

THE COHERENCE FORMATION MODEL OF ILLUSTRATED TEXT
COMPREHENSION:
A PATH MODEL OF ATTENTION TO MULTIMEDIA TEXT

A Dissertation
Submitted to
the Temple University Graduate Board

in Partial Fulfillment
of the Requirements for the Degree
DOCTOR OF PHILOSOPHY

by
Shannon Leigh Fitzhugh
January, 2012

Examining Committee:

Thomas F. Shipley, Advisory Chair, Department of Psychology
Jennifer G. Cromley, Department of Psychological Studies in Education
Nora S. Newcombe, Department of Psychology
Jason Chein, Department of Psychology
Julie L. Booth, Department of Psychological Studies in Education
Adam Davey, Department of Public Health

© Copyright
Shannon L Fitzhugh
2012

Abstract

The study reported here tests a model that includes several factors thought to contribute to the comprehension of static multimedia learning materials (i.e. background knowledge, working memory, attention to components as measured with eye movement measures). The model examines the effects of working memory capacity, domain specific (biology) and related domain (geoscience) background knowledge on the visual attention to static multimedia text, and their collective influence on reading comprehension. A similar model has been tested with a previous cohort of students, and has been found to have a good fit to the data (Fitzhugh, Cromley, Newcombe, Perez and Wills, 2010). The present study tests the efficacy of visual cues (signaling) on the comprehension of multimedia texts and the effects of signaling on the relationships between cognitive factors and visual attention. Analysis of Covariance indicated that signaling interacts with background knowledge. Signaling also changes the distribution of attention to varying components of the multimedia display. The path model shows that signaling alters the relationship between domain specific background knowledge (biology) and comprehension as well as that of related background knowledge (geoscience) on comprehension. The nature of the relationships indicates that the characteristics of the reading material influence the type of background knowledge that contributes to comprehension. Results are discussed in terms of their application to a classroom setting.

Acknowledgements

Thanks to the following people who have helped me achieve this goal in my life: my parents Jeffrey A. Fitzhugh and Christa May, thank you for all of your support and tough love; my brother Jeffrey M. Fitzhugh for the letter on beautiful cloud paper, I read still read it; Lori Monahan for pushing me to finish my undergraduate degree without which I would never have gotten this far; Greg Zelinsky for teaching me eye tracking; Mark Neider for giving me a hard time and a helping hand as an undergraduate RA; Thomas F. Shipley for providing the opportunity to help start the eye tracking lab at Temple University; Nora Newcombe for the opportunity to be a part of SILC, for funding the equipment for the eye tracking lab, and many other things; Jason Chein for always asking me the tough questions; Adam Davey for having a sense of humor about S.E.M and for being a great teacher; Julie Booth for very insightful comments on previous versions of this paper; Dominique Dumay for taking over the lab duties and being a wonderful support person; Mike Cohen for reminding me that I am tough as nails; Alexandra Morrison for being my cheerleader; Karen Fecenko-Taka for letting me cry on your shoulder; Chris Barcia for always being able to make me smile; Jennifer Cromley, I never could have done this without your support and pep talks, I am forever in your debt and always your friend; my Lillian, you are the light of my life and my inspiration for everything I do.

TABLE OF CONTENTS

ABSTRACT	iii
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	ix
LIST OF FIGURES	x
CHAPTER 1 – INTRODUCTION	1
DESCRIPTION OF STUDY	1
SCIENCE LITERACY IN THE UNITED STATES	3
THE CONSTRUCTION INTEGRATION MODEL	5
INTEGRATED MODEL OF TEXT AND PICTURE COMPREHENSION	7
REVIEW OF EYE MOVEMENTS AND COGNITIVE PROCESSES	11
BUILDING THE PRELIMINARY STUDY	13
THE PRESENT STUDY	15
CHAPTER 2 – REVIEW OF THE LITERATURE	17
PATH MODELS	17
SELECTION AND DEFINITION OF VARIABLES	19
THE COHERENCE FORMATION OF ILLUSTRATED TEXT COMPREHENSION	21
Path 1a: Background Knowledge on Comprehension	23

CHAPTER 2 – REVIEW OF THE LITERATURE (continued)	
Path 1b: Background Knowledge on Attention to Text.....	23
Path 1c: Background Knowledge on Attention to Diagrams.....	24
Path 2a: Working Memory on Comprehension.....	26
Path 2b: Working Memory on Attention to Text.....	27
Path 2c: Working Memory on Attention to Diagrams.....	28
Path 3: Attention to Text on Comprehension.....	29
Path 4: Attention to Diagrams on Comprehension.....	30
Signaling on Attention to Text and Attention to Diagrams.....	31
CHAPTER 3 – METHODS.....	33
PARTICIPANTS.....	33
RESEARCH DESIGN.....	34
MATERIALS AND MEASURES.....	34
PROCEDURE.....	40
SCORING AND CODING.....	43
CHAPTER 4 – RESULTS.....	48
DESCRIPTIVES.....	48
First Order Variables.....	48
Second Order Variables.....	57
SCREENING FOR MULTIVARIATE ASSUMPTIONS.....	60
RESULTS.....	61

CHAPTER 4 – RESULTS (continued)

Analysis of Variance	62
Cluster Analysis	63
Analysis of Covariance	64
Regression Analysis	72
Factor Analysis	74
Path Model	76
Series of Models	78
CHAPTER 5 – DISCUSSION	85
RESEARCH QUESTION 1	85
RESEARCH QUESTION 2	86
RESEARCH QUESTION 3	87
RESEARCH QUESTION 4	89
RESEARCH QUESTION 5	90
RESEARCH QUESTION 6	91
RESEARCH QUESTION 7	93
LIMITATIONS OF THE STUDY	94
IMPLICATIONS OF THE STUDY	94
THEORETICAL CONTRIBUTIONS	96
CONCLUSION	97
REFERENCES	99

APPENDICES	114
Appendix A – IRB approval to use data set.....	114
Appendix B – Letter to Parents describing larger study.....	115
Appendix C – Consent form for Parents.....	116
Appendix D – Biology Background Knowledge Measure.....	117
Appendix E – Geoscience Background Knowledge Measure.....	121
Appendix F – Example stimuli from control group.....	123
Appendix G – Example stimuli from numbered group.....	124
Appendix H – Example stimuli from hyperlink group.....	125
Appendix I – Scoring Rubric for comprehension answers.....	131
Appendix J – Correlation Matrix of all variables.....	133

LIST OF TABLES

1. Table 1 – Descriptives for Raw Variables.....	50
2. Table 2 – Descriptives for Unstandardized Residuals	54
3. Table 3 – Chi-square difference test for models	79
4. Table 4 – Direct and Indirect Effects Variables on Comprehension.....	84

LIST OF FIGURES

1. Figure 1 – Schematic Representation of the Integrated Model of Text and Picture Comprehension	8
2. Figure 2 – Example diagram showing iconic and symbolic representation..	9
3. Figure 3- Example Path Diagram.....	19
4. Figure 4 – The Coherence Formation Model of Illustrated Text Comprehension.....	22
5. Figure 5 – Schematic of the Automated Working Memory Test.....	37
6. Figure 6 – Fitted Path Diagram with Standardized Estimates for Control Group.....	80
7. Figure 7 – Fitted Path Diagram with Standardized Estimates for Signal Group.....	81

CHAPTER 1 - RATIONALE

Description of the Study

The research described in the following manuscript tests a model that includes several factors thought to contribute to the comprehension of multimedia learning materials (i.e. background knowledge, working memory, attention to components as measured with eye movement measures). The model examines the effects of working memory capacity and domain background knowledge, on the amount of visual attention to text and accompanying diagrams, and their collective influence on reading comprehension. A similar model has been tested with a previous cohort of students from the same long term research grant, and has been found to have a good fit to the data (Fitzhugh, Cromley, Newcombe, Perez and Wills, 2010). In the study described here, the previous model is modified with the addition of a signaling manipulation for guiding text and diagram integration, thought to vary in efficacy with individual characteristics of the student (Bartholome & Bromme, 2009).

The study reported here seeks to; a) examine the data for the presence of “groups” as determined by the combination of working memory and background knowledge scores; b) confirm that the preliminary model holds in a new sample of participants; c) attempt to understand the way in which signaling influences the visual attention to multimedia components, and subsequent comprehension of materials.

The presence of “groups” based on working memory and background knowledge was tested using cluster analysis to determine if there were profiles of students based on the scores of these two measures. The fit of path model was used to determine if the

model replicates with a new group of students from the same school as the preliminary model. In addition, the fit of the model for both signaled and non-signaled groups was tested to determine if the relationships between variables were the same across experimental manipulations.

The signaling manipulation mimics that of Bartholome and Bromme (2009), who tested the efficacy of two different signaling conditions in undergraduate non-biology majors learning about plant classification. The two cues used as signals were; static numbered cues in running text and corresponding diagrams (or diagram components), and hyperlink cues, in which a portion of hyperlinked text highlighted the corresponding diagram portion when clicked. The numbered format lead to deeper understanding and thus higher scores on a classification of flora test (deep processing), and the hyperlink condition was found to be detrimental to performance on the same task. Based on subject self-reported confidence and ease in learning the materials, the authors hypothesized a shallower processing effect in the hyperlink condition brought about by the perceived lack of effort needed to coordinate textual and diagram components.

An alternative to the authors' hypothesis proposed here is that segmentation of the diagram in the hyperlink condition due to the highlighting of specific diagram portions leads to less global processing or comparisons between diagram components than does the numbered condition. The numbered condition, while providing signals, forces the learner to visually segment the diagram themselves, leading to more comparisons between diagram portions. To answer this question, eye tracking data will be examined for switches made from one portion of the diagram to another portion of the same diagram, a behavior which indicates relations between diagram parts were examined.

In the following sections, the need for research on improving comprehension of science text is examined and the selection of variables for the path model under examination is discussed. First, the Science, Technology, Engineering and Mathematics (S.T.E.M.) crisis is outlined. This includes a discussion of the learning materials used in typical American classrooms and the cognitive processes thought to contribute to their successful use.

Science Literacy in the United States.

A report published by the Program for International Student Assessment (PISA) examining science literacy in 15 year-old students across 57 countries showed that overall the U.S. has fallen to below average ranking (range 24 to 35 of 57) in science literacy worldwide (Baldi, Jin, Skemer, Green, & Herget, 2007). In addition, U.S. students are statistically over-represented in the lowest 2 levels of science literacy (below level 1 and level 1) and under-represented in the average levels of science literacy (levels 3 & 4). Generally, students in the United States have rudimentary to functional knowledge of scientific principles, but lack the ability to reason and problem solve using the principles of scientific exploration. This means that contributions from the United States to the scientific community in terms of research and development may be limited.

In addition to overall levels of understanding and applications of science, the PISA also examines efficacy in 3 content areas, Earth and Space systems, Living Systems (Biology/Chemistry) and Physical Systems (Physics). While the United States scored above average on Earth and Space systems, scores were at or below average for living and physical systems. Despite low scores, children in the United States have an above average awareness of environmental issues, such as ozone depletion and habitat

destruction. Attention to instruction and learning in biology can help American students acquire the knowledge needed to act upon their environmental awareness, encouraging environmentally aware individuals to become contributing members to the field of environmental technology. A greater understanding of living systems in conjunction with heightened awareness is a formula for innovative change in the technology and management of living systems of the Earth.

There currently exists a small body of research in the comprehension of illustrated science text, and the results have been informative. However, the participants consist primarily of undergraduate non-science majors, which is problematic for two reasons. First, the PISA report (2006) highlights the level of understanding of scientific principles needed for pursuit of careers in science (levels 5 and 6) is lacking at age 15, and unlikely to reach proficiency levels by the end of high school (age 18). Thus, examination of the effects of multimedia learning on college undergraduates does not address the question of how to help K-12 students improve in science literacy. Second, there is a sampling problem in that undergraduates have self-selected to attend institutions of higher learning and were selected for academic competency by the standards of the institution. Thus the results of these studies may not be generalized to less selective populations such as children attending public schools. More research needs to be done with middle and high school aged students in science classrooms in order to address the problem of teaching scientific competency.

The review of the existing literature begins in the next section with a brief discussion of the Construction Integration Theory (Kintsch, 1995) as a framework for how background knowledge aids in the formation of stable representations from reading.

The Model of Integrated Multimedia Comprehension (Schnotz, 2005) is based on Kintsch's CI theory and discusses how working memory aids in the construction of stable representations from text and visual information. Following the theoretical review, a brief review of eye tracking as it relates to cognition is presented.

The Construction-Integration Model.

The Construction-Integration (CI) model presented by Walter Kintsch (Kintsch and van Dijk, 1978) is a connectionist model of comprehension with two phases, a *construction* phase where all associations among stimuli are activated, and an *integration* phase by which only the relevant associations are kept and processed. Comprehension in this context is defined as occurring "when and if the elements that enter into the process achieve a stable state in which the majority of the elements are meaningfully related to one another and other elements that do not fit the pattern of the majority are suppressed" (Kintsch, 1998, pg 4). Specifically Kintsch distinguishes between the vernacular "perceive" as an isolated instance of perception and "understanding" which additionally involves the relationship between a concept or object and its context. The model as espoused by Kintsch can apply to cognition as a whole; however for the purpose of this review, will be discussed in terms of text comprehension.

Kintsch's model assumes dual coding theory (Paivio, 1971) is correct and that information is encoded as a modality specific representational unit, with the simultaneous processing of related information for a particular stimulus forming cross modal associations. In the first phase of text comprehension, reading a word activates the word, all of its meanings (vocabulary), all semantically related words (background knowledge) and any images associated with the word or associations. This process can be relatively

simple when background knowledge (experience) is low; not many associations will be activated. However, this can become extremely problematic when background knowledge is high and associations are many. Integration is the process by which the context helps determine the correct meaning of the word presented in the text.

