presented on this topic (e.g., projectile motion)?” Participants made their judgments on a 0% to 100% scale increasing in increments of 20%. A response of 0% indicated that participants thought they did not understand the content at all, whereas a response of 100% indicated that participants completely understood the content (see Appendix O). Participants were then asked to make an immediate JOL for the graph by answering, “How well do you understand the information presented in this graph?” These judgments were made on a 0% to 100% scale increasing in increments of 20% (see Appendix P). Once participants had provided their immediate text and graph JOLs, the iMotions Attention Tool 5.4 software displayed a stop sign image for 30 seconds. Following the 30 seconds, participants were again asked to make JOLs about the text and graph. These judgments followed the same format as the immediate text and graph JOLs (see Appendices O–P).

Participants were then asked to provide a response to the inference question. Participants typed their responses into a textbox presented in the middle of the screen and were given as much time as needed to respond to the question (see Appendix Q); and after they submitted their answer they were subsequently asked to make an RCJ by answering the question, “How confident are you that the answer you provided is correct?” Participants
made their judgment on a 50% to 100% scale increasing in increments of 10%. A judgment of 50% confidence indicated that participants simply guessed at the answer (indicating that participants believed they had a 50/50 chance of getting their answer correct), whereas a judgment of 100% indicated that participants were completely confident in their response (see Appendix R). This procedure was repeated for the remainder of the 12 trials presented during the experimental session. Following the completion of the experimental session, participants were administered the adapted PIQ (Schraw et al., 1995) to assess their perceived interest in the science content presented during the trials. Once the PIQ was completed, the experimenter turned off the iMotions Attention Tool 5.4 software and asked the participant to remove the EDA bracelet. Lastly, participants were debriefed and compensated $20.00 for their participation. The entire experimental session lasted approximately 1.5 hours.

**Coding and Scoring**

**Metacognitive judgments (EOLs, immediate and delayed text and graph JOLs, RCJs).** Each participant had a total of 72 judgments collected during the experimental session (12 each of EOLs, immediate text JOLs, immediate graph JOLs, delayed text JOLs, delayed graph JOLs, and RCJs). Three mean scores (one for each type of discrepancy: none, text, and text and graph) were calculated for each participant for each type of judgment, totaling 18 mean judgment scores per participant.
**Time to make metacognitive judgments.** The amount of time in seconds that participants spent making each of the 72 metacognitive judgments was automatically recorded by iMotions Attention Tool 5.4. Three mean scores were calculated for the time to make each type of judgment (e.g., EOLs, immediate and delayed text and graph JOLs, and RCJs), totaling 18 mean judgment times per participant.

**Metacognitive judgment accuracy.** A mean judgment accuracy score was calculated for each participant for each type of JOL (i.e., immediate and delayed text and graph JOLs). Calculating judgment accuracy scores for each trial involved subtracting each participant’s question response score, as a percentage from their JOL percentage responses. These calculations are referred to in the metacognitive literature as bias scores, bias calibrations, or over- and under-confidence measures (Dunlosky & Metcalfe, 2009; Pieschl, 2009; Schraw, 2009). These judgment accuracy means allowed the researchers to examine the direction of the metacognitive judgment, indicating how under- or over-confident a participant’s metacognitive judgments were. For example, if participants provided a JOL of 80% and scored a 33% on their response to the inference question, their response score was subtracted from their judgment score (80% – 33%), resulting in a measurement of accuracy. In this example, the participants were 47% overconfident in their understanding as their judgment was 47% greater than their response score. If participants provided a JOL of 20% and scored a 67% on their response to the inference question, their accuracy score would be –47%, meaning the participants were 47% underconfident in their response. This was done for each
of the four judgments of learning (immediate text and graph, and delayed text and graph) for each participant.

**Participants’ study time.** The amount of time (in seconds) each participant spent on each slide was automatically recorded by iMotions Attention Tool 5.4. A total of three mean study times (one for each type of discrepancy: none, text, and text and graph) was calculated per participant.

**Inference questions.** The inference question responses were scored using a thematic coding method, in which the researchers identified specific concept units found within the target material (Jackson & Trochim, 2002; Ryan & Bernard, 2000). Specifically, a concept unit was a phrase or sentence representative of one concept required to answer the specific inference question related to the science content presented in the trial. This study used the coding scheme developed by Burkett and Azevedo (2012), which outlined three concept units for each of the 12 trials (see Appendix S). Participants’ responses to the inference questions were scored according to this coding scheme for presence of these components. For example, in the trial that asked participants “Why does the nucleus of an atom have a positive charge?” the participant’s answer needed to include: (1) that protons are positive and neutrons are neutral, (2) that electrons are negative, and (3) that the nucleus only contains protons and neutrons and no electrons, to be awarded a 100% score. If the participant’s response contained none of the three measurable concept units, such as in this response, “The nucleus has a positive charge because it is divided into two sections and one has a positive
charge (the proton) and the other one has a negative charge (the neutron). Together they balance one another out,” the participant received a 0% score. If the participant’s response contained one of the three concept units, such as when a participant’s response included “The nucleus of an atom has two properties: protons and neutrons,” the participant was given a 33% score. An example of a response with a 67% score was: “An atom’s nucleus consists of neutrons with no charge and protons which contain one unit of positive charge and thus cannot be negative or neutral.” An example of a 100% scored response was: “The nucleus of an atom contains protons, which are positively charged, and neutrons, which are neutral. Therefore the net charge of the nucleus will be positive. The negative particles, electrons, reside outside of the nucleus in the electron cloud.” Lastly, three mean scores for each participant (by each discrepancy type: none, text, text and graph) were calculated.

**Interrater agreement.** A primary and secondary coder scored all of the 672 inference question responses. Any disagreements between the researchers were settled through discussion; the interrater reliability was calculated to be 95% (639/672).

**Perceived Interest Questionnaire (PIQ).** The participants’ perceived situational interest in the overall science content was calculated by summing participant’s individual responses to the PIQ. For example, a higher summed score (e.g., a score of 43 out of 50) indicated that participants reported more perceived situational interest in the science content, whereas a lower summed score (e.g., a score of 21 out of 50) would indicate participants reported less perceived situational interest in the science content.
**Eye-movements.** Participants’ eye-tracking data were analyzed by creating AOIs using the eye-tracking module of iMotions Attention Tool 5.4 (iMotions, 2014). AOIs were superimposed on specific components of the slides. More specifically, AOIs were drawn on the text, the sentences in the text that contained the discrepant information, the graph, and the inference question to determine whether or not participants detected discrepant information present in the text, the graph, or both (see Appendices T–V for examples of drawn AOIs). AOI analyses were then conducted by the eye-tracking module of Attention Tool 5.4, which determined the number of fixations, the proportion of time participants spent viewing the AOI in relation to the total slide exposure time, and the number of revisits (e.g., how many times the participants fixated on an AOI, or fixated on a different area of the screen and then returned to the AOI) per AOI. Additionally, the number of fixations for the discrepant content (i.e., the sentences in the text, or the graph that contained discrepant information) was calculated by subtracting the number of fixations on discrepant content AOIs from the total text and graph AOIs. This process was repeated to calculate both the proportion of time spent viewing the discrepant content and the number of revisits to the discrepant content. Next, three mean scores were calculated for the number of fixations on the AOIs (e.g., on the text, graph, and inference question) by discrepancy type (i.e., none, text, text and graph), resulting in nine mean scores per participant. Three mean scores were calculated for the proportion of time spent viewing the text, graph, and inference question, separately. Lastly, three mean scores were also calculated for the number of revisits to the text, graph, and inference by
discrepancy type. In total, 27 mean scores were calculated for the eye-movement metrics collected during this study, per participant.
RESULTS

To investigate the effect of discrepancy type (none, text, and text and graph) on participants’ accuracy of responses to inference questions, study-time allocation, metacognitive judgment scores, and eye-movements (i.e., fixations on the AOIs, proportion of time spent viewing AOIs, and revisits to AOIs), a series of analyses was conducted. Separate one-way repeated measures ANOVAs were conducted to determine if discrepancy type influenced the accuracy of participants’ responses to inference questions and study-time allocation. Next, Pearson product–moment correlations were calculated to examine the relationships among inference question scores, study-time allocation, and the metacognitive judgments (i.e., EOLs, immediate and delayed text and graph JOLs, and RCJs; see Table 3). In addition, a series of one-way repeated measures ANOVAs was conducted to determine if discrepancy type impacted participants’ metacognitive judgments (see Table 4). Next, the relationships between participants’ eye-movements and inference question scores, study-time allocation, and metacognitive judgments were examined with a Pearson product–moment correlation (see Appendix W). Differences in the number of participants’ fixations on the text, graph, and inference question by discrepancy type were examined with three one-way repeated measures ANOVAs. Furthermore, differences in the proportion of time participants spent viewing the text, graph, and inference question were examined using a series of three one-way repeated measures ANOVAs. Three one-way repeated measures ANOVAs were also run to determine if discrepancy type influenced the number of participants’ revisits to
the text, graph, and inference question (see Table 5 for all eye-movement ANOVAs). Lastly, participants’ responses to the PIQ showed a mean of 31.00 ($SD = 6.82$).

**Research Question 1:** Does discrepancy type influence the accuracy of inference question responses during multimedia learning?

A one-way repeated measures ANOVA was conducted using participants’ inference question scores to determine if type of discrepancy (none, text, text and graph) impacted the accuracy of participants’ inference question responses. Results revealed a significant effect of discrepancy on the accuracy of responses to inference questions: $F(2, 54) = 7.45, p < .001, \eta^2_p = .22$. A Bonferroni pairwise comparison indicated scores were significantly higher for content with a text discrepancy ($M = 0.48, SD = 0.16$) than either content with no discrepancy ($M = 0.41, SD = 0.15$) or content with a text and graph discrepancy ($M = 0.41, SD = 0.15$). Results revealed no significant differences between participants’ inference question scores for content with no discrepancy and content with a text and graph discrepancy.

**Research Question 2:** Does discrepancy type influence study-time allocation during multimedia learning?

A one-way repeated measures ANOVA was conducted to determine if discrepancy type (none, text, and text and graph) influenced participants’ study-time allocation. Mauchly’s test indicated that the assumption of sphericity had been violated, $\chi^2(2) = 8.68, p < .05, \varepsilon = .87$, and therefore Greenhouse-Geisser corrected tests are reported. The results
revealed a significant effect of discrepancy on participants’ study-time allocation: $F(1.74, 95.80) = 4.08, p < .05, \eta^2_p = .07$. A Bonferroni pairwise comparison revealed that participants allocated more time to study content that with a text and graph discrepancy ($M = 106.50, SD = 39.61$) than content with a text discrepancy ($M = 96.64, SD = 38.35$). Results revealed no significant differences in participants’ study-time allocation for content with no discrepancy ($M = 100.63, SD = 37.86$) than either content with discrepancies.