To demonstrate the construction-integration process, let us turn to an example. A participant is presented with a sentence “*The earthquake destroyed all the buildings in the town except the mint*” (Kintsch, 1998; pg 95). At first read, all meanings of the decoded word “*mint*” (place to store money, making money or coins, flavoring added to lamb and chocolate) are activated. Increased background knowledge results in more nodes being activated and thus provides more opportunity for the construction of meaning. However, with all meanings of *mint* activated, the representation is garbled and incoherent. This is where the integration portion of the model becomes critical.

In our example sentence, *mint* has several possible meanings, which are all suppressed by the context of *earthquake* and *building*, except one (place to store money). These words act as constraints by activating the semantically related meanings of *mint* and suppressing the semantically unrelated meanings. This produces a more stable representation of the meaning of the text, or what is termed the text base; a propositional representation derived directly from the text. It has been shown that while the text base formation is relatively automatic, the text base enables only the answering of factoid type questions and is not a stable form of knowledge (Kintsch, Britton, Fletcher & Kintsch, 1993). The more stable form of knowledge, one that enables inferences and elaboration of the text is called the situation model.

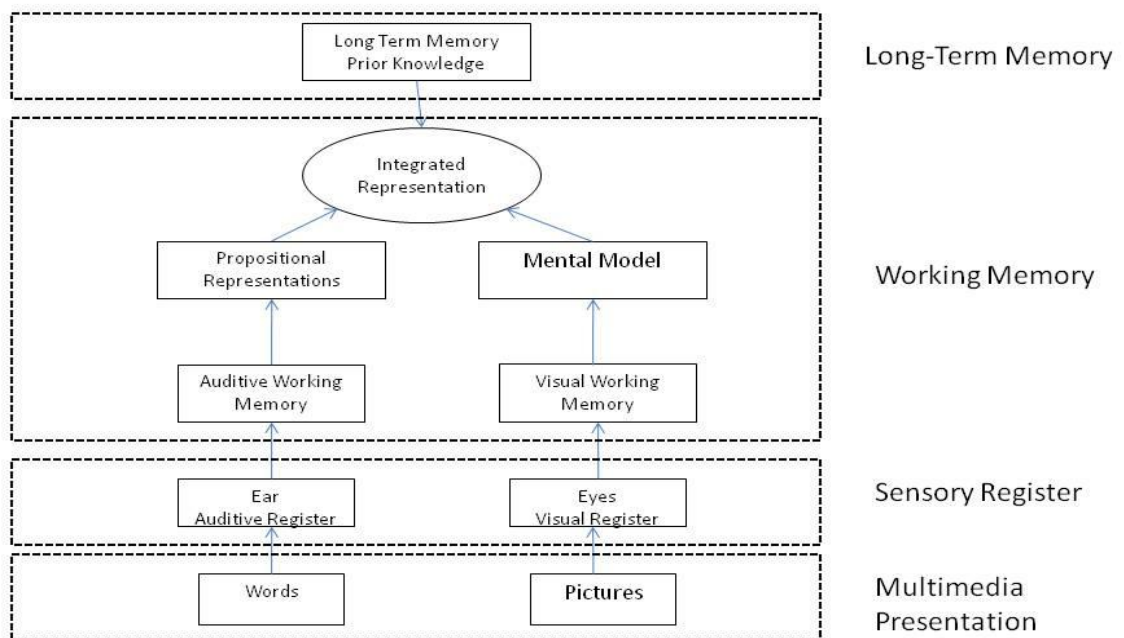
The situation model is defined as the complete structure of the text to be stored in long term memory, incorporating text derived propositions and propositions or mental models from long term memory (Kintsch, 1998). Understanding or comprehension of the text resides in the situation model, the formation of which relies heavily on Background Knowledge. Thus the model predicts that those with low levels of background knowledge will have fewer associations active in their network and thus have less opportunity for integration, forming a text base and situation model that are very similar. Those with high background knowledge form a richer situation model which leads to better understanding, more elaboration and more inferences (for studies supporting the model see Britton & Gulgoz, 1991; Wiley & Voss, 1999; McNamara, Kintsch, Songer and Kintsch, 1996; Grasser, Kessler, Krouoz & McLain-Allen, 1998; E. Kintsch, 1990; Mannes & Kintsch, 1987; MaNamara, 2001; Otero & Campanario, 1990; Singer & Halderson, 1996; Otero & Kintsch, 1992; Schmalhofer, McDaniel, & Keefe, 2002; Singer & Halldorson ,1996; Singer & Kintsch, 2001).

In the next section, the Integrated Model of Text and Picture Comprehension proposed by Schnotz (2005) is a theory that extends Kintsch's CI theory to comprehension of illustrated texts. This theory accounts for some of the individual differences found in the comprehension of multimedia displays. While much of the research focuses on the use of animations, the general principles can be applied to static representations as well.

Integrated Model of Text and Picture Comprehension.

Schnotz (2005) proposed the Integrated Model of Text and Diagram comprehension which shares many aspects with Mayer's model (see Figure 1). Like

Mayer's model of Multimedia Comprehension (Mayer, 2005), information is presented to both modalities (verbal and visual), is selected for processing and enters working memory. The Integrated Model of Text and Picture Comprehension distinguishes the sensory channel from the representational channel. Like dual coding theory, Schnotz's theory proposes that each sensory channel creates a different representational form for the information.



Schematic representation of Mayer's Theory of Multimedia Learning. Integration occurs after two stable models, propositional and mental schematic model are created separately for each type of stimuli. Information from Long term memory is used in the formation of the integrated model.

Figure 1 – Schematic Representation of the Integrated Model of Text and Picture Comprehension (Schnotz, 2005)

Text and descriptive information, whether in auditory (spoken) or visual format, form symbolic representations which bear no physical resemblance to their referent (i.e. the word bird does not resembles an actual bird). Symbolic representations are primarily

responsible for the coding of abstract knowledge. The description of what a bird of prey may eat (i.e. mammals or small reptiles) encompasses a large range of animals with similar characteristics (size, structure, etc.). However, the pictorial representation is iconic in nature, that is, it is tied to the referent by similarity or other structural commonalities (see Figure 2). While the depiction is only able to show an individual food item (mouse), it provides information not included in the text such as size and shape of the prey animal, size and shape of the hunter (or parts such as talons, beak etc.) and other spatial relationships which are useful in making inferences (Kosslyn, 1994).



Figure 900 - This owl uses echolocation and keen eyesight to catch prey. Owls eat small mammals, reptiles and even insects.

Figure 2 – Example diagram showing iconic and symbolic representation

Schnotz (2005) proposes integration occurs in the *formation* of the verbal model (or propositional representations) and the mental model (schematic visual representations). Thus relevant portions of the visual material will be used in the

construction of the propositional model and relevant portions of the verbal material will be included in the mental model. The information from the two models is then integrated further with information from long term memory. Thus although two separate models are created, they are created with *integrated* information rather than domain specific information. This view is consistent with Paivio's dual coding theory in which representational formats are activated simultaneously for stimuli (when presented simultaneously).

The process is thought to be automatic; however, the level of automaticity is altered by background knowledge. Thus for those with low background knowledge the process is more effortful, thus occupying much of their working memory resources. Those with higher background knowledge have more internal structure (i.e. event schemas, mental models, vocabulary, etc) with which to build their knowledge and thus construct more elaborate and more accurate inferences.

Coherence, or the idea that words and pictures should be semantically related in order for them to contribute to the same mental model, has been shown to be a necessary condition for text and picture integration to occur (for evidence see Mayer, Bowe, Bryman, Mars and Tapangco, 1996; Harp and Mayer, 1997, 1998; Moreno and Mayer, 1998; Mayer, Heiser and Lonn, 2001). Additionally, prior knowledge has been shown to be a leading factor in determining the success of use of MERs. Poor readers (i.e. those with low background knowledge) often have limited sources for construction of mental models (situation models), thus adding a picture to the text provides another resource for that construction (Cooney and Swanson, 1987; Levie and Lentz, 1982; Mastropieri and Schruggs, 1989). In addition, those with low background knowledge have difficulty

forming accurate mental models from visual representations without accompanying text information (Kalyuga, Chandler and Sweller, 2000); however, those with high background knowledge can actually benefit from it.

Research in this area is in its infancy. Thus while behavioral studies have been able to uncover individual differences in the comprehension of illustrated science text, the mechanisms of these differences remain largely unknown. The addition of eye tracking measures allows for the moment to moment monitoring of the attention process, allowing researchers to relate the number, duration and patterns of fixations with outcomes. These types of analysis help elucidate the process by which students learning from illustrated text construct meaning out of the materials presented. This is important for designing interventions and teacher aides which address the components of the comprehension process in the dose and sequence which promote understanding. In the following section, the relationship of eye movements to cognitive processes is reviewed and measures used for the study proposed here are defined.

Review of Eye Movements and Cognitive processes

The basic two characteristics of eye movements are saccades (the actual moving of the eyes from one place to another) and fixations (a pause on an object of interest). Saccades generally last 40-60ms (Abrams, Meyer & Kornblum, 1989) and can vary greatly in size depending on the task (Rayner, 1998). The longer a saccade, the faster the eye tends to move (saccadic velocity; Inhoff and Radach, 1998). The length of saccades is an indicator of difficulty of processing. However, since no new information is encoded during saccades (Rayner, 1998) they are often not a focus of analysis.

Cumulative fixation duration is a measure of the total time spent fixating an object and indicates the total visual attention an object received. It is a course grained measure and is often succeeded by finer grained measures. *Fixation duration* is the duration of any single fixation on an object, and is indicative of the time spent encoding that object (Antes, Chang, & Lenzen, 1985; Irwin, 1998; Goldberg & Kotval, 1999). Longer fixation durations indicate difficulty encoding an object and increase with less frequent words (Jacobroon & Dodwell, 1979), longer words (Rayner and Pollatsek, 1982) and complexity of the object (Rayner, 1998).

Number of Fixations indicates the number of times a particular object was fixated. More fixations on an area of interest indicate that area received a lot of visual attention. Interpretation of this increased attention is also dependent on behavioral performance or comparisons with other measures of eye movements. Typically, the distribution of fixations tends to be more informative. *The Proportion of Fixations* on a region both indicates the distribution of fixations across a stimulus display and allows for direct comparisons between groups that make different overall numbers of fixations.

Eye movements have been shown to be directly related to cognitive operations in that they reflect the moment to moment processing of symbols when the referent is visible (Just and Carpenter, 1976, 1985). Thus the locus of a fixation indicates what is currently being processed and the order of fixations indicates the order of the processing. More recent research has shown that in the absence of a referent, such as in imagery tasks, eye movements reflect the spatial content of the imagined scene (Brandt and Stark, 1997).

In addition to reflecting cognitive processes, studies have shown that eye movements can aid in cognitive operations. Grant and Spivey (2003) studied eye movements during a problem solving task (tumor problem). They found a specific pattern of eye movements was correlated with successful solutions. They also found that highlighting the conceptually relevant portions of the diagram increased accurate solutions. In a study of guided eye movements, Thomas (2003) found that guiding eye movements with a cue in a manner consistent with a problem solving solution resulted in faster and more accurate solutions than randomly guided eye movements.

The following section will briefly describe the main findings of the preliminary study conducted for this report. The study incorporated cognitive measures of background knowledge, working memory and eye movement measures of attention to text and diagram components to test a model of comprehension of illustrated instructional texts. This preliminary model was the basis for the dissertation research, and for the modifications made in this report.

Building on the preliminary study.

The Coherence Formation Model of Illustrated Text Comprehension was tested in an attempt to account for how coherence formation occurs when reading illustrated science text. The data for the preliminary study were from Cohort I (Spring 2009) participants from the same 3 year research project as the data presented here. The data were not included in this analysis due to modifications in the comprehension paradigm from year 1 to year 2 of the study.

The preliminary study (see Fitzhugh et al. 2010 for more details) established that the Coherence Formation Model of Illustrated Text Comprehension (CF Model) was a

good fit to the data. Working memory and background knowledge were found to have direct positive effects on comprehension, and significant reciprocal relationships with attention to textual elements of the multimedia display. Increases in either working memory or background knowledge resulted in less attention to the text. However, background knowledge was found to have a significant and positive relationship with attention to the diagram (the relationship between WM and time in diagrams was non-significant).

The CF model was altered slightly from the preliminary study based on the results of the larger intervention study. It was hypothesized that visuo-spatial working memory would be more predictive of comprehension requiring the interpretation of diagram materials. This only visuo-spatial working memory was collected and was entered as an observed independent variable in the path model.

In addition, the effect of a signaling manipulation presented in Cohort III was tested. The manipulation involved two different types of signaling, hyperlink and numbered sequence. The manipulation has been shown to lead to differences in scores on a plant classification task, with numbered sequence participants having higher scores (Bartalome & Bromme, 2009). The model was modified to include the effects of the signaling manipulation by testing the fit of a two group model with Cohort II as the control group and Cohort III as the signaled group. Better fit of the two group model indicates the groups are different and the signaling manipulation alters relationships in the model.

The Present Study.

There is a dire need for research on the comprehension of science text by middle and high-school students. This study seeks to examine some of the relationships established with undergraduate research and test their relationships in a high school sample. As such this study used a data set from NSF REESE Diagrammatic Reasoning study to examine the relationships between working memory, background knowledge, attention to text, and attention to diagrams to comprehension of typical biology textbooks. This study capitalizes on data using ecologically valid stimulus materials in a natural school environment to test how these variables relate in a non-laboratory setting.

The research questions are:

1. Are there “groups” of learners with different profiles of Working Memory and Background Knowledge?
 - a. How many of these groups are there?
 - b. Does the classification of these groups capture the interactions between working memory and background knowledge?
2. Which measures of eye movements are best for examining the effects of attention to illustrated text on comprehension?
3. How does signaling effect the time spent on text and time spent on diagrams?
4. Does the signaling manipulation interact with any of the predictor variables?
5. Using a new sample, does the Coherence Formation Model of Illustrated Text Comprehension (CF Model) fit the data well?
6. Which predictors have the largest effect on attention to text and diagrams as measured with eye movements?

- a. Is the relationship the same between the groups identified?
7. What is the relationship between attention to differing parts of a multimedia display and comprehension of textbook materials?

The following chapter outlines the literature supporting the Coherence Formation Model, followed by an in depth review of the statistical method, including definitions of terms and conventions of notation. Each path in the model is then supported by a literature review of studies supporting the direction of the effect. Chapter III details the methods of data collection for the sample used here and the statistical procedures used in the analysis. Chapter IV presents the results of the current study, and Chapter V discusses the results in terms of the research questions.

CHAPTER 2 – REVIEW OF THE LITERATURE

Raising the level of scientific understanding in our youth is an undertaking that involves understanding many aspects of the learning environment (school, SES, parental education level, etc.) as well as those of the learner (e.g. cognitive, emotional, and motivational). To that end this study focuses on understanding how various cognitive factors of the learner interact to produce comprehension of illustrated science text. This chapter reviews the relevant literature on multimedia science text comprehension and the factors that have been shown to effect comprehension of text and diagram materials.

First discussed are the requirements of Path Models including terminology and conventions of representations. Next the selection of variables is discussed and variables are defined. Then the Coherence Formation Model of Illustrated Text comprehension is presented along with the corresponding path diagram. Finally, review of the literature supporting each of the paths in the model is presented.

Path Models

Path analysis is a technique for estimating the presumed causal relationships between observed variables. However, the analysis is based on the covariance structure of the observed variables. Essentially the model specified must attempt to explain why X and Y are correlated, including theoretically based assumptions of causation, as well as any known spurious relationships between variables.

An acceptable model fit does not mean the model is correct. To reasonably conclude that X causes Y, several challenging qualifications must be met. First is temporal precedence; X must precede Y in time. In experimental research with cross-

sectional measurement, this requirement cannot be met, as data are collected concurrently. Thus the determination of causal paths in a path model must have a substantive rationale. Second the direction of the effect must be correctly specified. Misspecification of the direction of the effect (X cause Y when actually Y causes X) can greatly affect model fit; as can exclusion of a common causal variable (A causes X and Y). The omission of a common causal variable is the most common problem associated with modeling techniques, but is easily detected with the presence of correlated exogenous variables. Third, the relationship between X and Y variables must not disappear with the inclusion of covariates. That is to say, the relationship between X and Y must not be a spurious one.

Identification. A model is identified when it is theoretically possible to derive unique estimates for every parameter in the model; that is to say that $df \geq 0$. The number of free parameters in the model is determined by the number of observations. The number of observations equals $\frac{v(v+1)}{2}$, where v is the number of observed variables. A model with $df = 0$ is said to be just-identified; however a just-identified model is a unique solution and will always fit the data perfectly. An over-identified model ($df > 0$) on the other hand, will not perfectly reproduce the data. Therefore, model testing for over-identified models analyzes the difference in model fit of alternatively specified models to determine which accounts best for the relationships in the input covariance matrix.

Path models have a set of notation practices for construction of path diagrams. Manifest (observed) variables are represented by rectangles and error variances of these variables are represented with circles (See Figure 3). Double headed arrows between variables represent the covariance between the two variables. Double headed arrows

from a variable back to itself indicate the variance of the variable. Straight single headed arrows represent the direction of causality. For example in Figure 3, variable X1 is said to cause variables Y1, Y2. An absence of symbols between two variables indicates the variables have no hypothesized association.

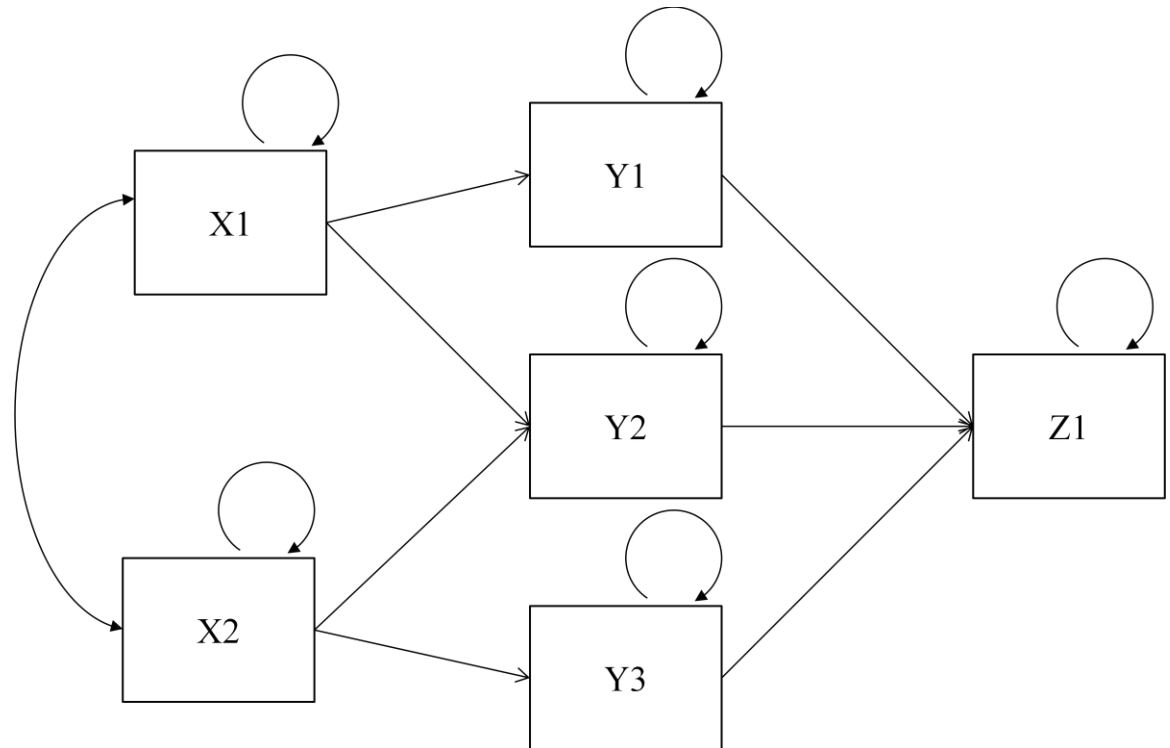


Figure 3. Example Path Diagram

Selection and definitions of variables.

A review of the literature on comprehension generally, and multimedia comprehension specifically, yielded 4 variables thought to be associated with comprehension. Each will be discussed briefly and defined.

Background Knowledge. Background knowledge is defined as the text relevant knowledge a reader possess based on prior experiences, either through reading or other

experiences. Readers can form connections between these experiences and the current text. Background knowledge is usually measured for the domain in which the learning material is situated (e.g. biology, mathematics).

Working memory. Working memory is a complex system of storage and processing components which is responsible for the coordination of processing activities during complex activities such as reading (Daneman and Carpenter, 1980) and problem solving (Kane et al., 2004). Working memory is typically measured with a class of tasks known as complex span tasks. This class of tasks consists of a processing task, interleaved with memory items and has been shown to be related to measures of general intelligence (Kane, Hambrick, Tuholski, Wilhelm, Payne & Engle, 2004).

Attention to text. Attention to text is defined as visual attention to the textual materials and measured by both number of fixations and amount of time spent viewing the particular text. The number of fixations can be analyzed spatially to determine distribution across the display; or they can be analyzed temporally to determine the time course of the visual inspection of text. Fixation durations can be analyzed individually or aggregated by sequences or areas of interest (AOIs; Just and Carpenter, 1976). Many studies examine attention to text in both fixations and time in order to triangulate the locus of processing.

Attention to diagrams. Attention to diagrams is defined as visual inspection of the visual-spatial materials and is measured by both number of fixations and amount of time spent viewing the particular representation. The number of fixations can be analyzed spatially to determine distribution across the display; or they can be analyzed temporally to determine the time course of the visual inspection of the display. Fixation

durations can be analyzed individually or aggregated by sequences or areas of interest (AOIs; Just and Carpenter, 1976). Many studies examine attention to diagrams in both fixations and time in order to better triangulate the locus of processing.

Comprehension. Comprehension is defined as the ability to understand the meaning or importance of something. Comprehension of text is distinguished from retention, which is the ability to recall specific propositions of the text, or recognize the main idea of the proposition. Retention is associated with a stable text-base, while comprehension is associated with a stable situation model. Comprehension is usually assessed with open ended questions or with complex questions which require inference. Similar to BK, comprehension is typically measured in the domain of interest.

Signaling. Signaling refers to the addition of cues to a stimulus which draw attention to the components to which the cue is attached. Signaling can be accomplished through numbering (Bartholome & Bromme, 2009), coloring (Boucheix & Guinard, 2005; Ozelick, Arslean-Ari, & Cagiltay, 2010) and arrows (Boucheix & Lowe, 2010) to name a few. For the purpose of this analysis, signaling was tested through the grouping of participants into groups and group differences were evaluated in the model.

The Coherence Formation Model of Illustrated Text Comprehension.

Coherence formation, as stated in the last section, is the process by which images or pictures are semantically related to the text with which they are presented. The Coherence Formation Model of Illustrated Text Comprehension is an attempt to synthesize a patchwork of research on multimedia comprehension with the goal of understanding how coherence formation occurs. Presented below is a schematic of the CF model of comprehension (see Figure 4). The model predicts that working memory

has a direct effect on comprehension and direct effects on both the attention to text and the attention to diagrams. There are also hypothesized indirect effects of working memory through the two attention variables. The model predicts similar effects for background knowledge. The model also predicts that attention to text and attention to diagrams have direct effects on the comprehension of the materials.

Path 1 Background Knowledge (BK) effects. In the Coherence Formation model of comprehension proposed here, Background Knowledge is thought to affect comprehension both directly and indirectly through its effects on attention to components of multimedia displays, i.e. text and diagrams. Path 1a from BK to comprehension will be discussed first, followed by Path 1b from BK to attention to text, and Path 1c from BK to attention to diagrams.

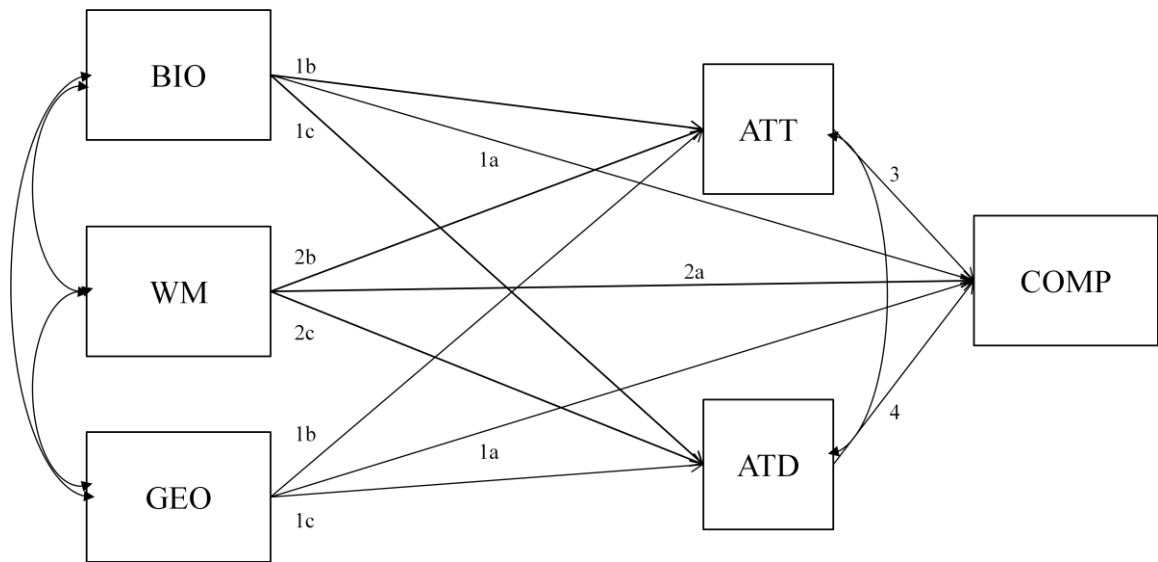


Figure 4 – The Coherence Formation Model of Illustrated Text Comprehension

Path 1a: Background Knowledge on Comprehension. Background knowledge has been studied in the context of comprehension of coherent and incoherent text. McNamara, Kintsch, Songer and Kintsch (1996) showed that high background knowledge resulted in better learning from incoherent as opposed to coherent text. Low knowledge learners have considerable trouble with incoherent text, as they lack the necessary components to infer the macro structure of the text (Kintsch, 1995). Conversely, high background knowledge allows for the “filling in” of missing causal statements which help construct a stable representation.