**Research Question 3:** Does discrepancy type influence metacognitive judgments during multimedia learning?

In order to better understand the relationships among metacognitive judgments, inference question scores, and study-time allocation, a series of Pearson product–moment correlations was run (see Table 3). Results revealed that all metacognitive judgments (i.e., EOLs, immediate and delayed text and graph JOLs, and RCJs) were significantly correlated with each other—indicating that as participants judged the content as being easier to learn, they also reported understanding the content better and were more confident in their responses to the inference questions. Results also revealed that both immediate and delayed text JOLs and RCJs were significantly negatively correlated with study time. This indicated that as participants judged their understanding of the text-based content better, they also spent less time studying the content and reported more confidence in their responses to inference questions. Furthermore, results revealed that all metacognitive judgments were significantly correlated with inference question scores. More specifically, as participants
judged the content as being easier to learn and reported understanding the content better, they scored higher on and reported more confidence in their responses to the inference questions.

**Metacognitive judgments.** In order to determine if discrepancy type (none, text, and text and graph) influenced participants’ metacognitive judgments, a series of six one-way repeated measures ANOVAs was run on each of the six metacognitive judgment types (EOLs, immediate and delayed text and graph JOLs, and RCJs). For each of these analyses, if the Mauchly’s test statistic was significant and the assumption of sphericity had been violated, the Greenhouse-Geisser statistics are reported. Additionally, if the result of the ANOVA was significant, a Bonferroni pairwise comparison was conducted to compare the individual levels of the discrepancy type (none, text, and text and graph; see Table 4).

**Ease of learning judgments.** Results indicated a significant effect for discrepancy type on EOLs, $F(2, 54) = 9.00, p < .001, \eta_p^2 = .25$, such that EOLs were significantly higher for content with no discrepancy ($M = 0.72, SD = 0.15$) and content with a text discrepancy ($M = 0.73, SD = 0.15$) than content with a text and graph discrepancy ($M = 0.68, SD = 0.14$). Results indicated no significant differences for EOLs between content with no discrepancy and content with a text discrepancy.

**Immediate text judgments of learning.** Results indicated participants’ immediate text JOLs were not significantly different, $F(2, 54) = 0.13, p > .05, \eta_p^2 = .005$.

**Immediate graph judgments of learning.** Results indicated a significant effect of discrepancy type on participants’ immediate graph JOLs, $F(1.57, 86.20) = 28.05, p < .001,$
\( \eta_p^2 = .34 \), such that participants’ immediate graph JOLs were significantly higher for content with a text discrepancy \( (M = 0.77, SD = 0.19) \) than either content with no discrepancy \( (M = 0.71, SD = 0.18) \) or content with a text and graph discrepancy \( (M = 0.60, SD = 0.15) \). Results also revealed that participants’ immediate graph JOLs were significantly higher for content with no discrepancy than content with a text and graph discrepancy.

**Delayed text judgments of learning.** Results indicated that participants’ delayed text JOLs were not significantly different, \( F(2, 54) = 0.58, p > .05, \eta_p^2 = .02 \).

**Delayed graph judgments of learning.** Results indicated a significant effect of discrepancy type on delayed graph JOLs, \( F(1.71, 93.93) = 22.52, p < .001, \eta_p^2 = .30 \), such that delayed graph JOLs for content with a text discrepancy \( (M = 0.76, SD = 0.19) \) were significantly higher than either content with no discrepancy \( (M = 0.70, SD = 0.18) \) or content with a text and graph discrepancy \( (M = 0.60, SD = 0.16) \). Results also indicated that delayed graph JOLs were significantly higher for content with no discrepancy than content with a text and graph discrepancy.

**Retrospective confidence judgments.** Results revealed a significant effect of discrepancy on RCJs, \( F(2, 54) = 18.30, p < .001, \eta_p^2 = .40 \), such that participants’ RCJs were significantly higher for content with a text discrepancy \( (M = 0.85, SD = 0.11) \) than either content with no discrepancy \( (M = 0.80, SD = 0.10) \) or content with a text and graph discrepancy \( (M = 0.80, SD = 0.10) \). Results indicated that there were no significant
differences between RCJs for content with no discrepancy and content with a text and graph discrepancy.

Overall, results indicated that immediate and delayed text JOLs were not influenced by discrepancy type. However, results revealed participants’ EOLs, immediate and delayed text graph JOLs, and RCJs were influenced by discrepancy type. Specifically, EOLs were higher for content with no discrepancy and a text discrepancy than for content with a text and graph discrepancy. For immediate and delayed graph JOLs and RCJs, participants’ judgments were significantly higher for content with a text discrepancy than either content with no discrepancy or content with a text and graph discrepancy.

**Time to make metacognitive judgments**. In order to determine if discrepancy type influenced the amount of time it took participants to make metacognitive judgment, a series of one-way repeated measures ANOVAs was run on each of the six judgment types (EOLs, immediate and delayed text and graph JOLs, and RCJs). For each of the six repeated measures ANOVAs on each judgment type, results indicated no significant differences ($p > .05$).

**Metacognitive judgment accuracy**. To determine if there was a difference between the accuracy of immediate and delayed JOLs, two paired samples $t$-tests were run. The first paired samples $t$-test compared the accuracy of immediate and delayed text JOLs. Results indicated there was no significant difference between participant’s judgment accuracy for immediate ($M = 0.29, SD = 0.15$) and delayed ($M = 0.29, SD = 0.16$) text JOLs, $t(55) = 1.40,$
The second paired samples \( t \)-test compared the accuracy of immediate and delayed graph JOLs. Results indicated there was no significant difference between the accuracy of immediate (\( M = 0.26, SD = 0.16 \)) and delayed (\( M = 0.25, SD = 0.16 \)) graph JOLs, \( t(55) = 1.60, p > .05 \).

Overall, the results indicated that there were no significant differences in the accuracy between immediate and delayed text JOLs, and immediate and delayed graph JOLs.

**Research Question 4:** Does discrepancy type influence the number of fixations on the text, graphs, and inference questions presented during multimedia learning?

In order to better understand the relationships among participants’ eye-movements (i.e., the number of fixations on, the proportion of time spent viewing, and the number of revisits to the text, graph, and inference question), metacognitive judgments, inference question scores, and study-time allocation, a series of Pearson product–moment correlations was run. Results indicated that as participants had more fixations on the discrepant text and the discrepant graphs, they also allocated more time to studying the content (for a correlation matrix with these variables, see Appendix W).

Although the number of fixations on, proportion of time spent viewing, and number of revisits to the discrepant content (i.e., the sentences that contained discrepancies, or the discrepant graph) were calculated, these data were not used in the following analyses. Not all discrepancy types were included on each slide (e.g., two discrepant sentences versus no discrepant sentences), and as such, they could not be compared.
To determine if discrepancy type (i.e., none, text, and text and graph) impacted the number of participants’ fixations on the text, graph, and inference question, a series of 3 one-way repeated measures ANOVAs was conducted using a subsample ($n = 30$) of the total participants (see Table 5).

**Fixations on text.** Results revealed that discrepancy type did not significantly influence the number of participants’ fixations on the text, $F(2, 28) = 1.61, p > .05, \eta^2_p = .10$.

**Fixations on graph.** Results revealed there was a significant effect of discrepancy type on the number of participants’ fixations on the graph, Wilk’s $\lambda = 0.41, F(2, 28) = 20.35, p = .000$. However, Mauchly’s test was significant, and therefore the assumption of sphericity had been violated, $\chi^2(2) = 58.30, p = .000, \varepsilon = 0.53$. As such, the Greenhouse-Geisser corrected tests were examined, and they indicated that the number of participants’ fixations on the graph was not significantly impacted by the discrepancy type, $F(1.07, 30.92) = 1.38, p > .05, \eta^2_p = .05$.

**Fixations on inference question.** Results revealed that there was a significant effect of discrepancy type on the number of participants’ fixations on the inference question, Wilk’s $\lambda = 0.75, F(2, 28) = 4.64, p < .05, \eta^2_p = .25$. However, Mauchly’s test was significant and indicated the assumption of sphericity had been violated, $\chi^2(2) = 186.31, p < .001, \varepsilon = .50$. Therefore, the Greenhouse-Geisser corrected tests were examined, and they revealed that the number of participants’ fixations on the inference question was not significantly influenced by discrepancy type, $F(1.00, 29.02) = 2.71, p > .05, \eta^2_p = .09$. 

Overall, the results suggest that the number of participants’ fixations on the text, graph, and inference question were not significantly influenced by discrepancy type (none, text, and text and graph).

**Research Question 5:** Does discrepancy type influence the proportion of time spent viewing the text, graphs, and inference questions presented during multimedia learning?

To determine if discrepancy type (none, text, and text and graph) influenced the proportion of time (in seconds) participants spent viewing the text, graph, and inference question, a series of three, one-way repeated measures ANOVAs was conducted using a subsample (n = 30) of the total participants. Additionally, if the result of the ANOVA was significant, a Bonferroni pairwise comparison was conducted in order to compare the individual levels of the discrepancy type (see Table 5).

**Proportion of time viewing the text.** Results revealed that the proportion of time (in seconds) participants spent viewing the text was significantly influenced by discrepancy type, $F(2, 28) = 5.57, p < .05, \eta^2_p = .29$, such that the proportion spent viewing the text was significantly higher for both content with no discrepancy ($M = 61.26, SD = 7.75$) and a text discrepancy ($M = 61.81, SD = 10.48$) than content with a text and graph discrepancy ($M = 58.91, SD = 9.90$). Additionally, results revealed there were no significant differences between the proportion spent viewing the text for content with no discrepancy and content with a text discrepancy.
**Proportion of time viewing the graph.** Results revealed a significant effect of discrepancy type on proportion of time (in seconds) participants spent viewing the graph, $F(2, 28) = 19.00, p < .001, \eta^2_p = .58$, such that the proportion spent viewing the graph was significantly higher for content with a text and graph discrepancy ($M = 10.30, SD = 4.34$) than either content with no discrepancy ($M = 7.19, SD = 3.17$) or content with a text discrepancy ($M = 6.77, SD = 3.44$). Furthermore, results revealed no significant differences between the proportion spent viewing content with no discrepancy and content with a text discrepancy.

**Proportion of time viewing the inference question.** Results revealed a significant effect of discrepancy type on proportion of time (in seconds) spent viewing the inference question, $F(2, 28) = 5.15, p > .05, \eta^2_p = .27$, such that the proportion spent viewing the inference question was higher for both content with no discrepancy ($M = 3.10, SD = 2.01$) and content with a text discrepancy ($M = 3.18, SD = 2.20$) than content with a text and graph discrepancy ($M = 2.32, SD = 1.60$). Furthermore, results revealed no significant difference in the proportion spent viewing the inference question for content with no discrepancy and content with a text discrepancy.