Overall, research has shown a facilitative effect of background knowledge on comprehension of illustrated science text (Munzer, Seufert and Brunken, 2009). The inclusion of diagrams improves the quality of mental models, but high knowledge participants construct more complete and accurate mental models compared to those with low background knowledge (Butcher, 2006). The type of representation added has been shown to interact with background knowledge; animations tend to benefit low BK participants over high BK participants (Ollerenshaw, Aidman & Kidd, 1997; Boucheix & Guignard, 2005). However, interactions in the opposite direction have been reported for animated diagrams (Kriz & Hegarty, 2007) and graphs (Kaluga, 2007).

Path 1b Background Knowledge on Attention to Text. There is precious little research in the area of background knowledge effects on attention to components of multimedia displays. Schwonke, Berthold and Renkl (2009) report that attention to textual components of multimedia displays is positively related to comprehension for those with high and low levels of background knowledge. Vauras, Hyona and Niemi (1992) examined eye movements while reading coherent vs. incoherent text and found

that the latter attracted the most visual attention. In addition they found that difficulty comprehending the incoherent text slowed down the reading process. While the study did not examine levels of background knowledge, the poor comprehension scores with incoherent text suggest prior knowledge was low.

Verbal protocols and participant constructed diagrams have shown differences in reported components of the workings of complex systems between experts and novices. Experts (High BK) tend to report and produce more integrated representations, often incorporating structural, functional and behavioral elements of the system, while novices (Low BK) tended to report and draw perceptually available static components of the system (Hmelo-Silver and Pfeffer, 2004). Direct support for attention to varying portions has been shown with eye tracking, which directly measures visual attention; higher background knowledge resulted in more time spent on relevant vs. irrelevant portions of the display (Canham & Hegarty, 2010).

The small number of studies reviewed here highlights the lack of research being conducted in this area. The differential findings from reported studies highlight the importance of understanding why and how background knowledge affects attention to components of multimedia displays.

Path 1c: Background Knowledge on Attention to Diagrams. Pictures are not perfect representations, nor are they easy to understand (Benson, 1995), and it is often unclear what portion of the visual representation is to be examined in the absence of relevant background knowledge. Adding captions to photographs can increase the amount of attention on relevant portions of the diagram (Pozzer-Argheidi, 2004). Background knowledge functions much in the same way by guiding the learner to the

appropriate portions of the display. Low background knowledge participants select salient features as opposed to thematically relevant materials (Lowe, 1999), but can benefit from additional time on graphical overviews of the materials (Salmeron, Baccino, Canas, Madrid and Fajardo, 2009). There is evidence that learners use prior knowledge to select relevant information, with those higher in background knowledge selecting more relevant information from the provided graphics. (Cook, Krajick & Vardas, 2006; Jarodzka, Scheiter, Gerjets and van Gog, 2010).

The selection of relevant information can greatly increase the chance for a correct inference. Grassear, Lu, Olde, Cooper-Pye & Whitten (2005) presented participants with diagrams of mechanical systems (lock, dishwasher). The diagrams included text explaining the basics of how the system functioned. They induced cognitive disequilibrium by the introduction of a fail statement such as “the key turns, but the bolt does not move”. Those participants higher in technical knowledge fixated the appropriate “fault” regions at above chance levels (fault region: area of mechanical breakdown that could cause the fail statement to be true). The number, percentage, and total time of fixations on the fault regions were positively correlated with comprehension scores and verbal think aloud protocols indicated more inferences were generated by those high in technical knowledge. Schwonke, Berthold and Renkl (2009) reported similar results with multiple representational displays. Despite decreased attention on the diagrams as compared to low prior knowledge participants, attention to the tree diagrams was related to higher outcome scores for high background knowledge participants. This is consistent with the hypothesis that only relevant areas were inspected by high background knowledge participants.

Path 2: Working Memory effects on Comprehension. Similar to Background knowledge, working memory is hypothesized to have a direct effect on comprehension of science text (Path 2a) and direct effects on attention to text (Path 2b) and diagrams (Path 2C).

Path 2a: Working Memory on Comprehension. Generally, the role of working memory during reading comprehension can be thought of as the selection of information for processing and formation of storage cues for retrieval in LTM. The construction integration model of comprehension outlined in previous sections, assumes a large role for working memory in the comprehension of text. Ericsson and Kintsch (1995) argue that the storage components of the memory system are much too limited to carry out complex tasks such as problem solving and reading. They introduce a Long Term Working memory component (LT-WM), which is a storage space for retrieval cues for information in long term memory. The cues stored in this space are more stable than information in STM and thus do not require immediate sustained attention.

Support for LT-WM comes from studies of interrupted reading, where participants reading a passage are interrupted for a span of time, sometimes with intervening tasks. If stable retrieval cues were not available, comprehension from the point of interruption would be impaired. Studies have shown that comprehension is not impaired after interruption, or interruption with additional task (Fischer & Glanzner, 1986; Glanzner et al., 1981). The only effect of the interruption was an increase in reading time for the first sentence after interruption of about 450 ms, which was similar to the time course of retrieval from LTM (approximately 400 ms; Ericsson & Kintsch, 1995).

In text only comprehension, it may be assumed that verbal working memory is predominantly responsible for success. However, it has been shown that, in multimedia comprehension, both verbal and visual-spatial working memory systems are needed for constructing a stable situation model (Glyselink, Jamet, & Dubois, 2009; Kintsch, 1995; Daneman & Carpenter, 1980). The research on multimedia comprehension is discussed in the next two sections. First the role of working memory on the attention spent on text is discussed, followed by a discussion of the role of working memory in attention to visual materials.

Path 2b: Working memory on Attention to Text. Differences in working memory capacity have been shown to be related to differences in complex processing tasks such as problem solving (Kane et al., 2004) and reading comprehension (Daneman & Carpenter, 1980). Eye tracking has enabled researchers to tap online attention processes to help understand these differences. Kaakinen, Hyona, and Keenan (2003) studied perspective effects on reading times for high and low WMC individuals (perspective effect: memory for a text is specific to the perspective of the reader). The perspective effect was greater for low WMC individuals than for high WMC individuals thus the introduction of a perspective results in the concentration of attention resources to relevant materials only, for those with lower WMC.

In general, those with lower WMC have difficulty selecting relevant materials when cues are not provided. In one of the few studies of multi-media comprehension in middle school students, Hannus and Hyona (1999) examined the differences in attention to illustrated science texts. There were no group differences in the time spent inspecting illustrations. High WMC students trended for longer reading times and more time spent

in relevant portions of the text. In another study, think aloud protocols were collected while participants read perspective texts. Question asking in the think aloud data corresponded with longer first pass fixations on the sentence (Kaakinen & Hyona, 2005). The authors conclude that deeper processing, as reflected by questioning and inference, was coupled with increased time spent processing the sentence.

Path 2c: Working Memory on Attention to Diagrams. Overall, low WMC individuals, like those with lower background knowledge, have trouble attending to the relevant and most informative parts of the diagram. Working memory studies of attention to diagrams have shown similar patterns of eye movements as with text. Hannus & Hyona (1999) studied eye movements of middle school students reading textbook materials and found no overall difference in viewing times on illustrations between those of high and low WMC. However, the distribution of fixations on relevant and irrelevant parts of the display differed; high ability participants spent more time on the relevant portions of the images while low WMC participants spent more time fixating irrelevant areas (blank white spaces between diagrams).

Hegarty (1992) used a sentence verification task to examine how people inspect static diagrams of mechanical systems (rope and pulley). Learners spent the most time looking at the referent of the sentence, followed by those nodes preceding it in the causal chain. This pattern indicates that learners infer the motion of the pulley system from a series of inferences made on the causal chain of the apparatus.

The seductive details effect of multimedia learning states that students are distracted by photographs and other visual representations, and that these distracting representations impair comprehension of the text presented. This effect is greater for

those of low WM capacity (Sanchez and Wiley, 2006), and is consistent with the executive control theories of working memory (Kane et al., 2004). Given the prevalence of these types of representations in standard high school textbooks (Pozzer & Roth, 2003) the opportunity of distraction is high and is magnified for those of lower working memory capacity.

Path 3: Attention to Text on Comprehension. Reading is a complex skill that is influenced by a variety of factors such as working memory, word fluency and background knowledge and thus the relationship between attention devoted to the textual elements of a multimedia display and comprehension is a complicated one. Each of these factors influences the amount of time it takes to encode and process the text under consideration. Textual factors also influence the time spent on the text such as syntactic and conceptual complexity (Rayner, 1998), and coherence (Kintsch, 1995).

Despite these complexities, studies have found positive effects of attention to textual elements on comprehension of learning materials (Hannus & Hyona, 1999; Kintsch, 1995; Pozzer-Ardenghi, 2004; Pozzer & Roth, 2003). In a study of perspective effects on reading comprehension, deeper processing, as evidenced by questioning in think aloud protocols, was coupled with increased time spent reading the sentence (Kaakinen & Hyona; 2005). Finally, in a study of the effects of background knowledge on MERs instructional materials, participants spent more time on text than on the two other representations (tree diagram and equation) and time spent on the text was positively related to learning outcomes (Schwonke et al., 2009). Some studies of attention to different parts of multimedia displays have not found this positive relationship between time on text and learning success. Jarodzoka et al. (2010) found

that despite differences in the attention devoted to textual elements between high and low WMC participants, this attention was not related to learning success.

Individual differences in attention to textual elements often correspond with distributional difference in the allocation of attention to varying parts of the MERs. Briefly summarized, those with higher levels of background knowledge (Schwonke et al., 2009, Canham & Hegarty, 2010) and/or working memory (Hyona & Keenan, 2003), allocate more visual attention to relevant portions of the display. This study will examine total time on all elements and time on relevant information only to examine the effects of individual differences on measures of eye movements.

Research on eye tracking and comprehension has been mixed. Often times the eye movement record suggests something (e.g. detection of error) but it does not get reported by the individual. Results such as these call into question the interpretation of the eye movements and their meaning. The two paths from time spent on text and time spent on diagrams to comprehension will hopefully help to elucidate some of the conflicting evidence regarding the interpretation of eye movements in terms of working memory and background knowledge.

Path 4: Attention to Diagrams on Comprehension. Despite the relatively small body of research on the attention given to diagrams in MER's, some interesting results have emerged. Generally speaking, fixations on visual representations elicit longer fixation durations than does reading text (Rayner, Portello, Stewart, Keir and Duffy, 2001). However, the visual attention to diagrams can be different based on individual cognitive characteristics such as background knowledge and working memory. Schwonke et al., (2009) found that attention to diagrams was positively related to

learning outcomes, but only for students with high prior knowledge scores. Graesser, Lu, Olde, Cooper-Pye and Whitten (2005) found that background knowledge influenced the number of fixations on the “fault” regions (areas of potential breakdown). Yet regardless of levels of background knowledge, all three measures of attention to “fault” regions (number of fixations, percentage of fixations and total time) were significantly related to device comprehension scores.

Signaling on Attention to Text and Diagrams. Learners can have trouble selecting relevant information for processing when presented with multimedia learning materials. Several studies have examined the efficacy of adding signals to the display in directing the attention of the learner to the relevant stimulus materials in animated displays. The addition of color coding or highlighting of naming labels (Ozcelik, Asla-Ari & Cagiltay, 2010), and of color coded arrows (Boucheix and Lowe, 2010) increased visual attention to relevant parts of the display.

With respect to static displays, the research is even more limited. Bartholome and Bromme (2009) studied the effects of two types of signaling manipulations, numerical labels vs. hypertext, on the identification and classification of plants. The learning environment was computerized and consisted of a diagram of a plant with accompanying text. The text contained either numbered cues (number in a circle) or was highlighted and underlined to signify a hyperlink. In the numbered condition, the text number corresponded to a number on the portion of the diagram described by the text. In the hyperlink version, clicking the hyperlink highlighted the corresponding portion of the diagram in yellow. The numbered version resulted in better learning for classification (the transfer task), than did the hyperlink condition. It was argued that the hyperlink

condition produced shallower processing by removing the need to visually segment the diagram. An alternative hypothesis is that the hyperlink condition, by virtue of the segmentation of the relevant portions, reducing global processing of the diagram, which may have impaired the performance on the classification task.

We expect that the effects of the signaling manipulation will manifest in an increase of attention to the signaled portions of the text and corresponding signaled portions of the diagrams. Readers of illustrated text rely heavily on the text to guide their learning, attending to text first (Hegarty & Just, 1993) and spending more time on it as compared to the diagram (Schmidt-Weigand et al. 2010). Signaling learners to focus on portions of the diagram should result in more looks on the diagram compared to no-cueing. The effects on the time spent looking at the diagram may vary according to which manipulation was received. The hyperlink condition may result in fixations on the diagram, but at lesser durations than the numbered condition.

The literature reviewed in this chapter led to the formation and testing of the CF model with Cohort I data from the same school as the data used for this study. The model proved to be a good fit to the data. The question of how the model would fit with the inclusion of a signaling manipulation prompted the present study. The following chapters first describe the procedures and materials used in the data collection phase of the study, followed by the methods of data screening used to determine the variables for use in the fitted model. Chapter V details the results of the exploratory analyses, group differences and finally the testing of the CF model.

CHAPTER 3 - METHODS

Participants.

Data for the study were collected from an urban school in Philadelphia, Pennsylvania during two waves of data collection within a larger study on Diagrammatic Reasoning. The school itself is a K-12 quasi-public school for high achieving students from single parent households in the tri-state area. This study focuses on the entire population of 10th grade biology students for two cohorts of students (Cohort II data collected spring 2010, Cohort III data collected fall 2010). Both cohorts of students were composed of sixty-three participants, taken from 4 intact biology classrooms, one of which was an honors class. The mean age was between 15-16 years of age, (Cohort II = 15.5, Cohort III = 15.5). The distribution of gender is 55-60% female. The distribution of race is 85-88% African-American, with the remainder of the students divided among White, Asian, Hispanic, and Other/mixed race. Socioeconomic status was estimated from education levels of the custodial parent, and is considered relatively low. For example, 53% of mothers had graduated from high school or less and 85% of fathers had graduated from high school or less (averaged across both years for which data were collected).

The academic demographics are similar to state and city averages for percentage of advanced students and percentage of basic level students; higher for percentage of proficient level students, and lower for percentage of below basic students. The 8th grade state mandated high-stakes test scores for Cohort II and III students were distributed as follows: 36% at the Advanced level (vs. a mean of 46% in the state and 21% in the city), 50% at the Proficient level (vs. a mean of 29% in the state and 28% in the city), 12% at

the Basic level (vs. a mean of 13% in the state and 21% in the city), and 2% at the Below Basic level (vs. a mean of 12% in the state and 30% in the city). Thus, the sample was selected from a socio-economically disadvantaged sample that is relatively high-achieving compared to the city and medium-achieving compared to the state.

Research Design.

The data set used for this study was collected in the context of a larger study examining the efficacy of a diagrammatic reasoning intervention aimed at improving students' ability to use diagrams effectively. The intervention consisted of a series of workbook pages, created from the students' current textbook, which taught conventions of diagrams (use of color, use of symbols etc), and either self-explanation (explanation of the text to a "confused" workbook character, Cohort II) or student-constructed diagrams (students completing partial diagrams, Cohort III) depending on the experimental group.

To examine possible process changes for Cohort III, think aloud data were collected during the eye tracking measure for both pre and post testing. The think aloud data will not be analyzed for this study. Within the larger study, participants are assigned to a diagrammatic reasoning intervention by their inclusion in a particular class. In addition to the intervention, a signaling manipulation was included in the pre/post comprehension measure for Cohort III. Participants were randomly assigned to a prompting condition as they entered the pre-test area, independently of their intervention condition.

Materials and Measures.

The data for this study were obtained with a combination of researcher-developed and existing measures. Pre-tests were given in regular classrooms during a single biology

class meeting in lieu of standardized test preparation. All measures given in the pre-test are described here. Those included in the analysis for this paper are described first and in greater detail, additional measures are described briefly.

Spatial working memory. For assessment of spatial working memory we used the automated version of the Symmetry Span task (Unsworth et al, 2005). The task belongs to a class of test known as “operation span tasks” consisting of a processing component and a storage task preformed concurrently. The automated version provides three practice blocks, one for each part of the task separately (processing and storage) and one to practice performing the tasks concurrently. For the symmetry judgment portion (processing), participants were presented with a matrix array and were instructed to decide whether the array is symmetrical along the vertical axis. Reaction time for each trial was recorded and an average time for solution of symmetry judgment was computed, and stored for use as the presentation speed of the actual test trials.

In the memory portion of the practice participants are presented with a 4x4 matrix of squares, one of which is colored red for 800ms. The numbers of matrices presented were between 4 and 12, each with one red square indicated. Participants are asked to remember the positions of the red squares in the order they were presented for later recall. Recall of positions was assessed at the end of the set of matrices by a blank 4x4 matrix; participants were required to click on the positions colored red in the order they were presented. Accuracy scores were presented after each trial in the form of “2 out of 5” for practice trials only.

The combined task interleaved the two tasks previously described such that the symmetry matrix was presented for processing for the average time for solution on the

practice version, followed immediately by the answer screen. After answering the symmetry judgment, a memory matrix with one red colored square was presented for 800ms. The sequence was between 4 and 12 interleaved trials with recall assessed at the end of the set. Recall of positions was assessed as before, but at the end of the sequence. If participants' response time on the processing task exceeded 2 standard deviations from their average response time on the practice portion, the trial was terminated and scored as an error. This is programmed into the task and was done to prevent verbal rehearsal during the processing task. In addition, participants were instructed to keep their accuracy for the processing task at 85% or above to prevent participants from "ignoring" the processing task in favor of the memory task. The task takes approximately 20 minutes to complete (see Figure 5 for example sequence)

Prior knowledge. To assess prior knowledge in the topic area of biology, a 25-item, 8-minute researcher developed assessment was used. The measure used plain language, and test students' knowledge of concepts of biology that are linked to the diagrams in the biology diagrammatic reasoning measure. The measure has been shown to have good reliability (Cronbach's Alpha = .83).

To assess prior knowledge in the topic area of geoscience, a 10-item, 4-minute researcher developed assessment was used. The measure used plain language, and test students' knowledge of concepts of geoscience that are linked to the diagrams in the geoscience diagrammatic reasoning transfer measure. The measure has been shown to have good reliability (Cronbach's Alpha = .87).

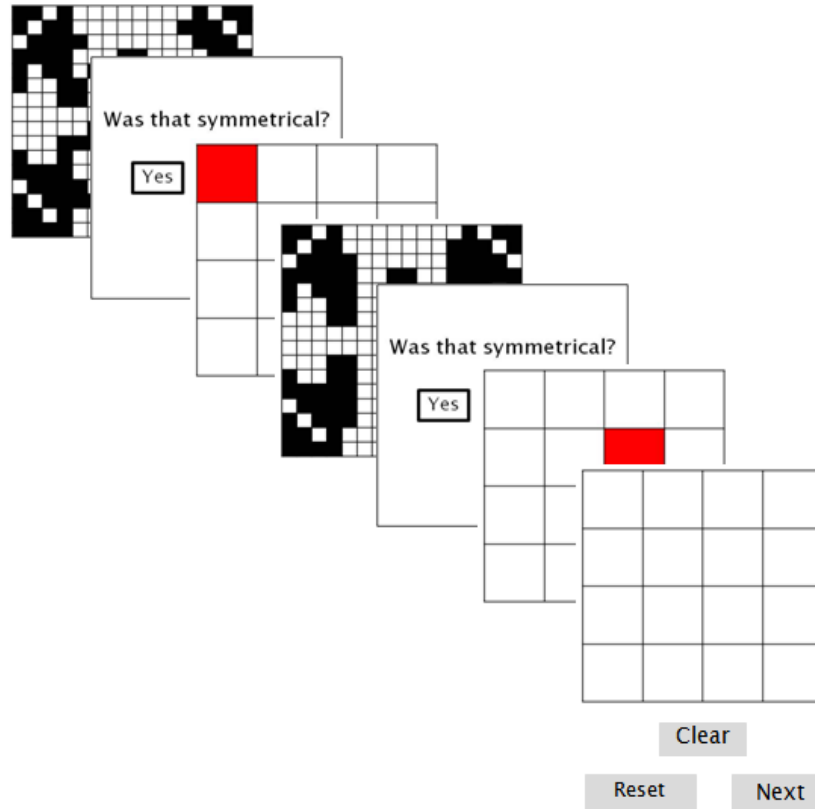


Figure 5 – Schematic of the Automated Working Memory Test (Unsworth et al., 2005)

Comprehension Measure. To assess student comprehension of the reading materials, an experimenter developed test of diagram and text comprehension was used. The test had been used in previous studies (Fitzhugh et al., 2010) and is comprised of scanned pages of the students' biology textbook which contain text and at least 1 diagram. Students were asked to read the page as if reading for homework, then answer 3 questions; 1) text based question (can be answered from the information found in the text), 2) Diagram based question (can be answered from the information found in the diagram), 3) Integration question (requires the integration of textual and diagram information to answer correctly). Participants were instructed to respond to questions verbally. All participant responses were recorded on a digital voice recorder then

transcribed and coded for accuracy offsite (see coding section for details of coding procedure).

Eye tracking. Eye movement measures were recorded during the completion of the comprehension measure described above using the Tobii T60 remote eye tracker (described in detail in the next section). The eye tracking data were analyzed for the following first order independent variables: number of fixations, fixation duration, cumulative fixation duration, reaction time and transitions between types of AOI's.

Think Aloud Protocols. Participants in Cohort III were also asked to think aloud while they learned from the materials presented to them. The instructions were given at the start of the comprehension measure as part of the regular instructions for the experiment. Participants were told that their “inner thoughts” while reading were important for learning and that the research team were interested in these thoughts. The instructions were repeated as verbal prompts by the experimenter when a period of silence of 5 seconds was encountered. Think aloud protocols have been shown to slow the pace of reading (Rayner, 1988), thus the control group may have different reaction time and reading time as compared to the two signaled groups.

Diagrammatic reasoning. Biology Diagrammatic Reasoning is a 25-item, 20-minute researcher-developed measure of diagrammatic reasoning. The measure consisted of diagrams scanned from high school biology textbooks. Diagrams were chosen by the researchers based on several criteria; a) maintaining the distribution of diagram component type (i.e. line drawing, photograph, etc.) found in textbook; b) the convention of diagrams to which the diagram applied; c) the content area from which the diagram was taken. Questions were devised to assess the level of ability to reason with diagrams

from verbatim reporting of facts (at the lower levels) to inference generation (at the higher levels). The measure has shown good internal consistency, reliability and concurrent validity (internal consistency reliability = .877 correlation with background knowledge $r(289) = .62$ with undergraduate students; Cromley, Snyder, Luciw, & Tanaka, 2007)

Geoscience Diagrammatic Reasoning measure consisted of diagrams scanned from high school biology textbooks. Diagrams were chosen by the researchers based on several criteria; a) maintaining the distribution of diagram component type (i.e. line drawing, photograph, etc.) found in textbook; b) the convention of diagrams to which the diagram applied; c) the content area from which the diagram was taken. Questions were devised to assess the level of ability to reason with diagrams from verbatim reporting of facts (at the lower levels) to inference generation (at the higher levels). The measure has shown good internal consistency, reliability and concurrent validity (internal consistency reliability = .80, correlation with background knowledge $r(289) = .57$ with undergraduate students; Cromley, Snyder, Luciw, & Tanaka, 2007)

Spatial ability. The Hidden Figures Test and the Mental Rotations Test (Form A, Peters, 1995) were administered as covariates for the larger study but scores are not analyzed in this study. Please see (Cromley et al, year) for details.

Equipment. Paper and pencil measures were presented in 1" three-ring binders with tabs separating the measures. The diagrammatic reasoning measures (biology and geoscience) were presented in color with the remaining measures presented in black and white. All responses are recorded on Scantron answer sheets and scored electronically,

with the exception of the spatial tests, for which answers were recorded directly on the sheet¹.

All computer based measures were presented on a standard 17" laptop running E-Prime 2.0 software (Psychology Software Tools, 2009). The laptop display was split and was displayed to participants on a Tobii T60 remote eye tracker. The Tobii T60 has a refresh rate of 60 Hz and a 17" monitor. During eye tracking (only during comprehension measure) the Tobii T60 samples the pupil and head position of the participants at a rate of 60 Hz with an error of .5 degrees of visual angle. Head movement is tolerated in the range of 44 x 22 x 30 cm.

Procedure.

Paper and pencil measures Paper and pencil measures were administered in intact classes during a 48 minute class period of biology instruction. Participants were given a booklet of test materials and a Scantron sheet pre-numbered with a subject number. Students were asked to record their names on a separate paper attached to the Scantron. After data entry, participants' names (i.e. attached papers) were removed and stored separate from the data. Students were then given verbal instructions for each test and a time limit for completion. Tests were given in the following order: 1) Biology Background knowledge (8 min), 2) geoscience background knowledge (5 min), 3) Biology Diagrammatic Reasoning (18 min), 4) Geoscience Diagrammatic Reasoning (6 min) 5) Embedded Figures Test (12 min), 6) Mental Rotation Test Form –A (3 min).

¹ Previous research by the Principal Investigator of the larger study had shown scores to drop significantly when students had to transfer answers on the spatial tests to Scantron sheets. Thus participants wrote directly on the forms for these tests and they were entered onto Scantron sheets for scoring by the research team.

Computer based measures. Computer based measures were administered individually in a separate room in the school building. Participants were asked to volunteer in the beginning of each biology class as permitted by instructional demands (i.e. not during testing or lab activities). Participants were led to the testing room and seated in front of the computer screen. Participants were then presented with a letter of informed assent and asked to read and sign the letter to indicate they would like to continue.

The reading comprehension measure was administered individually by computer in a dedicated room in the high school. All participants were presented with scanned pages of their current biology textbooks selected from later chapters of the book which had not been covered at the time of pre-testing (See Appendix F). The prompting intervention was included in the comprehension measure for Cohort III only. Participants from Cohort III were additionally presented with either numbered text prompts which correspond to numbered diagram parts (numbered condition, see Appendix G), or they were presented with hyperlinked text portions which highlighted the corresponding diagram portion when clicked (hyperlink condition, see Appendix H). The number and hyperlinks occurred at the same point in the text in both conditions. All text materials were identical with the exception of the prompting condition, thus Cohort II will serve as the control condition for examination of signaling effects. Participants answered comprehension questions verbally and these answers were recorded on digital recorders with a small personal microphone. Eye movements were recorded as they read and answered questions.

The working memory assessment was administered individually on computer using a program developed by Unsworth and colleagues (Automated Symmetry Span; Unsworth, Heitz, Schrock, and Engle, 2005). Participants were given verbal instructions and a verbal explanation of the working memory measure. After an opportunity for questions, the working memory measure was administered, and instructions were reissued via the E – Prime experimental program before practice blocks began. The experimenter stayed seated at a separate table in the room to monitor performance and answer any questions relating to navigation through the programs.

Upon completion of the working memory assessment the participant was given a verbal description of the comprehension measure. The participant was then asked to sit approximately 60 cm from the screen and the Tobii T60 was adjusted for calibration using the guidelines in the Tobii T60 manual (version 3, 2009). The calibration procedure is automated and uses 9 points of calibration. The participant was recalibrated if any of the 9 calibration points was missed or the error fell outside the calibration point radius.