**Research Question 6:** Does discrepancy type influence the number of revisits to the text, graphs, and inference questions during multimedia learning?

In order to determine if discrepancy type (none, text, and text and graph) influence the amount of revisits (e.g., the amount of times the participants’ eyes moved out of and
returned to an AOI) to the text, graph, and inference question, a series of 3 one-way repeated measures ANOVAs was conducted using a subsample \((n = 30)\) of the total participants. Additionally, if the result of the ANOVA was significant, a Bonferroni pairwise comparison was conducted to compare the individual levels of the discrepancy type (see Table 5).

**Revisits to the text.** Results revealed there was not a significant effect of discrepancy type on the amount of revisits to the text, \(F(2, 28) = 0.56, p < .05, \eta^2_p = .04\).

**Revisits to the graph.** Results revealed that Mauchly’s test indicated that the assumption of sphericity had been violated, \(\chi^2(2) = 8.52, p > .05, \varepsilon = .79\), and therefore the Greenhouse-Geisser corrected tests are reported. The results showed a significant effect of discrepancy type on the number of revisits to the graph \(F(1.58, 45.94) = 20.66, p < .001, \eta^2_p = .42\), such that the number of revisits to the graph was significantly higher for content with a text and graph discrepancy \((M = 15.26, SD = 9.33)\) than either content with no discrepancy \((M = 10.30, SD = 7.14)\) or content with a text discrepancy \((M = 9.25, SD = 7.00)\). Furthermore, results revealed no significant difference in the number of revisits to the graph for content with no discrepancy and content with a text discrepancy.

**Revisits to the inference question.** Results revealed there was not a significant effect of type of discrepancy type on the number of revisits to the inference question, \(F(2, 28) = 3.07, p > .05, \eta^2_p = .18\).

Overall, results indicated that the number of fixations on the text, graph, and inference question was not influenced by discrepancy type. In addition, neither the number of
revisits to the text nor the number of revisits to the inference question was influenced by discrepancy type. However, results revealed that the proportion of time spent viewing the text, graph, and inference question, as well as the revisits to the graph, was significantly influenced by discrepancy type. More specifically, the proportion of time spent viewing the text and inference question was higher for content with no discrepancy than content with discrepancies. The proportion of time spent viewing the graph was higher for content with a text and graph discrepancy than either content with no discrepancy or content with a text discrepancy. Lastly, revisits to the graph were significantly higher for content with a text and graph discrepancy than either content with no discrepancy or a text discrepancy.
DISCUSSION

The purpose of this study was to enhance our current understanding of the relationship between discrepancies in multimedia content and accuracy of responses to inference questions, study-time allocation, and metacognitive judgments based on behavioral, self-report, and trace data (e.g., eye-movements) during multimedia learning. Overall, results demonstrated that participants’ accuracy on inference questions, study-time allocation, and metacognitive judgments (i.e., EOLs, JOLs, delayed JOLs, and RCJs) were all significantly correlated with one another. Furthermore, results indicated that discrepancy type (i.e., none, text, text and graph) influenced accuracy of responses to inference questions, study-time allocation, and most metacognitive judgments except for immediate and delayed text JOLs. Eye-tracking data revealed that the number of fixations on discrepant text and graphs were significantly correlated with study-time allocation. In addition, discrepancy type influenced the proportion of time participants spent viewing text, graphs, and inference questions as well as the number of revisits to the graph. However, discrepancy type did not influence the number of fixations on text, graphs, and inference questions, or the number of revisits to text and inference questions.
Research Question 1

Does discrepancy type influence the accuracy of inference question responses during multimedia learning?

This research question addressed the influence of discrepancy type (i.e., none, text, text and graph) on the accuracy of responses to inference questions during multimedia learning. Results indicated that response accuracy was higher for content with text discrepancies than content with either no discrepancies or text and graph discrepancies. Additionally, results showed no significant differences between accuracy for content with no discrepancies and accuracy for content with text and graph discrepancies. Both findings support prior literature, suggesting that the presence of discrepancies influences accuracy of responses to inference questions during multimedia learning (Burkett & Azevedo, 2012). These findings can be partially explained by the low response accuracies across all discrepancy types (no discrepancies: $M = 0.41, SD = 0.15$, text discrepancies: $M = 0.48, SD = 0.16$, text and graph discrepancies: $M = 0.41, SD = 0.15$) and that the science content with text discrepancies could have possibly been less difficult than the content in the slides with no discrepancies or text and graph discrepancies.

Research Question 2

Does discrepancy type influence study-time allocation during multimedia learning?

This research question addressed the influence of discrepancy type on study-time allocation (in seconds) during multimedia learning. Results indicated that study-time
allocation was significantly higher for content with text and graph discrepancies than content with text discrepancies. No significant differences were found between study-time allocation for content with no discrepancies and content with discrepancies. Although inconsistent with some literature that suggests the presence of discrepancies does not influence study-time allocation (Burkett & Azevedo, 2012), these results suggest participants detected the discrepancies and allocated more time to study content that was poorly understood. This interpretation is in line with Mayer’s CTML (2014b), which suggests that text and graph discrepancies may have interfered with participants’ organization and integration of the text and graphs, which may contribute to longer study-time allocation.

**Research Question 3**

**Does discrepancy type influence metacognitive judgments during multimedia learning?**

This research question addressed the influence of discrepancy type on metacognitive judgments (i.e., EOLs, immediate and delayed JOLs, and RCJs) during multimedia learning. A discussion regarding the influence of discrepancy type on each metacognitive judgment is provided below.

**Ease of learning (EOL) judgments.** Results showed that discrepancy type influenced EOLs. More specifically, EOLs were significantly higher for content with no discrepancies and with text discrepancies than content with text and graph discrepancies. These results are inconsistent with Nelson and Narens’ (1990) metamemory framework since EOLs are made prior to learning and therefore they should not be influenced by the presence
of discrepancies. Furthermore, these findings are not supported by previous literature, demonstrating that discrepancy type does not influence EOLs during multimedia learning (e.g., Burkett & Azevedo, 2012). A possible explanation for the results found with the no discrepancy and text discrepancy slides could be explained by participants perceiving the content in these slides, based on the inference question they viewed, as more difficult to learn than content in slides with text and graph discrepancies.

**Immediate and delayed text judgments of learning (JOLs).** Results indicated that discrepancy type did not influence immediate text JOLs, which is not in line with prior research (e.g., Burkett & Azevedo, 2012) and several theories (i.e., Koriat’s cue-utilization framework, 1997; Mayer’s CTML, 2014b). More specifically, Mayer’s CTML (2014b) suggests that text containing discrepancies might interfere with the ability to form a coherent cognitive representation and should be judged as less understood. Additionally, Koriat’s (1997) cue-utilization framework suggests that the discrepancies in the text should be used as intrinsic, mnemonic, and extrinsic cues (see Introduction) during multimedia learning. A plausible explanation for these findings is that participants did not pay attention to or detect the presence of text discrepancies, consistent with research on comprehension regulation (Otero et al., 2002).

Results also indicated that discrepancy type did not influence delayed text JOLs. As prior literature and theory have reported on the increased accuracy of delayed JOLs, these results are unexpected because they conflict with the delayed JOL effect (Nelson &
Dunlosky, 1991) and are inconsistent with results from Burkett and Azevedo (2012). More specifically, the monitoring dual memories (MDM) principle suggests the delay should facilitate more accurate judgments of the discrepancies since content was accessed from participants’ long-term memory as opposed to their short-term memory. As immediate JOLs were not influenced by discrepancy type, it is possible that participants did not attend to or notice the discrepancy. Additionally, given the accuracy of participants’ responses to inference questions across discrepancy type, it is also possible that the 30-second delay between making the immediate and delayed JOLs was not enough time to make more accurate judgments.

**Immediate and delayed graph judgments of learning (JOLs).** Results showed that immediate graph JOLs were significantly lower for content with text and graph discrepancies than content with no discrepancies and content with text discrepancies. In addition, results indicated immediate graph JOLs were lower for content with no discrepancies than content with text and graph discrepancies. Though inconsistent with the nonsignificant results from Burkett and Azevedo (2012), results are in line with those from Jaeger and Wiley (2014) who demonstrated a significant influence of illustrations on participants’ JOLs. Additionally, Mayer’s CTML (2014b) states that information presented in the pictorial channel (i.e., graph) is processed differently than visual (i.e., text) information, which can help explain why discrepancy type influenced immediate graph JOLs as opposed to immediate text JOLs.
Results also indicated delayed graph JOLs were significantly lower for content with text and graph discrepancies than content with no discrepancies and content with text discrepancies. Additionally, results demonstrated delayed graph JOLs were significantly lower for content with no discrepancies than content with text discrepancies. These results are inconsistent with the nonsignificant results from Burkett and Azevedo (2012). However, Mayer’s CTML (2014b) suggests the pictorial modality of the graph discrepancies was processed differently than the text discrepancies. As such, the discrepant graphs could have been processed faster than discrepancies in the text alone, which can help explain why discrepancy type influenced delayed graph JOLs and not delayed text JOLs. These results augment this study’s findings of longer study-time allocation for content with text and graph discrepancies, suggesting that participants detected and were aware of the discrepancies in the text and graph slides.

**Retrospective confidence judgments (RCJs).** Results indicated that discrepancy type influenced RCJs in that they were higher for content with text discrepancies than content with either no discrepancies or text and graph discrepancies. Results also demonstrated no significant differences between RCJs for content with text and graph discrepancies and content with no discrepancies. These results are expected based on previous literature, suggesting participants’ RCJs are largely inaccurate and stable across a learning session (Dunlosky & Metcalfe, 2009; Koriat & Goldsmith, 1996; Mengelkamp & Bannert, 2010). More specifically, these results suggest that participants were metacognitively unaware
whether their answers were correct or incorrect when responding to inference questions for science content across all discrepancy types.

**Accuracy of judgments of learning.** Results showed no significant differences between immediate and delayed text and graph JOLs. These findings are inconsistent with the delayed JOL effect (i.e., Nelson & Dunlosky, 1991) and previous literature (e.g., Burkett & Azevedo, 2012). More specifically, these results conflict with the delayed JOL effect, indicating the increased accuracy of JOLs made after a delay (Nelson & Dunlosky, 1991). The MDM principle emphasizes the roles of long-term and short-term memory in making JOLs. Given that participants were required to generate inferences in order to successfully answer response questions, it is possible that more emphasis was placed on combining information in the text and graphs rather than memorizing it. Additionally, in this study participants were asked how well they understood the information and not asked how likely they would be able to recall it at a later time, as is more frequently used in JOL literature (Koriat, Nussinson, & Ackerman, 2014; Metcalfe & Finn, 2008; Nelson & Dunlosky, 1991).

**Time to make metacognitive judgments.** Results indicated that discrepancy type did not influence the amount of time to make metacognitive judgments (i.e., EOLs, immediate and delayed text and graph JOLs, and RCJs). Previous literature has neglected to examine the amount of time it takes participants to make metacognitive judgments, although it can be considered analogous to study-time allocation following EOLs. It was expected that discrepancy type would influence the amount of time to make metacognitive judgments in
that the more complex the discrepancy (e.g., text and graph > text only), the longer it would take to make judgments. Prior knowledge in the science topics might have also influenced the amount of time, such that higher prior knowledge would facilitate faster metacognitive judgments. However given the low scores on the science pretests and the low accuracy of responses to inference questions, it is understandable that discrepancy type did not influence the amount of time to make metacognitive judgments.

**Research Question 4**

**Does discrepancy type influence the number of fixations on the text, graphs, and inference questions presented during multimedia learning?**

This research question addressed the influence of discrepancy type on the number of fixations on text, graphs, and inference questions during multimedia learning. Results indicated that the number of fixations on discrepant text and graphs was significantly related to study-time allocation. However, results also indicated that discrepancy type did not influence the number of fixations on text, graphs, and inference questions. The results for the number of fixations on the text are inconsistent with eye-movement and multimedia research, which suggests a bias toward fixating more on text than graphs or diagrams (Hegarty & Just, 1993; Johnson & Mayer, 2012; Schmidt-Weigand et al., 2010). Given this text bias, the number of fixations on graphs is consistent with previous literature. It is also possible that graphs were not seductive enough to attract attention (cf. Jaeger & Wiley, 2014). Lastly, little research provides a basis to interpret the nonsignificant influence of discrepancy type on the
number of fixations on inference questions. However, it was expected that more accurate metacognitive monitoring would result in more fixations on the inference questions and discrepancies to enhance multimedia learning.

**Research Question 5**

**Does discrepancy type influence the proportion of time viewing the text, graphs, and inference questions presented during multimedia learning?**

This research question addressed the influence of discrepancy type on the proportion of time spent viewing the AOIs (i.e., text, graphs, and inference questions) presented during multimedia learning. A discussion is provided below regarding the influence of discrepancy type on proportions of time spent viewing each AOI.

**Text.** Results showed that discrepancy type influenced the proportion of time spent viewing text as the text proportion was greater for content with no discrepancies and with text discrepancies than content with text and graph discrepancies. Additionally, results indicated no significant differences between the text proportions for content with no discrepancies and with text discrepancies. As such, these results are consistent with eye-movement and multimedia learning research that indicates participants spend more time viewing text than diagrams or graphs (Hegarty & Just, 1993; Johnson & Mayer, 2012; Schmidt-Weigand et al., 2010).

**Graphs.** Results showed that discrepancy type influenced the proportion of time spent viewing the graphs, as the graph proportion was greater for content with text and graph
discrepancies than content with no discrepancies and content with text discrepancies. Additionally, results indicated no significant differences in the graph proportion with no discrepancies and content with text discrepancies. These results support the research question and are augmented by the increased study time and the immediate and delayed JOLs for content with text and graph discrepancies found in this study. Furthermore, these results corroborate Mayer’s CTML (2014b) suggestion that graphs are processed differently than text, which can help explain why graph discrepancies were detected.

Inference questions. Results indicated that discrepancy type influenced the proportion of time viewing inference questions, such that the inference question proportion was longer for content with no discrepancies than content with discrepancies. Results also showed that the inference question proportion for content with text discrepancies was not significantly different than content with text and graph discrepancies. These results are inconsistent with metacognitive monitoring expectations, which suggest the proportion of time viewing inference questions should be longer for content with discrepancies. However, this result can be partially explained by the possible influence of the content in the no discrepancy slides. More specifically, the no discrepancy slide asking “What would happen if you pulled up on both ends of a stretched, taut string at the same time?” resulted in less accurate responses ($M = 0.08$), suggesting participants were unable to understand both the content and inference question of that slide.
Research Question 6

Does discrepancy type influence the number of revisits to the text, graphs, and inference questions during multimedia learning?

This research question addressed the influence of discrepancy type on the number of revisits to the AOIs (i.e., text, graphs, and inference questions) presented during multimedia learning. More specifically, this question addressed how many times participants fixated on an AOI, fixated on a different area of the screen, and returned to the original AOI. A discussion is provided below regarding the influence of discrepancy type on the number of revisits to each AOI.

**Text.** Results indicated that discrepancy type did not influence the number of revisits to the text. These results are inconsistent with prior literature, suggesting revisits are second-pass indicators and therefore indicative of intentional cognitive processing (i.e., comprehension; Mason, Pluchino, & Tornatora, 2013, 2015; Mason, Pluchino, Tornatora, & Ariasi, 2013; Rayner, 2009; Scheiter & Eitel, 2015). As such, these results suggest that participants did not detect or cognitively process text discrepancies.

**Graphs.** Results showed that discrepancy type influenced the number of revisits to the graph, such that the number of revisits was greater for content with text and graph discrepancies than content with no discrepancies and content with text discrepancies. Results also demonstrated no significant differences between content with no discrepancies and content with text discrepancies. These results are consistent with the research question and
previous literature, demonstrating revisits are indices of intentional cognitive processing and suggesting that participants detected and attempted to comprehend discrepancies in the graphs (Mason, Pluchino, & Tornatora, 2013, 2015; Mason, Pluchino, Tornatora, & Ariasi, 2013; Rayner, 2009; Scheiter & Eitel, 2015).

**Inference questions.** Results indicated that discrepancy type did not influence the number of revisits to inference questions. These results are consistent with prior literature that suggested that more revisits would occur on discrepant content that requires more cognitive processing (Mason, Pluchino, & Tornatora, 2013, 2015; Mason, Pluchino, Tornatora, & Ariasi, 2013; Rayner, 2009; Scheiter & Eitel, 2015).

In summary, the results from this study suggest a relationship among metacognitive judgments (e.g., immediate and delayed graph JOLs), study-time allocation, and eye-movements during learning with multimedia content with discrepancies. More specifically, results indicated that content with text and graph discrepancies led to longer study-time allocation, lower immediate and delayed graph JOLs, longer proportions of time spent viewing the graph, and more revisits to the graph.

**Limitations of Current Study**

The current study had several limitations. First, as with any study that uses a within-subjects methodology, it is always possible the participants noticed that some of the information was discrepant and made their judgments based on that expectation. Though there are benefits to running this experiment as a within-subjects design (e.g., reduction in
error variance associated with individual differences), it would be beneficial to see if the results remain consistent when using a between-subjects design. Furthermore, the graphs were all created by the researcher using Microsoft Office Excel 2013 to ensure uniformity of graph features across trials. However, this may have biased participants to think that the information represented in the graphs was not as important as the information in the text because the graphs may have seemed like a superficial addition. Also, participants’ low prior knowledge (based on pretest scores) in several science domains may explain low accuracy of responses to inference questions as well as their general inability to detect and resolve discrepancies due to cognitive load imposed by selecting, organizing, and integrating the multiple representations in the multimedia content.

Another possible limitation of this study is reflected by the options available to participants when making their metacognitive judgments, as they were only provided a scale increasing in increments of 20%. As such, it is possible that discrepancy type influenced metacognitive judgments but the scales used in this study were too global and not sensitive enough to detect such differences. The analyses and integration of participants’ facial expressions data could have enhanced our understanding of the role of affect during multimedia learning. For example, evidence of confusion (expressed facially) while reading or inspecting discrepancies (in both text and graphs) or revisits between discrepancies may be indicative of discrepancy detection. Lastly, despite the use of eye-tracking data, we can
only make very limited interpretations of general control strategy use because this study only
looked at study-time allocation as a measure of control processes.

**Contributions to Knowledge and Future Directions**

The results of this study enhance our current understanding of the relationship between discrepancies in multimedia content and the accuracy of responses on inference science questions, study-time allocation, and metacognitive judgments based on behavioral, self-report, and trace data (e.g., eye-movements) during multimedia learning. More specifically, these findings extend those of previous studies in metacomprehension and multimedia learning (e.g., Burkett & Azevedo, 2012; Jaeger & Wiley, 2014; Serra & Dunlosky, 2007, 2010) and eye-movement and multimedia learning (e.g., Hegarty & Just, 1993; Mason et al., 2015; Schmidt-Weigand et al., 2010) by examining the relationship between these processes as opposed to examining them in isolation. The discussion that follows relates the conceptual, theoretical, methodological, and analytical issues raised by the results of this study.

Conceptually and methodologically, results from this study suggest a difference between immediate and delayed text and graph JOLs when learning with multimedia content containing discrepancies. As such, future research in this area should focus on cognitive and metacognitive mechanisms employed during these JOLs and how they impact these judgments. More specifically, how does the 30-second delay before making a delayed JOL influence organizing information in the text and the graphs (as results from this study
indicated a difference between the two)? Also, as it has been neglected in prior literature, future research should explore what could influence (e.g., prior knowledge or more accurate metacognitive monitoring) the amount of time it takes participants to make metacognitive judgments and their accuracies.

Theoretically, future research should also explore the use of different theories and models (e.g., Haider and Frensch’s information reduction hypothesis, 1999; Schnozt’s integrated model of text and picture comprehension, 2014) to account for and analytically integrate process data channels, such as eye-tracking during multimedia learning. More specifically, researchers should theoretically integrate and temporally delineate the cognitive processes inferred from eye-movements as participants integrate and comprehend multimedia content. Furthermore, future research should explore how process data channels (e.g., eye-tracking) can augment models of metacognitive monitoring (e.g., Nelson & Narens, 1990; Winne & Hadwin, 1998), as results from this study (i.e., lower immediate and delayed graph JOLs and higher proportions of time spent viewing graphs) suggest a significant association between the two.

Analytically, eye-tracking data suggested that the number of fixations was the least informative of the influence of discrepancy type on detection and comprehension. As such, these results are in line with prior research (e.g., Scheiter & Eitel, 2015) that suggests first-pass indicators are indicative of differences in visual salience or basic reading processes rather than intentional cognitive processes related to comprehension or integration of
information. As such, future research should continue distinguishing (e.g., initial fixations on vs. revisits to AOIs) between these indicators to provide a more comprehensive examination of the cognitive processes underlying multimedia learning.