Once successfully calibrated, the participant was fitted with a personal microphone clipped to the collar of the shirt. To test the microphone and identify their transcript, the participant was asked to say “This is participant XXX” and speak their participant number out loud into the microphone, which activated a voice based tape recorder. Once the audio test was completed, participants began the comprehension/eye tracking measure. Participants were then presented with written instruction on how to bring up the questions (press 1, 2 or 3 on the keyboard for example) and advancing the page (click the left mouse button). They were then presented with a practice trial to

familiarize them with the keystroke requirements and verbal answering procedure. The experimenter remained seated at a separate table in the room to answer any procedural questions but instructed participants that they would not answer any content related questions. The entire computer based measure lasted approximately 45 minutes.

Scoring and Coding

The data set used in the present study is comprised of pre-test data from a larger study examining the efficacy of an intervention designed to instruct high school students in diagram comprehension. All behavioral measures were collected prior to the intervention as was eye movement data from Cohort III. Eye Movement data for Cohort II was collected concurrent to the implementation of the intervention due to scheduling constraints at the school where the sample was taken. Thus, days of exposure to the intervention was recorded as the number of calendar days the intervention had been implemented at the time of eye tracking data collection. While workbooks were not done every day, they were done several times a week and thus this is a good approximation of how much exposure was received. This variable was then entered into a regression for each variable affected (e.g. all eye tracking measures and the comprehension measure), and unstandardized residuals were saved. All analysis of variance and analysis of covariance tests were performed on these residuals.

The data used for this study is a subset of the data collected for the larger study. The scores for Diagrammatic Reasoning, Paper Folding, Mental Rotations Test and Embedded Figures Test were not examined and thus are not discussed here (for a discussion of these variables Cromley et al, 2010).

Comprehension Measure. The comprehension questions are answered verbally onto audio tapes, which are then transcribed by the experimenter. After transcription, the data is coded using a rubric developed by the research team. The rubric assigns a score for each response based on the correctness and completeness of the answer (i.e. 0 = incorrect, 1 = Partial credit, 2 = correct). Partial credit is assigned to incomplete responses such as only answering 1 part of a 2 part question, or correct for one portion but not the other. No response due to unintelligible or absent answer was scored as a 0 and coded as a missing value. Coding was completed independently by 3 members of the research team. All disputes regarding coding were resolved with a group discussion. Inter-rater reliability for the coding scheme was 88% (see Appendix I for coding rubric).

In previous research using this method, there has been a trend for students to answer questions requiring inference with “I don’t know” and thus be scored incorrect. Most of the questions requiring inference are text and diagram integration questions, and thus were more difficult. The tendency to not attempt and answer resulted in very little variability for the scores of that question type. Thus we have found that adjusting scores for the difficulty level by weighting them by the total possible points produces a better distribution of scores. All scores were weighted to correct for the tendency to not attempt of the question using the following formula:

$$WS = s * \sqrt{\frac{1}{s/2}}$$

where WS = the weighted score, s = the score for that item, and 2 is the total score possible for each item. The formula weights the correct score of questions such that more

credit is given to correctly answered difficult questions than is given to correctly answered, verbatim questions. Weighted scores were then entered into data analysis.

Intraclass correlation index was computed for the comprehension measure. The index is a composite measure of both inter-rater (systematic) and intra-rater (non-systematic) reliability. Typically correlations greater than .70 are considered acceptable. The Cronbach's alpha level across all trials and questions types was within acceptable parameters ($\alpha = .79$).

Background Knowledge. Scoring for the biology and geoscience background knowledge measures were scored in the same manner and so will be discussed together. The tests are multiple choice tests each with one discrete correct answer choice out of four possibilities (A, B, C, and D). Answers are recorded on Scantron sheets and processed through Temple University's Measurement and Research Center. Data are converted to electronic format in the form of a text document, which is imported into SPSS. Answer choices are reported and assigned a 1 if correct and a 0 for incorrect. Correct answers are summed for each subject and percentages are computed by dividing the subject's total correct by the total possible (24 points for Biology and 10 points for Geoscience).

Visuo-spatial Working Memory. The automated program produces two scores. The liberal scoring method awards one point for each matrix position recalled in its correct serial position. Thus a participant can earn a score of 2 by recalling 2 of the 4 matrix positions in the correct serial location. The second or conservative score awards a point only if all locations are recalled in the correct serial position. Thus in the previous example, although 2 of the 4 items were correctly recalled, the trial would be scored as

incorrect because not all locations were recalled correctly. The liberal scoring method is often used for examining errors such as serial position, primacy and recency effects. We will use the conservative score as it has been shown to be more discriminating and a better predictor of general intelligence measures (Unsworth, 2007a)

Reaction Time. Reaction Time, the time from the start of the trial until the subject pressed the key to terminate the trial, was calculated automatically by the computer program running the comprehension measure.

Fixation Duration. The average fixation duration was calculated by averaging the duration of each valid fixation, across trials and across AOI types.

Number of Fixations. The total number of fixations was calculated by summing the number of valid fixations for each trial.

Proportion of Fixations on Relevant AOI's. Valid fixations were coded for whether the fixation was on a portion of the display relevant to the question. Relevance was determined a priori. The total number of fixations on relevant areas was then summed and divided by the total number of fixations made.

Switching Across AOI Type. A switch was identified as a fixation on one AOI type followed by a fixation on a different AOI type. Data were coded for 10 different types of switches to account for switches between all pairs of AOI's in the display. Number of across AOI type switches was then summed across trials.

Switching Within AOI Type. A within AOI type switch was identified as a fixation on one AOI type followed by a fixation on the same AOI type. For example, a fixation on the diagram followed by a fixation on another portion of the diagram, without

moving out of the AOI was considered a switch within AOI type. Within type switching was coded for each of the four AOI types. These data were then summed across trials.

Proportion Data. In addition to the previous first order statistics, a series of second order statistics was calculated for each first order variable. Overall Comprehension score and Reaction Time were partitioned into subscales by the type of question (i.e. text, diagram and integration). Fixation Duration, Number of Fixations, Proportion of Fixations on Relevant AOI's, Switches Across AOI Type and Switches Within AOI Type were divided into proportions on each type of AOI (i.e. text, diagram, signal, and question). These data thus explore the distribution of eye movements across different types of AOI's to examine possible differences in the distributions between groups. These data also correct for differences in absolute time and number due to extraneous variables (i.e. strategies or reading speed), thus making group comparisons possible.

CHAPTER 4 - RESULTS

Descriptives

Means and standard deviations were calculated for all variables in raw form (see Table 1). All dependent and independent variables were screened for univariate normality and homogeneity of variances. The eye data for Cohort II was collected while the intervention was implemented due to scheduling constraints of the school. Thus, Days Exposure to the intervention was recorded for each participant at the time of data collection. This variable was removed from all eye data using regression. Cohort III students' data was collected prior to the start of the intervention, so the Days Exposure for this group was evaluated at a value of zero. Unstandardized regression residuals were saved as variables and analysis proceeded on these values. Descriptives for these data are presented in Table 2.

First Order Variables

Behavioral Measures. Means and standard deviations were calculated for all behavioral measures. All behavioral measure scores were normally distributed (Comprehension, Biology Background, Geology Background, and Working Memory). The Geoscience Background scores indicated a possible violation of homogeneity of variance. A variance ratio test on the ratio of the highest variance group to the lowest variance group was below the cutoff and thus within acceptable limits. No transformation of the variable was performed.

Eye Movement Data; Reaction Time (RT): The unstandardized residuals of reaction time with days exposure removed exhibited a positive skew, which is common in reaction time data. Thus a square-root transformation was applied to the raw RT for each

trial, and the unstandardized residuals computed from the transformed variable. The resulting variable was normally distributed and homogeneity of variance between groups was established.

Number of Fixations. Means and standard deviations were calculated for average number of fixations across trials (see Table 1). There were no violations of normality and the homogeneity of variance assumption was satisfied.

Fixation Duration. Means and standard deviations are presented in Table 1. There were no violations of normality for the unstandardized residuals of average fixation duration on the display. The variance was homogeneous between groups. Thus no transformations were required.

Proportion of Fixations on Relevant AOI's. Means and standard deviations are presented in Table 1. No violations of univariate normality were found and variance was equivalent across groups.

Across AOI Type Switches. Means and standard deviations are presented in Table 2. No violations of univariate normality were found. A significant Levene statistic indicated a possible violation of the homogeneity of variance assumption. Thus a variance ratio test was conducted on SWTYPE and the ratio of the largest variance group to the smallest variance group was found to be within acceptable limits (<5). Thus no transformations were made on this variable.

Table 1 – Descriptives for Raw Variables

<i>Variable</i>	<i>Mean (SD)</i>	<i>Skew</i>	<i>Kurt</i>	<i>Mean (SD)</i>	<i>Skew</i>	<i>Kurt</i>	<i>Mean (SD)</i>	<i>Skew</i>	<i>Kurt</i>
	<i>Control Group N=63</i>			<i>Numbered Signals N=33</i>			<i>Hyperlink Signals N=30</i>		
<i>Working Memory</i>	27.00(7.85)	0.42	0.34	27.66(6.21)	0.22	0.89	27.13(6.95)	0.49	0.01
<i>Biology Background Knowledge</i>	5.41(1.55)	0.05	0.03	6.57(2.86)	0.72	0.59	6.34(2.99)	0.13	0.44
<i>Geoscience Background Knowledge</i>	7.25(2.04)	0.27	-0.52	4.39(1.71)	-0.06	-0.19	3.58(1.76)	0.89	0.43
<i>Comprehension</i>	9.79(4.22)	0.30	-0.08	9.78(3.27)	-0.23	-0.49	8.86(2.87)	0.15	-0.59
<i>Reaction Time</i>	3.56(1.63)	-0.91	0.67	5.32(1.38)	0.68	0.08	5.39(1.38)	0.69	-0.50
<i>Duration</i>	1.62(0.58)	1.45	3.39	1.55(0.39)	0.51	-0.39	1.81(1.33)	2.43	5.88
<i>Fixations</i>	98.37(36.64)	0.59	0.17	141.33(65.5)	-0.15	0.21	147.15(65.5)	-0.41	-0.8
<i>Switches Across Type</i>	36.53(14.96)	0.299	-0.26	69.45(25.22)	0.63	1.19	72.93(30.01)	-0.29	-0.68
<i>Switches Within Type</i>	61.70(25.87)	1.07	1.67	71.88(25.50)	-0.17	-0.22	74.22(39.48)	0.00	-0.80

Table 1 – continued

Variable	Mean (SD)	Skew	Kurt	Mean (SD)	Skew	Kurt	Mean (SD)	Skew	Kurt
	Control Group N=63			Numbered Signals N=33			Hyperlink Signals N=30		
<i>Relevant Fixations</i>	0.65(0.16)	-0.76	1.35	0.56(0.09)	-0.62	-0.03	0.51(0.08)	0.13	1.15
<i>Comprehension Score (Text)</i>	1.62(0.74)	-1.01	0.17	1.43(.49)	-0.49	-0.85	1.19(0.40)	0.26	-0.53
<i>Comprehension Score (Diagram)</i>	1.51(0.73)	-0.66	-0.33	1.01(0.52)	0.41	-0.63	1.09(0.55)	-0.09	-0.39
<i>Comprehension Score Integration</i>	1.46(0.93)	-0.04	-1.05	1.04(0.51)	-0.16	-0.37	0.84(0.58)	0.22	-0.63
<i>Reaction Time Text</i>	2.40(1.15)	0.16	0.21	0.79(0.34)	0.61	0.21	0.68(0.40)	1.41	1.88
<i>Reaction Time Diagram</i>	2.42(1.19)	0.08	0.24	0.83(0.37)	1.28	3.00	0.65(0.41)	1.59	3.66
<i>Fixations on Text</i>	0.35(0.15)	0.44	-0.07	0.37(0.12)	0.36	-0.53	0.27(0.08)	0.63	1.17
<i>Fixations on Diagrams</i>	0.32(0.14)	-0.07	-0.44	0.39(0.13)	-0.16	0.48	0.32(0.10)	-0.41	-0.10
<i>Fixations on Signals</i>	--	--	--	0.02(0.03)	1.55	1.80	0.28(0.12)	-0.25	-1.12
<i>Fixations on Questions</i>	0.28(0.12)	0.98	2.40	0.20(0.06)	0.53	0.41	0.35(0.07)	1.14	2.19

Table 1 – continued

Variable	Control Group N=63			Numbered Signals N=33			Hyperlink Signals N=30		
	Mean (SD)	Skew	Kurt	Mean (SD)	Skew	Kurt	Mean (SD)	Skew	Kurt
<i>Duration on Text</i>	0.37(0.16)	-0.16	-0.58	0.45(0.14)	-0.41	0.23	0.35(0.16)	0.02	-0.39
<i>Duration on Diagrams</i>	0.26(0.12)	0.05	-0.63	0.26(0.12)	0.78	1.10	0.24(0.08)	0.11	-1.03
<i>Duration on Signals</i>	--	--	--	0.01(0.01)	4.41	22.13	0.01(0.01)	0.75	0.21
<i>Duration on Questions</i>	0.32(0.13)	0.17	-0.05	0.28(0.09)	-0.12	-0.70	0.39(0.14)	0.61	-0.44
<i>Switches Between Text and Diagrams</i>	0.26(0.12)	0.73	1.29	0.34(0.11)	-0.02	-0.63	0.22(0.08)	0.45	0.54
<i>Switches Between Text and Signals</i>	--	--	--	0.06(0.07)	1.37	0.82	0.10(0.07)	-0.23	-1.44
<i>Switches Between Text and Questions</i>	0.24(0.15)	1.07	2.17	0.25(0.10)	0.63	0.22	0.25(0.10)	1.16	2.09
<i>Switches Between Diagrams and Signals</i>	--	--	--	0.01(0.02)	1.37	0.52	0.02(0.02)	0.57	-0.73
<i>Switches Between Diagrams and Questions</i>	0.32(0.17)	0.58	0.08	0.22(0.11)	-0.05	-0.90	0.28(0.10)	0.62	0.70
<i>Switches Between Signals and Questions</i>	--	--	--	0.01(0.02)	3.14	9.32	0.02(0.02)	1.25	1.31