In addition, future studies should also investigate the use of different statistical methods (e.g., multilevel models such as random coefficient regression models to determine the amount of variance in eye-movements accounted for by discrepancy) or process mining techniques to delineate eye-movement patterns reflective of greater comprehension in order to appropriately capture differences in the influence of discrepancy type on eye-movements during multimedia learning. As this study was limited in its reliance on traditional statistical analyses, using multilevel modeling (Raudenbush & Bryk, 2002), could account for and compare the differences in the number (e.g., two discrepant sentences in the text vs. no text discrepancies) of discrepancies within the stimuli used in this study.

Future research should explore why and how participants detected graph discrepancies as opposed to those in the text and the impact of these discrepancies on the accuracy of responses to inference questions. More specifically, did participants’ level of prior knowledge influence their ability to detect and comprehend text discrepancies? Or was it participants’ inability to appropriately deploy cognitive strategies to resolve the discrepancies? Could cueing participants’ attention to discrepancies and providing them with the correct information to resolve them answer these questions and lead to more accurate responses and better metacognitive monitoring? Future research should further investigate
the influence of inaccurate cognitive monitoring, metacognitive monitoring, and metacomprehension processes and if potential cognitive (e.g., Bergey, Cromley, & Newcombe, 2015) or metacognitive (e.g., Azevedo, 2014; Veenman, 2014) training regimens could be of benefit to participants during multimedia learning.

Lastly, future research should consider the potential impact of affective responses (e.g., confusion or frustration) by incorporating facial expressions and EDA as they could be associated with and influence the accuracy of participants’ metacomprehension. As affect data (i.e., facial expressions of emotions and EDA) were collected but not analyzed for this study, future research will investigate the influence of participants’ affective processes on their metacomprehension for multimedia material with discrepancies. More specifically, future research will explore the influence of facial expressions of learner-centered emotions (e.g., confusion and frustration) on participants’ metacognitive judgments (i.e., EOLs, immediate and delayed JOLs, and RCJs) and eye-movements (i.e., revisits to and the proportion of time viewing AOIs). For example, expressions of confusion or frustration may have interfered with the accuracy of participants’ metacognitive judgments and their ability to resolve the detected discrepancies. Including this channel will allow us to answer questions such as the following. (1) When participants viewed the inference questions and were asked to make EOL judgments, did they evidence a facial expression of confusion, and could this help explain the EOL judgment findings in this study? (2) When participants detected or failed to detect discrepancies, what emotions did they facially express and how
did these emotions influence the accuracy of their responses to inference questions (e.g., confusion interfering with the ability to generate inferences)? (3) Did participants’ potential confusion interfere with their retrieval processes when making a delayed text JOL? (4) Are there emotions consistently expressed when making metacognitive judgments and are they indicative or predictive of judgment accuracy? Examining the influence of affective processes and other process measures (i.e., eye-tracking) will help augment our current understanding of the cognitive, affective, and metacognitive processes underlying metacomprehension.
Table 1

*Demographic Characteristics of Participants in Sample (N = 56)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>M</th>
<th>SD</th>
<th>n</th>
<th>%</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>16</td>
<td>28.6</td>
<td>28.6</td>
<td>71.4</td>
<td></td>
</tr>
<tr>
<td>Female</td>
<td>40</td>
<td>71.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Age</td>
<td>20.39</td>
<td>1.88</td>
<td>18–27</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Classification</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Freshman</td>
<td>12</td>
<td>21.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sophomore</td>
<td>11</td>
<td>19.6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Junior</td>
<td>21</td>
<td>37.5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Senior</td>
<td>9</td>
<td>16.1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other</td>
<td>3</td>
<td>5.4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Major</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Psychology</td>
<td>22</td>
<td>39.3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Prev. Science Courses</td>
<td>3.30</td>
<td>2.74</td>
<td>52</td>
<td>92.85</td>
<td>0–16</td>
</tr>
<tr>
<td>Prev. Math Courses</td>
<td>2.03</td>
<td>1.06</td>
<td>54</td>
<td>96.42</td>
<td>0–5</td>
</tr>
<tr>
<td>Prev. Work Exp.</td>
<td>8</td>
<td>14.28</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Note.* Prev. Science = Previous science courses taken at the university level; Prev. Math = Previous math courses taken at the university level; Prev. Work. Exp. = Previous work experience in the fields related to chemistry, physics, biology, physical science or math.

Table 2

*Participants' Percentage Scores on the Pretests*

<table>
<thead>
<tr>
<th>Pretest</th>
<th>M</th>
<th>SD</th>
<th>Range</th>
</tr>
</thead>
<tbody>
<tr>
<td>Science</td>
<td>69.64</td>
<td>14.23</td>
<td>35–100</td>
</tr>
<tr>
<td>Disc.</td>
<td>55.65</td>
<td>16.21</td>
<td>17–92</td>
</tr>
<tr>
<td>Graph</td>
<td>78.72</td>
<td>10.65</td>
<td>25–92</td>
</tr>
</tbody>
</table>

*Note.* Science Pretest = Pretest related to science content (e.g., chemistry, physics, biology, or physical science); Disc. = Discrepancy related science content pretest; Graph = Graph comprehension pretest.
Table 3

*Pearson Product–Moment Correlations among Metacognitive Judgments, Response Scores, and Study-Time Allocation*

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Ease of learning judgments (EOLs)</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2. Text judgments of learning (JOLs)</td>
<td>0.45*</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3. Graph judgments of learning (JOLs)</td>
<td>0.33*</td>
<td>0.45*</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Delayed text judgments of learning (JOLs)</td>
<td>0.43*</td>
<td>0.88*</td>
<td>0.47*</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Delayed graph judgments of learning (JOLs)</td>
<td>0.30*</td>
<td>0.45*</td>
<td>0.92*</td>
<td>0.50*</td>
<td>—</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6. Retrospective confidence judgments (RCJs)</td>
<td>0.42*</td>
<td>0.63*</td>
<td>0.42*</td>
<td>0.64*</td>
<td>0.43*</td>
<td>—</td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Study time</td>
<td>−0.05</td>
<td>−0.15*</td>
<td>−0.02</td>
<td>−0.14*</td>
<td>−0.04</td>
<td>−0.18*</td>
<td>—</td>
<td></td>
</tr>
<tr>
<td>8. Inference question response scores</td>
<td>0.11*</td>
<td>0.24*</td>
<td>0.14*</td>
<td>0.24*</td>
<td>0.17*</td>
<td>0.36*</td>
<td>0.03</td>
<td>—</td>
</tr>
</tbody>
</table>

* *p < .01.*
Table 4

Means, Standard Deviations and ANOVA Results for Six Metacognitive Judgments, by Condition for Univariate Analyses

<table>
<thead>
<tr>
<th>Judgment</th>
<th>No Discrepancy</th>
<th>Text Discrepancy</th>
<th>Text and Graph Discrepancy</th>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>df</td>
</tr>
<tr>
<td>Ease of learning judgment (EOL)</td>
<td>0.72 (0.15)</td>
<td>0.73 (0.15)</td>
<td>0.68 (0.14)</td>
<td>2, 54</td>
</tr>
<tr>
<td>Immediate text judgment of learning (JOL)</td>
<td>0.73 (0.13)</td>
<td>0.72 (0.16)</td>
<td>0.73 (0.14)</td>
<td>2, 54</td>
</tr>
<tr>
<td>Immediate graph judgment of learning (JOL)</td>
<td>0.71 (0.18)</td>
<td>0.77 (0.19)</td>
<td>0.60 (0.15)</td>
<td>1.57, 86.20</td>
</tr>
<tr>
<td>Delayed text judgment of learning (JOL)</td>
<td>0.71 (0.13)</td>
<td>0.72 (0.16)</td>
<td>0.73 (0.14)</td>
<td>2, 54</td>
</tr>
<tr>
<td>Delayed graph judgment of learning (JOL)</td>
<td>0.70 (0.18)</td>
<td>0.76 (0.19)</td>
<td>0.60 (0.16)</td>
<td>1.71, 93.93</td>
</tr>
<tr>
<td>Retrospective confidence judgment (RCJ)</td>
<td>0.80 (0.10)</td>
<td>0.85 (0.11)</td>
<td>0.80 (0.10)</td>
<td>2, 54</td>
</tr>
</tbody>
</table>

Note. ND = No Discrepancy; TD = Text Discrepancy; TGD = Text and Graph Discrepancy.
* p < .001.
Table 5

*Means, Standard Deviations and ANOVA Results for the Number of Fixations on, the Proportion of Time Spent Viewing, and the Number of Revisits to the Text, Graph, and Inference Question, by Condition for Univariate Analyses (n = 30)*

<table>
<thead>
<tr>
<th>Variable</th>
<th>No Discrepancy</th>
<th>Text Discrepancy</th>
<th>Text and Graph Discrepancy</th>
<th>ANOVA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>M (SD)</td>
<td>df</td>
</tr>
<tr>
<td>Fixations</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>1525.33 (4726.60)</td>
<td>279.90 (101.60)</td>
<td>297.26 (107.70)</td>
<td>2, 28</td>
</tr>
<tr>
<td>G</td>
<td>44.21 (93.32)</td>
<td>29.77 (14.31)</td>
<td>52.88 (27.32)</td>
<td>1.07, 30.92</td>
</tr>
<tr>
<td>Q</td>
<td>121.16 (363.55)</td>
<td>16.36 (11.03)</td>
<td>13.85 (11.63)</td>
<td>1.00, 29.02</td>
</tr>
<tr>
<td>Proportion of Time</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>61.26 (7.75)</td>
<td>61.81 (10.48)</td>
<td>58.91 (9.90)</td>
<td>2, 28</td>
</tr>
<tr>
<td>G</td>
<td>7.19 (3.17)</td>
<td>6.77 (3.44)</td>
<td>10.30 (4.34)</td>
<td>2, 28</td>
</tr>
<tr>
<td>Q</td>
<td>3.10 (2.01)</td>
<td>3.18 (2.20)</td>
<td>2.32 (1.60)</td>
<td>2, 28</td>
</tr>
<tr>
<td>Revisits</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>56.33 (23.82)</td>
<td>55.13 (27.36)</td>
<td>58.05 (20.82)</td>
<td>2, 28</td>
</tr>
<tr>
<td>G</td>
<td>10.30 (7.14)</td>
<td>9.25 (7.00)</td>
<td>15.26 (9.33)</td>
<td>1.58, 45.94</td>
</tr>
<tr>
<td>Q</td>
<td>6.84 (4.67)</td>
<td>6.08 (3.95)</td>
<td>5.40 (4.00)</td>
<td>2, 28</td>
</tr>
</tbody>
</table>

*Note. T = Text; G = Graph; Q = Inference Question. ND = No Discrepancy; TD = Text Discrepancy; TGD = Text and Graph Discrepancy.*

*p < .05.

**p < .001.
REFERENCES


Hertzog, C., Hines, J. C., & Touron, D. R. (2013). Judgments of learning are influenced by multiple cues in addition to memory for past test accuracy. *Archives of Scientific Psychology, 1*, 23–32.


APPENDICES
Appendix A. Recruitment Flyer

RESEARCH OPPORTUNITY

Compensation: $10/hr

CRITERIA:

- 18+ years of age
- Undergraduate Student
- 20/20 Vision (cannot be wearing glasses; contacts are fine)
- No head coverings blocking any part of the face
- Available for 1-1.5 hours

LOCATION:
740 Poe Hall
2310 Stinson Drive

S.M.A.R.T. Lab

metacompstudy@gmail.com
Nicholas Mudrick
740 Poe Hall
2310 Stinson Drive
Appendix B. Consent Form

The Relationship among Affective Processes, Metacomprehension, Study-Time Allocation, and Inferences during Multimedia Learning

Statement of Age of Participant
By signing this form you are stating that you are over 18 years of age, and wish to participate in a program of research being conducted by Dr. Roger Azevedo at North Carolina State University, Department of Psychology.

Purpose
The purpose of this research is to examine the interactions among affective processes, metacognitive judgments, inference generation and study-time allocation while learning about complex science topics in multiple representations.

Procedures
The procedures will involve 1 session, during which you will be asked to learn about a variety of science topics in multimedia. During the session, you will be asked to (1) provide metacognitive judgments at various time intervals [i.e., assessing your understanding of what you are learning], and (2) answer questions pertaining to the material that you are learning. The session will last approximately 1 hour. The experiment will be video recorded. Your eye-gaze patterns and emotional states will be captured using an eye-tracking device and cameras, which are integrated and connected to the computer monitor. You will be wearing a wireless biosensor on your wrist, which will record your skin conductance information. If you leave the study at any time, you may ask for all your data to be withdrawn.

Confidentiality
All information collected in the study is confidential within the limits allowed by law, and your name will not be identified at any time. Your face will be recorded on the videotape. A numeric code will be used as identification on data collection materials. Once data are collected, this code will be used for maintenance and analysis of data. Pseudonyms will be used in publications and conference papers. The data will be kept in a locked file drawer and a password-protected computer in Dr. Azevedo’s research lab located in room 740 of Poe Hall (2310 Stinson Drive). According to APA guidelines, after 5
years, all data will be destroyed. If you are willing to give us permission to have your video, eye-tracking data, and face recordings shown at academic conferences and presentations, or for screenshots from this data to be used in academic publications, please provide your signature here: (________________). You do not have to give permission for this usage in order to participate in the experiment.

**Risks**

There may be minimal risks associated with this experiment, including fatigue from reading science multimedia materials and stress from repeatedly being asked to complete confidence judgments. The experiment is not designed to help you personally, but that we hope to learn more about the interactions among affective processes, metacognitive judgments, inference generation, and study-time allocation while learning about complex science topics in multiple representation. You will receive $20.00 for your participation in this study. Participation is voluntary and you are free to ask questions and/or to withdraw from participation at any time without penalty or loss of benefits to which you are otherwise entitled. You also have the right to refuse to answer any question. If you have any questions or concerns regarding your rights or welfare as a participant in this research study, please contact the NC State IRB Administrator at 919-515-4514 (Ms. Debra Paxton), or at debra_paxton@ncsu.edu.

Nicholas Mudrick
Dr. Roger Azevedo
North Carolina State University
Department of Psychology
2310 Stinson Drive, room 740
Raleigh, NC, 27695-7650
Tel: 847-532-5475
Email: nvmudric@ncsu.edu

Printed Name of Participant
__________________________________________________

Signature of Participant
__________________________________________________

Research Signature
__________________________________________________

Date
__________________________________________________
Appendix C. Demographic Questionnaire

Name: ______________________________

Gender ________________________

Age: _________________________

College: _______________________

Major: _______________________ 

Classification:
• Freshman
• Sophomore
• Junior
• Senior
• Other

Previous Science courses taken (at the University Level): Please include the course title, number and if you remember any of the specific topics covered.


Previous Math courses taken (at the University Level): Please include the course title, number and if you remember any of the specific topics covered.


Relevant work experience related to the fields of Chemistry, Physics, Biology (including health or medicine), Physical Science, or Math.
Appendix D. Science Knowledge Pretest

Adapted from Burkett and Azevedo, 2012.

Please circle the response that BEST answers each of the following questions. Choose ONE and only one answer for each question. Please be sure to answer EVERY question to the best of your ability.

1. The BEST reason for including protein in a healthy diet is because it is the main source of
   a. Energy for the body.
   b. Fiber for digestion.
   c. Raw materials for cell growth and repair.
   d. Vitamins for fighting disease.

2. The Sun is an example of which of the following?
   a. Comet.
   b. Planet.
   c. Galaxy.
   d. Star.

3. An atom of Silicon has a mass number of 29 and an atomic number of 14. How many neutrons does it have?
   a. 14
   b. 15
   c. 43
   d. 29

4. According to the Ideal Gas Law, which of the following may decrease when the volume of a gas is increased?
   a. The temperature of the gas.
   b. The number of molecules in the gas.
   c. The pressure of the gas.
   d. The average kinetic energy of the gas molecules.
5. Most underground caves are formed by the action of water on
   a. Granite.
   b. Limestone.
   c. Sandstone.
   d. Shale.

6. Which is a chemical change?
   a. Element 1 is polished to form a smooth surface.
   b. Element 2 is heated and evaporates.
   c. Element 3 develops a white, powdery surface after standing in air.
   d. Element 4 is separated from a mixture by filtration.

7. The walls of a building are to be painted to reflect as much light as possible. What color should they be painted?
   a. White.
   b. Red.
   c. Black.
   d. Pink.

8. How many gametes can be produced by an individual with the genotype XXYyZz?
   a. 4
   b. 6
   c. 8
   d. 12

9. Almost all interactions of matter result from the operation of which of the following forces?
   a. Gravitational force.
   b. Electromagnetic force.
   c. Nuclear force.
   d. All of the above.

10. Oxygen, hydrogen, and water are substances. Which of these substances are elements?
    a. Oxygen, hydrogen and water.
    b. Oxygen and hydrogen only.
    c. Oxygen only.
    d. Water only.
11. Which of the following phrases best describes convection?
   a. An exchange of particles that increases the disorder in a system.
   b. An exchange of heat between a hot solid object and a cold solid object.
   c. An exchange of heat between warmer and cooler regions in a gas or liquid.
   d. An exchange of gas particles between higher pressure and lower pressure regions.

12. Tissues are found in living things. What is the definition of a tissue?
   a. A group of cells with similar structure and function.
   b. A group of cells with different structure and function.
   c. A group of organelles contained inside a cell.
   d. A group of substances that make up the walls of a cell.

13. Three gases found in Earth’s atmosphere are carbon dioxide, nitrogen, and oxygen. What is their order of abundance from greatest to least?
   a. Nitrogen, oxygen, carbon dioxide.
   b. Nitrogen, carbon dioxide, oxygen.
   c. Oxygen, nitrogen, carbon dioxide.
   d. Carbon dioxide, oxygen, nitrogen.

14. People get energy from the food they eat. Where does the energy stored in food come from?
   a. Fertilizer.
   b. The Sun.
   c. Vitamins.
   d. The soil.

15. When a person sees something, what carries the message from the eyes to the brain?
   a. Arteries
   b. Muscles
   c. Nerves
   d. Veins

16. A powder made up of both white specks and black specks is likely to be
   a. A solution.
   b. A pure compound.
   c. A mixture.
   d. An element.
17. Mitosis and meiosis are similar because they both:
   a. Result in the production of gametes in humans.
   b. Result in the production of two identical daughter cells.
   c. Involve replication of DNA.
   d. Have two cell divisions.

18. Most of the chemical energy released when gasoline burns in a car engine is not used to move the car, but is changed into
   a. Electricity.
   b. Heat.
   c. Magnetism.
   d. Sound.

19. The magnetism of a substance is due essentially to which of the following?
   a. The magnetic properties of its atoms.
   b. The arrangement of its atoms.
   c. The position of its poles.
   d. Both A & B.

20. What is the empirical formula of a compound that contains 0.05 moles of Magnesium (Mg), 0.05 moles of Sulfur (S) and 0.20 moles of oxygen?
   a. MgSO
   b. MgSO₂
   c. MgSO₃
   d. MgSO₄
Appendix E. Pretest Related to the Science Content Discrepancies

Please circle the response that BEST answers each of the following questions.
Choose ONE and only one answer for each question.
Please be sure to answer EVERY question to the best of your ability.

1. When pulling on both ends of a stretched, taut string at the same time, what would happen?
   a. The string remains straight and nothing would happen.
   b. The string will bend and a wave will permeate throughout the string.
   c. A small wave between the ends of the string will occur.
   d. You have to pull up on both ends to create a wave.

2. Why is the melting point of a substance not considered to be a chemical property?
   a. Because chemical properties are observed when a substance takes part in chemical reaction, i.e., when salt mixes with water.
   b. Because chemical properties are observed when a substance takes part in a chemical reaction, i.e., when water turns to steam.
   c. Because physical properties are observed when the chemical identity of a substance is changed.
   d. Because the melting point of a substance results in a change.

3. Why would you expect to find a lower level of relative humidity during the summer than during the spring?
   a. As the temperature goes down, relative humidity goes up.
   b. As the temperature goes up, relative humidity goes down.
   c. Temperatures are lower in the spring so that the relative humidity is lower.
   d. Colder atmospheres result in more water vapor.

4. Why doesn’t the rabies virus affect all species?
   a. Viruses can only enter cells that have specific target molecules and only specific animals have the same target molecules.
   b. Viruses can enter all molecules and all animals have the same target molecules.
   c. Rabies can only affect animal species.
   d. Mammals have target molecules in common.
5. What is the relationship between the free-fall acceleration of an object and the object's mass?
   a. Acceleration is independent of the object’s mass, density or shape.
   b. Acceleration is dependent on the object’s mass, density or shape.
   c. There are multiple forces at work.
   d. An object free-falls at different rates and speeds.

6. Why is the temperature in a valley higher, on average, than the temperature at the peak of the corresponding mountain?
   a. The lower part of the atmosphere is warmer, and decreases with increasing altitude.
   b. The lower part of the atmosphere is colder, and increases with increasing altitude.
   c. Radiation is gained at higher altitudes.
   d. Radiation at the peak of a mountain is lost.

7. How are energy and the function of enzymes related?
   a. Enzymes speed up the rates of reaction and lower the energy required to start the reaction.
   b. Enzymes slow down the rates of a reaction and raise the energy required to start the reaction.
   c. Some enzymes decrease reaction rates.
   d. Enzymes are proteins that impact reactions.