Table 1 – continued

Variable	Control Group N=63			Numbered Signals N=33			Hyperlink Signals N=30		
	Mean (SD)	Skew	Kurt	Mean (SD)	Skew	Kurt	Mean (SD)	Skew	Kurt
<i>Switches Within the Text</i>	0.38(0.19)	0.33	-0.54	0.37(0.19)	0.33	-0.64	0.22(0.13)	0.90	1.34
<i>Switches Within the Diagram</i>	0.32(0.18)	0.12	-0.59	0.47(0.20)	-0.27	-0.21	0.36(0.17)	-0.45	-0.19
<i>Proportion of Relevant Text Fixations</i>	0.22(0.14)	0.54	-0.52	0.17(0.07)	0.27	-0.82	0.15(0.08)	0.16	0.09
<i>Proportion of Relevant Diagram Fixations</i>	0.18(0.11)	0.74	0.12	0.18(0.09)	0.21	-0.34	0.15(0.07)	-0.24	-0.08
<i>Proportion of Relevant Signals Fixated</i>	--	--	--	0.01(0.02)	1.87	3.79	0.02(0.02)	0.42	-0.94
<i>Proportion of Relevant Question Fixations</i>	0.25(0.12)	0.87	3.22	0.20(0.06)	0.65	0.46	0.19(0.08)	0.71	-0.29

Table 2 – Descriptives of Unstandardized Residuals

Variable	Mean (SD)	Skew	Kurt	Mean (SD)	Skew	Kurt	Mean (SD)	Skew	Kurt
	Control Group N=63			Numbered Signals N=33			Hyperlink Signals N=30		
<i>Comprehension Score</i>	0.11(0.46)	-0.53	0.50	-0.05(0.39)	-0.72	0.89	-0.15(0.37)	-0.39	0.83
<i>Reaction Time</i>	-0.39(1.74)	-0.87	0.41	0.32(1.37)	0.73	0.11	0.40(1.32)	0.74	-0.41
<i>Duration</i>	0.01(0.58)	1.44	3.29	-0.12(0.39)	0.51	-0.39	0.13(1.33)	2.43	5.88
<i>Fixations</i>	10.17(39.99)	0.65	0.12	6.54(45.56)	-0.15	0.21	12.35(65.53)	-0.41	-0.84
<i>Switches Across Type</i>	-7.58(19.13)	0.00	0.02	5.26(25.22)	0.63	1.19	8.73(30.01)	-0.30	-0.68
<i>Switches Within Type</i>	-2.66(26.20)	1.14	1.70	1.34(25.50)	-0.17	-0.22	3.68(39.48)	0.00	-0.80
<i>Number of Relevant Fixations</i>	0.02(0.16)	-0.92	2.11	0.00(0.09)	-0.62	-0.03	-0.05(0.08)	0.13	1.15
<i>Comprehension Text Questions</i>	0.07(0.59)	-1.04	0.38	0.04(0.55)	-0.61	-0.37	-0.18(0.45)	-0.25	0.36
<i>Comprehension on Diagram Questions</i>	0.11(0.62)	-0.48	0.53	-0.13(0.54)	0.26	-0.60	-0.06(0.58)	-0.12	-0.49
<i>Comprehension on Intergration Question</i>	0.14(0.81)	-0.23	-0.39	-0.06(0.55)	-0.19	-0.59	-0.21(0.59)	0.25	-0.68

Table 2 – continued

<i>Variable</i>	<i>Mean (SD)</i>	<i>Skew</i>	<i>Kurt</i>	<i>Mean (SD)</i>	<i>Skew</i>	<i>Kurt</i>	<i>Mean (SD)</i>	<i>Skew</i>	<i>Kurt</i>
	<i>Control Group N=63</i>			<i>Numbered Signals N=33</i>			<i>Hyperlink Signals N=30</i>		
<i>Reaction Time Text</i>	0.30(1.21)	0.23	-0.04	-0.21(0.34)	0.66	0.26	-0.33(0.40)	1.40	1.88
<i>Reaction Time Diagram</i>	0.31(1.31)	-0.18	-0.50	-0.20(0.38)	1.23	2.80	-0.40(0.42)	1.49	3.39
<i>Reaction Time Integration</i>	0.16(0.99)	-0.51	-0.49	-0.54(0.54)	1.57	3.27	0.14(1.15)	1.40	3.31
<i>Fixations on Text</i>	-0.01(0.15)	0.36	-0.27	0.05(0.12)	0.36	-0.53	-0.05(0.08)	-0.63	1.17
<i>Fixation on Diagram</i>	-0.01(0.14)	-0.08	-0.49	0.04(0.13)	-0.16	0.48	-0.03(0.10)	-0.41	-0.10
<i>Fixations on Signals</i>	---	---	---	0.00(0.03)	1.55	1.80	0.02(0.03)	-0.25	-1.12
<i>Fixations on Questions</i>	0.01(0.12)	1.03	2.76	-0.08(0.06)	0.53	0.41	0.07(0.07)	1.14	2.19
<i>Duration on Text</i>	-0.01(0.16)	-0.15	-0.63	0.05(0.14)	-0.41	0.23	-0.05(0.16)	0.02	-0.39
<i>Duration on Diagrams</i>	0.00(0.12)	0.06	-0.63	0.01(0.12)	0.78	1.10	-0.01(0.08)	0.11	-1.03
<i>Duration on Signals</i>	---	---	---	0.00(0.01)	4.41	22.13	0.01(0.01)	0.75	0.21

Table 2 – continued

Variable	Mean (SD)	Skew	Kurt	Mean (SD)	Skew	Kurt	Mean (SD)	Skew	Kurt
	Control Group N=63			Numbered Signals N=33			Hyperlink Signals N=30		
<i>Duration on Questions</i>	0.01(0.13)	0.32	0.12	-0.05(0.09)	-0.12	-0.70	0.06(0.14)	0.61	-0.44
<i>Switches Between Text and Diagrams</i>	-0.01(0.12)	0.73	1.28	0.07(0.11)	-0.02	-0.63	-0.05(0.08)	0.45	0.54
<i>Switches Between Text and Signals</i>	---	---	---	0.00(0.07)	1.37	0.82	0.04(0.07)	-0.23	-1.44
<i>Switches Between Text and Questions</i>	-0.01(0.15)	1.06	2.06	0.01(0.10)	0.63	0.22	0.01(0.10)	1.16	2.09
<i>Switches Between Diagrams and Signals</i>	---	---	---	0.00(0.02)	1.37	0.53	0.01(0.02)	0.57	-0.73
<i>Switches Between Diagrams and Questions</i>	0.03(0.17)	0.60	0.08	-0.06(0.11)	-0.05	-0.90	0.00(0.10)	0.62	0.70
<i>Switches Between Signals and Questions</i>	---	---	---	0.00(0.02)	3.14	9.32	0.01(0.02)	1.25	1.31

Within AOI Type Switching. Means and standard deviations are presented in Table 1. No violations of univariate normality were found. A significant Levene statistic indicated the possibility of a violation of homogeneity of variance. A variance ratio test was conducted and the ratio of the largest variance group to the smallest variance group was found to be within acceptable limits (<5). Thus no transformations were made on this variable.

Second Order Data.

Comprehension Scores. Scores on the comprehension measure were divided into scores for each type of question presented. The resulting three variables (COMPTxt, COMPDia, COMPInt) were screened for normality. Severe violations of kurtosis were found for the COMPTxt and COMPDia variables. Examination of the distribution revealed a ceiling effect for both types of questions. It was possible to correctly answer these two types of questions with very little or no inference. Therefore all that was required was for the student to locate the appropriate text phrase or diagram portion. The COMPInt was normally distributed and exhibited no violations of homogeneity of variance. The main focus of this analysis is to investigate the efficacy of students' comprehension of science text. As such the integration question requires the most comprehension as it cannot be answered verbatim from the text or diagram. Thus the comprehension data for the text and diagram questions were removed from analysis. All further analysis included COMPInt only.

Proportions of Reaction Time. Average reaction time across the trial was parsed into average reaction time for each type of question (RTTtxt, RTDia, and RTInt). Similar to the RT data, the variables exhibited a positive skew, but also showed violations of

kurtosis. A square-root transformation of the raw data, with days exposure then removed, corrected the distributions for RTTtxt and RTDia, but did not correct the distribution violation for RTInt. As the comprehension data for text and diagram questions had to be excluded due to violations of normality, the corresponding RT data also must be discarded. RTInt was also excluded from further analysis due to the level of kurtosis in the data.

The non-normal distributions for this variable may be accounted for by strategic differences in the way the participants approached the comprehension task. Many student press the question buttons to bring up the questions prior to reading. This strategy is one suggested by their instructor and results in very short RT for these participants. However, reading strategies are not a focus of this paper and thus will not be examined here.

Proportion of Fixations. The total number of fixations made on the display was decomposed into proportions of fixations on each of the AOI types (FIXT, FIXD, FIXS, and FIXQ). There were no violations of assumptions for FIXT, FIXD and FIXQ. The proportion of fixations on signal exhibited severe kurtosis. Examination of the histograms showed a floor effect due to the fact that the hyperlink group was the only group for whom this behavior could be quantified. The control group did not have any external signals, but did include a figure reference in each paragraph, and thus hits on the figure reference were coded as signal hits. The numbered group either did not directly fixate the numbers, or they gazed briefly in the course of reading, examining the signal through non foveated vision. Due to these conditions, the signaling data was excluded from all further analysis.

Proportions of Fixation Duration. The total duration on the display was parsed into the proportion of that total duration on each of the components of the display. Average proportion of duration on the text (DURT) and diagram (DURD) exhibited no violations of normality. Proportion of fixations on signal exhibited severe kurtosis and skew. Examination of the graphs indicated a lack of fixations captured for the control group and the numbered group, with very low incidence of behavior for the hyperlink group. Thus, the variable was excluded from further analysis. Proportion of fixations on questions (DURQ) was normally distributed and the homogeneity of variance assumption was satisfied. No transformations were performed on these variables.

Proportion of Fixations on Relevant AOI's. The number fixations on relevant parts of the display were parsed into the proportion of relevant fixations on each AOI type (except signal data which was removed from analysis). There were no violations of normality and all variances were homogeneous between groups. No transformations were done on these variables.

Proportion of Switches Across AOI type. The number of total across AOI type switches was coded for switches between each of the four AOI's, resulting in 12 patterns. Since direction of the switch is not a focus of this study, these were further aggregated into 6 switch patterns; switches between text and diagram (SWTD), switches between text and signal (SWTS), switches between text and question (SWTQ), switches between diagram and signal (SWDS), switches between diagram and question (SWDQ), and switches between signal and question (SWSQ). Similar to previous signal data, the proportion of switches between signals and other components was close to zero. Thus these variables were excluded from further analysis. The remaining three patterns

showed no violations of normality. Switches between diagram and text had a significant Levene statistic. A variance ratio test of the largest group variance to the lowest group variance was within acceptable limits. Thus no transformations were performed on the variables.

Proportion of Within AOI Switches. The total number of within AOI switches was parsed into the proportion of within switches made for each of the AOI types; switches within the text (WSWT), switches within the diagram (WSWD), switches within signals (WSWS), and switches within the question (WSWQ). Signal data was removed from further analysis due to low occurrence of fixations, resulting in three variables. There were no violations of normality; however proportion of switches within text and proportion of switches within diagram both had a significant Levene statistic. Variance ratio tests performed on both variables showed the ratio of the largest group variance to the smallest group variance was within acceptable limits. No transformations were made on the variables.

In summary, 38 variables were screened for univariate normality. Data that exhibited uncorrectable violations from normality were dropped from further analysis. Thus 13 variables were removed from analysis. Due to the non normality of the individual question data, the total comprehension scores was removed in favor of the comprehension score for integration questions only. The remaining analysis will proceed with the remaining 25 variables.

Screening for multivariate assumptions.

Scatter plots of each bivariate combination of variables were evaluated for linearity and homoscedasticity. Only the second order proportion data exhibited any

violations of these assumptions. Proportions of eye tracking behavior related to attention on textual components were positively related to switching behavior containing textual components (both across and within). Proportion of attention to textual elements was negatively related to the attention to diagram components and switching behavior related to attention to diagrams (both across and within). Homoscedasticity was present in one of the bivariate combinations of variables; Proportions of Duration on diagram and proportion of relevant fixations on diagrams. These relationships will be taken into consideration in the factor analysis.

Correlations for each variable are listed in Appendix J. These correlations were examined for possible multicollinearity problems with the data. Variables with high correlations ($\geq .800$) were flagged as possible problems and were considered for aggregation into factors during factor analysis.