8. Why is there a positive charge in the nucleus of an atom?
   a. The nucleus only contains protons and neutrons and no electrons.
   b. The nucleus only contains electrons and no protons and neutrons.
   c. Electrons have a positive charge.
   d. The components of an atom contain neutrons, protons and electrons.

9. How does brightness of a star relate to its distance?
   a. Stars at a farther distance will appear fainter; and those closer will appear brighter.
   b. Stars at a farther distance will appear brighter; and those closer will appear fainter.
   c. Difference in brightness is related to how close the star is to the sun.
   d. The brightness of a star is called the absolute magnitude.
10. Why do we refer to ions as “charged” molecules?
   a. Ions are formed by an atom gaining or losing electrons.
   b. Ions are formed by an atom gaining or losing neutrons.
   c. Cations are negatively charged particles.
   d. Molecules are partially held together by ionic bonds.

11. How is the acceleration of an object easily increased?
   a. Increasing the force of an object increases its acceleration.
   b. Decreasing the force of an object decreases its acceleration.
   c. Newton’s first law says velocity increases as a result of no force.
   d. Adding force to an object affects its acceleration.

12. What is considered to be the third stage of demographic transition?
   a. During the third stage of demographic transitions, birth rates fall.
   b. During the third stage of demographic transition, birth rates rise.
   c. It is a transitional period seen in less developed countries.
   d. The difference between birth and death rates is small.
Appendix F. Graph Comprehension Pretest

Taken from Burkett and Azevedo, 2012.

Please circle the response that BEST answers each of the following questions. Choose ONE and only one answer for each question. Please be sure to answer EVERY question to the best of your ability.

Please use the following graph to answer questions 1–4:

![Graph of Average Temperatures in U.S. Cities](image)

1. What is displayed on the **y**-axis of the graph?
   a. Temperature.
   b. Seasons.
   c. U.S. cities.
   d. This graph does not have a **y**-axis.

2. What is displayed on the **x**-axis of the graph?
   a. Temperature.
   b. Seasons.
   c. U.S. cities.
   d. This graph does not have an **x**-axis.
3. According to the graph, which city has the highest average temperature during the summer?
   a. Memphis.
   b. Boston.
   c. Los Angeles.
   d. This information is not represented on this graph.

4. According to the graph, which U.S. city has an average temperature that is consistently cooler than the other cities, regardless of the season?
   a. Boston.
   b. Los Angeles.
   c. Memphis.
   d. None of the cities represented on the graph have consistently cooler temperatures.

For numbers 5–6 please plot the following points on the graphs provided:

5. (−4,2)  

6. (5,−3)
Please use the following graph to answer questions 7–8:

7. What was the trend of coffee sales during the weeks surrounding final exams during the fall semester?
   a. Coffee sales increased dramatically.
   b. Coffee sales decreased marginally.
   c. Coffee sales remained constant.
   d. This information is not represented on this graph.

8. In which month did college students purchase the least amount of coffee?
   a. August.
   b. September.
   c. October.
   d. November.
9. In the graph below, no axis or origin is shown. If point B’s coordinates are (4,5), which of the following coordinates would most likely belong to point A?

- a. (2,2)
- b. (−2,2)
- c. (2,−6)
- d. (1,2)

Please use the following graph to answer questions 10–12:

10. Between which two months did the infant gain the most weight?
   - a. April and May.
   - b. June and July.
   - c. August and September.
11. How much weight did the infant gain during the 6-month period?
   a. 13.7 pounds.
   b. 6.5 pounds.
   c. 7.2 pounds.
   d. 6.7 pounds.

12. Between which two months did the infant gain the least weight?
   a. May and June.
   b. August and September.
   c. July and August.
   d. June and July.
Appendix G. Perceived Interest Questionnaire (PIQ)

Taken from Schraw, Bruning, and Svoboda, 1995.

In this part we want you to rate how you responded to each of the science topics overall. Please indicate how strongly you agree or disagree with each statement using the 5-point scale shown below.

<table>
<thead>
<tr>
<th>Strongly Disagree</th>
<th>Disagree</th>
<th>Neutral</th>
<th>Agree</th>
<th>Strongly Agree</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
</tr>
</tbody>
</table>

_____ 1. I thought the science content was very interesting

_____ 2. I’d like to discuss these science topics with others at some point

_____ 3. I would read this science content again if I had the chance

_____ 4. I got caught-up in the science content without trying to

_____ 5. I’ll probably think about the implications of this science content for some time to come

_____ 6. I thought the science content was fascinating

_____ 7. The science content was personally relevant to me

_____ 8. I would like to read more about these science topics in the future

_____ 9. This science content was one of the most interesting things I’ve read in a long time

_____ 10. The science content really grabbed my attention
Appendix H. Example Slide with No Discrepancy

A virus is a small, infectious agent that is simply genetic information (DNA or RNA) enclosed in a protein coat. A virus is much smaller than a cell. At about 10µm (microns) in diameter, an average human cell is perhaps one-tenth the diameter of a human hair. It would take millions of viral particles to form a dot the size of a printed period. A virus does not have a nucleus, organelles, ribosomes, or even cytoplasm and only a few types of viruses contain enzymes. All viruses share two features: genetic information and a protein coat.

All viruses contain genetic material (either DNA or RNA) that carries instructions to make their molecular components. The major criterion for classifying viruses is whether the genetic material is DNA or RNA. The capsid, or protein coat, surrounds the genetic material. The capsid's shape determines a virus's overall form, which is another characteristic used in classification.

The host range of a virus is composed of the kinds of organisms or cells that it can infect. Viruses can enter only cells that have a specific target attachment molecule, or receptor, on their surfaces. Animals, fungi, plants, protists and bacteria all get viral infections. The rabies virus can infect humans, skunks, raccoons, bats and dogs because these mammals have common target molecules on their cells. Some target molecules are on a very small subset of cells in an organism, whereas others occur in a group of related organisms.
Appendix I. Example Slide with Text Discrepancy Emphasized

The first evidence for the existence of subatomic particles came from studies of the conduction of electricity through gases at low pressures. J. J. Thomson showed in 1897 that colored rays of light (cathode rays) consist of a stream of negatively charged particles, which he called electrons. Every atom contains a definite number of these negatively charged particles.

A few years after Thomson’s groundbreaking work, a series of experiments carried out under the direction of Ernest Rutherford in 1911 shaped our ideas about the nature of the atom. Rutherford showed that the scattering of particles bombarded at a piece of thin gold foil was caused by a small positively charged nucleus at the center of the gold atom.

The nucleus of an atom can be considered to consist of two different types of particles: protons and neutrons. The proton has a mass nearly equal to that of an ordinary hydrogen atom. The proton carries a unit positive charge (+1), equal in magnitude to that of the electron (-1). The neutron is an uncharged particle with a mass slightly greater than that of a proton. Because protons and neutrons are much heavier than electrons, most of the mass of an atom (>99.9%) is concentrated in the nucleus, even though the volume of nucleus is much smaller than that of an atom. Electrons are found in the outer regions of the atom, where they form what is considered to be a cloud of positive charge.
Appendix J. Example Slide with Text and Graph Discrepancies Emphasized

Force, loosely speaking, is a push or pull on an object. The force is said to act on the object to change its velocity. For example, when a car slams into a telephone pole, a force on the car from the pole causes the car to stop. The relation between a force and the acceleration it causes was first understood by Isaac Newton.

Newton’s first law states that if no force acts on a body, the body’s velocity cannot change; that is, the body cannot accelerate. In other words, if the body is at rest, it stays at rest. If it is moving, it continues to move with the same velocity (same magnitude and same direction). A force is measured by the acceleration it produces. The greater the force, the greater the acceleration. When two or more forces act on a body, we can find their net force, or resultant force, by adding the individual forces.

A single force that has the magnitude and direction of the net force has the same effect on the body as all the individual forces together. This is known as the principle of superposition for forces. The world would be quite strange if, for example, you and a friend were to pull on the standard body in the same direction, each with a force of 1 Newton and yet somehow the net pull was 14 Newtons. Likewise, there may be multiple forces acting on a body, but if their net force is zero, the body cannot accelerate.

How could you easily increase the acceleration of an object?

![Acceleration vs. Force](image-url)
Appendix K. Example of Experimental Setup
Appendix L. Illustration of Experimental Procedure

Preselection (Completed Online)

- Pretests
  - Science knowledge
  - Discrepancy-related science knowledge
  - Graph comprehension

Participant Scheduled (Within 2 Days)

Experimental Session

- Set-Up
  - Informed Consent
  - Demographic Questionnaire
  - Eye-Tracker Calibration
  - EDA Baseline Collection
  - Experimental Instructions

- Question Presentation
  - "Why are ions called 'charged' molecules?"

- Ease of Learning Judgment

  - Content
    - "Ions are considered..."

  - Graph
    - (Graph depicting charges of molecules)

- Immediate Judgment of Learning

- Graph Judgment of Learning

Screen with Stop Sign (30 seconds)

- Delayed Judgment of Learning

- Graph Judgment of Learning

- Question Re-presentation
  - "Why are ions called 'charged' molecules?"

- Question Response

- Retrospective Confidence Judgment

Conclusion

- Debrief and pay participant
Appendix M. Example of an Ease of Learning (EOL) Judgment Layout

How easy do you think it will be to learn the information needed to answer this question?

<table>
<thead>
<tr>
<th>Confidence</th>
<th>0%</th>
<th>20%</th>
<th>40%</th>
<th>60%</th>
<th>80%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
<td>O</td>
</tr>
</tbody>
</table>
Appendix N. Example Slide as Presented to Students with Attention Tool 5.4

A projectile is a particle that moves in two dimensions. A projectile moves in a vertical plane with an initial velocity, but its acceleration is always the free-fall acceleration, which is a constant rate at which an object accelerates downward. The magnitude of the free-fall acceleration near Earth’s surface is 9.8 m/s². This acceleration is independent of the object’s characteristics such as mass, density or shape; it is the same for all objects. Therefore, if you were to allow a feather and an apple to free fall in a vacuum, with the absence of air, the apple would hit the bottom before the feather.

This two-plane motion is known as projectile motion. A projectile might be a tennis ball or baseball in flight, but it is not an airplane or duck in flight. Many sports involve the projectile motion of a ball, and much effort is spent trying to control that motion for an advantage. For example, the racquetball player who discovered the now famous Z-shot in the 1970s easily won his games because the ball’s peculiar flight to the rear of the court always perplexed his opponents.

In projectile motion, the horizontal motion and the vertical motion are independent of each other, that is, neither motion affects the other. This feature allows us to break up a problem involving two-dimensional motion into two separate and easier one-dimensional problems, one for the horizontal motion and one for the vertical motion.
Appendix O. Example of an Immediate and Delayed Text-Based Judgment of Learning (JOL) Layout

How well did you understand the information about the relationship between gravity and projectile motion presented in the text you just read?

<table>
<thead>
<tr>
<th>Confidence</th>
<th>0%</th>
<th>20%</th>
<th>40%</th>
<th>60%</th>
<th>80%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>O</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix P. Example of an Immediate and Delayed Graph-Based Judgment of Learning (JOL) Layout

How well did you understand the information about the relationship between gravity and projectile motion presented in the graph you just reviewed?

<table>
<thead>
<tr>
<th>Confidence</th>
<th>0%</th>
<th>20%</th>
<th>40%</th>
<th>60%</th>
<th>80%</th>
<th>100%</th>
</tr>
</thead>
</table>
Appendix Q. Example of a Response Input Layout

What is the relationship between the free-fall acceleration of an object and the object's mass?
Appendix R. Example Retrospective Confidence Judgment (RCJ) Layout

How confident are you that the answer you provided is correct?

<table>
<thead>
<tr>
<th>Confidence</th>
<th>50%</th>
<th>60%</th>
<th>70%</th>
<th>80%</th>
<th>90%</th>
<th>100%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Confidence</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
<td>○</td>
</tr>
</tbody>
</table>
**Appendix S. List of Concept Units**

Taken from Burkett and Azevedo, 2012.

<table>
<thead>
<tr>
<th>Text</th>
<th>Concept Units</th>
</tr>
</thead>
<tbody>
<tr>
<td>Projectile Motion</td>
<td>1) The object’s mass doesn’t play a role in free-fall acceleration</td>
</tr>
<tr>
<td></td>
<td>2) Gravity is the only force at work in free-fall acceleration</td>
</tr>
<tr>
<td></td>
<td>3) Objects drop at a constant rate</td>
</tr>
<tr>
<td>Demographic Transition</td>
<td>1) In the third stage birth rates fall</td>
</tr>
<tr>
<td></td>
<td>2) In the third stage the difference between birth rate and death rate is small</td>
</tr>
<tr>
<td></td>
<td>3) The world’s most developed countries are in stage 3 of the demographic transition</td>
</tr>
<tr>
<td>Molecules and Ions</td>
<td>1) Ions are formed when electrons are lost or gained</td>
</tr>
<tr>
<td></td>
<td>2) Ions that lose electrons are positive, ions that gain electrons are negative</td>
</tr>
<tr>
<td></td>
<td>3) Positive ions are called cations, negative ions are called anions</td>
</tr>
<tr>
<td>Humidity</td>
<td>1) As the temperature goes down, relative humidity goes up</td>
</tr>
<tr>
<td></td>
<td>2) The colder the atmosphere the more water vapor present because the evaporation rate decreases faster than the condensation rate</td>
</tr>
<tr>
<td></td>
<td>3) Temperatures are colder in Spring than in Summer</td>
</tr>
<tr>
<td>Structure of the Atmosphere</td>
<td>1) More radiation is lost at higher altitudes</td>
</tr>
<tr>
<td></td>
<td>2) The Earth is warmed from the ground up due to the Greenhouse effect</td>
</tr>
<tr>
<td></td>
<td>3) The peak of a mountain is a higher altitude than the valley</td>
</tr>
<tr>
<td>Force</td>
<td>1) Newton’s first law states that no force = no acceleration</td>
</tr>
<tr>
<td></td>
<td>2) If you increase the force on the object, you increase the acceleration</td>
</tr>
<tr>
<td></td>
<td>3) You can increase the force of an object by adding additional force in the same direction as the original force</td>
</tr>
<tr>
<td>Viruses</td>
<td>1) Viruses can only enter cells that have specific target molecules</td>
</tr>
<tr>
<td></td>
<td>2) Only certain animals have the specific target molecules</td>
</tr>
<tr>
<td></td>
<td>3) Animals who have common molecules can spread the disease to each other</td>
</tr>
</tbody>
</table>
| Components of the Atom | 1) Protons are positively charged and neutrons are neutral  
  2) Electrons are negatively charged  
  3) The nucleus of an atom contains only protons and neutrons, but no electrons |
|------------------------|---------------------------------------------------|
| Brightness of Stars    | 1) Astronomers calculate stars’ absolute magnitude to compensate for distance differences  
  2) Absolute magnitude is used calculate brightness based on differences of stars  
  3) The farther away the star is the fainter it is, the closer the star, the brighter |
| Properties of Substances | 1) Chemical properties are observed when a substance takes part in a chemical reaction  
  2) Physical properties are those for which there is no change in chemical identity  
  3) The melting point of an object only changes states (solid to liquid), not chemical identity |
| Waves                  | 1) If there is no wave the string remains straight  
  2) You have to pull up and then back down to create a wave  
  3) If you pulled up on both ends no wave would happen |
| Enzymes                | 1) Enzymes speed up the rate of the reaction  
  2) Enzymes lower the energy required to start a reaction  
  3) Enzymes are proteins whose function is to catalyze reactions |
Appendix T. Example of Drawn Areas of Interest (AOIs) for a No Discrepancy Slide

There are three main types of waves: mechanical waves, electromagnetic waves, and matter waves. Mechanical waves are most familiar because we encounter them almost constantly. Common examples include water waves and sound waves. These waves have two central features: They are governed by Newton’s laws, and they can exist only within a material medium, such as water and air.

A wave sent along a stretched, taut string is the simplest mechanical wave. If you give one end of a stretched string a single up-and-down jerk, a wave in the form of a single pulse travels along the string. This pulse and its motion can occur because the string is under tension. When you pull your end of the string upward, it begins to pull upward on the adjacent section of the string via tension between these two sections. As the adjacent section moves upward, it begins to pull the next section upward, and so on. Meanwhile, you have pulled down on our end of the string. As each section moves upward in turn, it begins to be pulled back downward by neighboring sections that are already on the way down.

Electromagnetic waves are less familiar, but you use them constantly: common examples include television waves, microwaves, x rays and radar waves. These waves require no material medium to exist. Matter waves are commonly used in modern technology, though they are probably very unfamiliar to you. These waves are associated with electrons, protons, and other fundamental particles, even atoms and molecules.
Appendix U. Example of Drawn Areas of Interest (AOIs) for a Text Discrepancy Slide

The first evidence for the existence of subatomic particles came from studies of the conduction of electricity through gases at low pressures. J. J. Thomson showed in 1897 that colored rays of light (cathode rays) consist of a stream of negatively charged particles, which he called electrons. Every atom contains a definite number of these negatively charged particles.

A few years after Thomson’s groundbreaking work, a series of experiments carried out under the direction of Ernest Rutherford in 1911 shaped our ideas about the nature of the atom. Rutherford showed that the scattering of particles bombarded at a piece of thin gold foil was caused by a small positively charged nucleus at the center of the gold atom.

The nucleus of an atom can be considered to consist of two different types of particles: protons and neutrons. The proton has a mass nearly equal to that of an ordinary hydrogen atom. The proton carries a unit positive charge (+1), equal in magnitude to that of the electron (-1). The neutron is an uncharged particle with a mass slightly greater than that of a proton. Because protons and neutrons are much heavier than electrons, most of the mass of an atom (>99.9%) is concentrated in the nucleus, even though the volume of nucleus is much smaller than that of an atom. Electrons are found in the outer regions of the atom, where they form what is considered to be a cloud of positive charge.

Why does the nucleus of an atom have a positive charge?

Protons and Neutrons of Some Common Elements

- Number of Protons
- Number of Neutrons

Hydrogen Carbon Sodium Potassium Calcium Iron

0 5 10 15 20 25 30

Q
Time spent: 1.6s
Fixations: 272

O
Time spent: 7.6s
Fixations: 1048
Appendix V. Example of Drawn Areas of Interest (AOIs) for a Text and Graph Discrepancy Slide

Force, loosely speaking, is a push or pull on an object. The force is said to act on the object to change its velocity. For example, when a car slams into a telephone pole, a force on the car from the pole causes the car to stop. The relation between a force and the acceleration it causes was first understood by Isaac Newton.

Newton’s first law states that if no force acts on a body, the body’s velocity cannot change; that is, the body cannot accelerate. In other words, if the body is at rest, it stays at rest. If it is moving, it continues to move with the same velocity (same magnitude and same direction). A force is measured by the acceleration it produces. The greater the force, the greater the acceleration. When two or more forces act on a body, we can find their net force, or resultant force, by adding the individual forces.

A single force that has the magnitude and direction of the net force has the same effect on the body as all the individual forces together. This is known as the principle of superposition for forces. The world would be quite strange if, for example, you and a friend were to pull on the standard body in the same direction, each with a force of 1 Newton and yet somehow the net pull was 14 Newtons. Likewise, there may be multiple forces acting on a body, but if their net force is zero, the body cannot accelerate.
Appendix W. Correlation Matrix with Eye-Movement Variables and Metacognitive Judgments

|               | 1    | 2    | 3    | 4    | 5    | 6    | 7    | 8    | 9    | 10   | 11   | 12   | 13   | 14   | 15   | 16   | 17   | 18   | 19   | 20   | 21   | 22   | 23   | 24   | 25   | 26   |
|---------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| Fixed time    | -    | 1.00* | -    | 0.05 | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    |
| Proportion   | 0.07 | -0.05 | 1.00* | -    | 0.05 | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    |
| Proportion   | 0.07 | -0.05 | 0.97** | 1.00* | -    | 0.05 | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    |
| Proportion   | 0.07 | -0.05 | 0.97** | -0.03 | -    | 1.00* | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    |
| Proportion   | 0.07 | -0.05 | 0.97** | -0.03 | -0.03 | -    | 1.00* | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    |
| Proportion   | 0.07 | -0.05 | 0.97** | -0.03 | -0.03 | -0.03 | -    | 1.00* | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    |
| Proportion   | 0.07 | -0.05 | 0.97** | -0.03 | -0.03 | -0.03 | -0.03 | -    | 1.00* | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    |
| Proportion   | 0.07 | -0.05 | 0.97** | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -    | 1.00* | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    |
| Proportion   | 0.07 | -0.05 | 0.97** | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -    | 1.00* | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    |
| Proportion   | 0.07 | -0.05 | 0.97** | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -    | 1.00* | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    | -    |
| Proportion   | 0.07 | -0.05 | 0.97** | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -0.03 | -    | 1.00* | -    | -    | -    | -    | -    | -    | -    | -    | -    |

Note. EOLs = ease of learning judgments; JOLs = judgments of learning; RCJs = retrospective confidence judgments.