Missing data. In the S.E.M framework the pattern of missing data in the multivariate sample must be evaluated to determine if it is systematic or not. However in this sample of 126, only 4 participants had missing data and the reason for missingness is known. In all cases data were missing due to an inability to track participants' eye movements. The reason for inability to track a participant is related to the physical structure of the eye or the presence of glasses. Neither of these factors is related in any way to their performance on the measures reported here, thus missingness was inferred to be missing at random.

Results

Results are discussed in sections grouped by the type of analysis examined. First discussed are the ANOVA's testing the difference in covariates for each of the signaling

groups. The purpose of these analyses are to establish the equality of groups on covariate measures prior to testing other hypotheses. Second, a Cluster Analysis to determine if groups exist based on WM and BK exist in the data set. The ANCOVA's will be discussed third as they examine the efficacy of the signaling conditions on the comprehension of the text presented and any possible interactions with covariates. Fourth, the regression analyses are discussed as they relate to the selection of variables for inclusion in the path model. Fifth, factor analysis for eye movement data is discussed. The purposes of these analyses are to examine the factor structure of the eye movement data for inclusion in the path model. Finally the path model results are presented.

Analysis of Variance.

Differences in Covariates between groups. Working Memory, Biology

Background and Geoscience Background scores were entered individually in three Group (3) by Score (1) ANOVA. There were no significant differences in working memory scores for the three groups. The omnibus test was significant for both Biology Background ($F(2,122) = 3.13, p = .05$) and Geoscience Background ($F(2,122) = 47.45, p < .001$). Planned contrasts for Biology Background found a significant difference between control group and numbered group ($t(122) = 2.03, p = .02$) and a non-significant trend for control group and hyperlink group ($p = .09$) in the same direction as the previous contrast. Essentially Biology Background scores were higher for the signaled groups (Cohort III) than for the control group (Cohort II). Planned contrasts for the Geoscience Background measure found the opposite trend in the data. Control group scored higher than both numbered group ($t(122) = 7.04, p < .001$) and the hyperlink group ($t(122) = 8.71, p < .001$). Given the statistically significant differences in Background knowledge

score between samples, these variables were examined in a cluster analysis to determine if participants exhibited distinct groups based on their scores. This analysis is discussed in the following section.

Cluster Analysis.

Given the statistically significant difference in groups on the Background Knowledge measures a cluster analysis was performed to investigate if groups of participants clustered around these scores. A hierarchical cluster analysis was performed on Working Memory, Biology Background Knowledge and Geoscience Background Knowledge using Ward's estimation of Squared Euclidean distance. This method finds the most similar pair of clusters p and q ($p > q$), denoting this similarity (s_{pq}). The algorithm then reduces the number of clusters by one through the merger of clusters p and q , labeling the new cluster t ($=q$) and updates similarity matrix to reflect revised similarities or dissimilarities between cluster t and all other clusters. Iterations are performed until all entities are in one cluster. Evaluation of the percentage of change when cluster are combined indicates the number of clusters present in the data. For example if the largest percentage difference is from cluster 1 to cluster 2, there is one cluster in the data.

Two clusters were identified using this method. As working memory was not significantly differently between groups, the cluster membership was primarily determined through the background knowledge scores. Examination of variables based on cluster membership indicated the two groups were defined by high/low background knowledge for both Biology Background Knowledge and Geoscience Background Knowledge. Cluster membership was distributed evenly across conditions. These results

indicate that participants did not differ within subjects on their scores for the background knowledge measures and serve as an indication that working memory and background knowledge may provide different loadings for different groups of participants.

Analysis of Covariance.

Analysis of covariance was run on each of the variables selected for inclusion in the path model to test for differences in these variables due to the inclusion or absence of signals in the display. The covariates for each analysis were Working Memory, Biology Background, and Geoscience Background. Planned contrasts to test for difference between groups were run regardless of the significance of the omnibus test and no corrections for multiple comparisons were made. Non-significant trends (between .05 and .10) in the expected directions are reported and discussed in terms of the bearing the trends may have on the path model.

Comprehension. Scores on the integration question were examined for differences between groups and interactions with covariates. There were no significant interactions with Working Memory, Biology Background Knowledge or Geoscience Background knowledge. There were no significant differences between conditions. All groups scored similarly on this measure. Thus the signaling manipulation did not affect the overall comprehension of the text as measured by the questions asked.

Number of Fixations. Number of fixations made on the display was examined for differences between groups and interactions with covariates. No significant effects of Working Memory, Biology Background Knowledge, or Geoscience Background Knowledge were found. In addition no effects of condition were found. Participants made similar numbers of fixations on the display regardless of the presence or absence of

signals in the display. Planned contrasts also showed no differences between type of signal present in the display.

Proportions of Fixations. The proportion of fixations allocated to each AOI type in the display was examined for differences between groups and interactions with covariates. There were no significant effects of covariates for Proportions of fixations on text, diagram or questions. Significant effects of condition were found for Proportion of fixations on text ($F(2,114) = 5.03, p = .008, \eta^2 = .09$) and Proportion of Fixations on questions ($F(2,114) = 17.42, p < .001, \eta^2 = .25$). There was also a non-significant trend for proportion of fixations on diagrams ($p = .08, \eta^2 = .05$).

Planned contrasts revealed significant difference between groups on the Proportion of Fixation on Text; both the control ($p = .05$), and the numbered group ($p = .002$) made a greater proportion of fixations on the text than did the hyperlink group. Proportions of Fixations on Diagrams showed non-significant trends; the numbered group tended to make a larger proportion of fixations on the diagram than either the control group ($p = .08$) or hyperlink group ($p = .06$). These trends may reflect in group separation in the path model. Finally the proportion of fixation on questions was significantly different in the groups. Control participants made a larger proportion of fixations on the question than did numbered participants ($p = .006$). The hyperlink group made the largest proportion of fixations on the questions, exceeding both the control group ($p = .03$) and the numbered group ($p < .001$). This indicates that the hyperlink group re-read the question more times than did the other two groups. Since there were no difference in the comprehension measure, it is difficult to know why this behavior occurred. It could be that the hyperlink group had more trouble with the questions, or

that they were more aware of their lack of understanding. Examination of the think aloud protocols collected during the comprehension measure may shed light on this phenomenon, but that analysis is beyond the scope of this paper.

Fixation Duration. Average total Fixation Duration on the display was examined for differences between groups and interactions with covariates. No significant effects of Working Memory, Biology Background Knowledge, or Geoscience Background Knowledge were found. In addition no effects of condition were found. Average length of durations on the display was similar regardless of the presence or absence of signals in the display. Planned contrasts also showed no differences between type of signal present in the display.

Proportions of Duration. Proportion of total duration on text was examined for differences between groups and interactions with covariates. No significant effects of Working Memory, Biology Background Knowledge, or Geoscience Background Knowledge were found. In addition no effects of condition were found. Participants spent equitable proportions of their overall time on the text regardless of the presence or absence of signals in the display. Planned contrasts also showed no differences between type of signal present in the display.

Proportion of total duration on the diagram was examined for differences between groups and interactions with covariates. No significant effects of Working Memory, Biology Background Knowledge, or Geoscience Background Knowledge were found. In addition no effects of condition were found. Participants spent similar amounts of time on the diagram regardless of the presence or absence of signals in the display. Planned contrasts also showed no differences between type of signal present in the display.

Proportion of total Duration on the questions was examined for differences between groups and interactions with covariates. No significant effects of Working Memory or Biology Background Knowledge were found. There was a non-significant trend for the Geoscience Background Knowledge measure ($p = .09$) which may reflect the group differences in scores on this measure. A significant difference was found for Condition ($F(2,114) = 6.29, p = .003, \eta^2 = .10$) indicating an effect of signaling on the proportion of time spent on the questions. Planned contrasts indicated this difference was driven by the two signaled groups; the hyperlink group spent a greater proportion of time on the question than did the numbered group ($p = .001$). This effect mirrors the effect found for the number of fixations made on the questions.

Number of Across AOI Type Switches. The number of switches made across AOI types was examined for differences between groups and interactions with covariates. No significant effects of Working Memory, Biology Background Knowledge or Geoscience Background Knowledge were found. There was a significant effect of Condition ($F(2, 114) = 3.85, p = .02, \eta^2 = .07$) indicating the presence of signals affected the number of across type switches made. Planned contrasts indicated the type of signal had an effect on across AOI type switching behavior. Specifically, both the numbered group ($p = .02$), and the hyperlink group ($p = .009$), made more switches across type than did the control group. Thus the presence of signals resulted in a greater number of switches across AOI types. The lack of significant difference in comprehension makes it difficult to ascertain whether these switches were helpful in the integration of information.

Proportion of Across AOI Type Switches. The proportion of switches across Text and Diagram components was examined for differences between groups and interactions

with covariates. No significant effects of Working Memory, Biology Background Knowledge or Geoscience Background Knowledge were found. There was a significant effect of Condition ($F(2, 114) = 9.40, p < .001, \eta^2 = .07$) indicating the presence of signals affected the number of across type switches made between the text and the diagram. Planned contrasts indicated the presence of a signal had an effect on Text and Diagram switching behavior. Specifically, both the numbered group ($p = .04$), and the hyperlink group ($p < .001$), made more switches across type than did the control group. Thus the presence of signals resulted in a greater number of switches between text and diagram.

The Proportion of switches between text and question was examined for differences between groups and interactions with covariates. No significant effects of Working Memory, Biology Background Knowledge or Geoscience Background Knowledge were found. No significant effects of Condition were found indicating the presence of signals does not affect the proportion of switches between the text and the questions. Planned contrasts showed the type of signal did not have an effect on this proportion.

The Proportion of switches between diagram and question was examined for differences between groups and interactions with covariates. No significant effects of Working Memory, Biology Background Knowledge or Geoscience Background Knowledge were found. No significant effects of Condition were found indicating the presence of signals does not affect the proportion of switches between the text and the questions. Planned contrasts showed a non-significant trend for the hyperlink group to make a larger proportion of switches between diagram and questions than did the

numbered group ($p = .07$). This trend is in the same direction as the previous proportion data indicating the hyperlink group attended more to the question than did the other two groups.

Within AOI type Switches. Proportion within AOI switches was examined for differences between groups and interactions with covariates. No significant effects of Working Memory, Biology Background Knowledge, or Geoscience Background Knowledge were found. In addition no effects of condition were found. Participants made equitable proportions of their switches within AOI's regardless of the presence or absence of signals in the display. Planned contrasts also showed no differences between type of signal present in the display.

Proportion of Within AOI Type Switches. The proportion of switches within Text was examined for differences between groups and interactions with covariates. No significant effects of Working Memory, Biology Background Knowledge or Geoscience Background Knowledge were found. There was a significant effect of Condition ($F(2, 114) = 6.34, p = .003, \eta^2 = .11$) indicating the presence of signals affected the proportion of within AOI type switches made on the text. Planned contrasts indicated the presence of a signal had an effect on within Text switching. Specifically, the control group made more switches within text than did the hyperlink group ($p = .03$) or the numbered group ($p < .001$). Thus the presence of signals resulted in a reduced number of within text switches.

The proportion of switches within diagram was examined for differences between groups and interactions with covariates. No significant effects of Working Memory or Biology Background Knowledge were found. There was a non-significant trend for

Geoscience Background Knowledge ($p = .07$) indicating there may be some influence of GBK on integration within the diagram. There was a significant effect of Condition ($F(2, 114) = 3.94, p = .02, \eta^2 = .07$) indicating the presence of signals affected the proportion of within AOI type switches made on the text. Planned contrasts indicated the presence of a numbered signal had an effect on within diagram switching. Specifically, the numbered group made more switches within diagram than did the hyperlink group ($p = .01$) or the control group ($p < .05$). This suggests the presence of numbered signals increases the number of within diagram integration behavior.

The proportion of switches within questions was examined for differences between groups and interactions with covariates. No significant effects of Working Memory, Biology Background Knowledge or Geoscience Background Knowledge were found. There was a significant effect of Condition ($F(2, 114) = 24.05, p < .001, \eta^2 = .31$) indicating the presence of signals affected the proportion of within question switches made on the text. Planned contrasts indicated the presence of a numbered signal had an effect on within question switching. Specifically, the control group made more switches within questions than did the numbered group ($p = .001$). The hyperlink group made more within question switches than both the numbered group ($p = .001$) and the control group ($p = .01$). This suggests the presence of hyperlink signals increases the number of within question switching behavior, thought to reflect rereading of the question.

Proportion of Fixations on Relevant AOI's. Proportion of fixations on relevant AOI's was examined for differences between groups and interactions with covariates. No significant effects of Working Memory, Biology Background Knowledge, or Geoscience Background Knowledge were found. In addition no effects of condition were

found. Participants spent equitable proportions of fixations on the relevant AOI's regardless of the presence or absence of signals in the display. Planned contrasts also showed no differences between type of signal present in the display.

Proportion of Relevant Fixations on AOI by Type. Proportion of relevant fixations on textual elements was examined for differences between groups and interactions with covariates. No significant effects of Working Memory, Biology Background Knowledge, or Geoscience Background Knowledge were found. In addition no effects of condition were found. Participants spent equitable proportions of relevant fixations on the textual AOI's regardless of the presence or absence of signals in the display. Planned contrasts also showed no differences between type of signal present in the display.

Proportion of relevant fixations on diagram elements was examined for differences between groups and interactions with covariates. No significant effects of Working Memory or Biology Background Knowledge. There was a non-significant trend for Geoscience Background Knowledge ($p = .08$), indicating there may be an effect of Geoscience Background Knowledge on the selection of relevant diagram components. No effects of condition were found; participants spent equitable proportions of relevant fixations on the diagram regardless of the presence or absence of signals in the display. Planned contrasts also showed no differences between type of signal present in the display.

Proportion of relevant fixations on questions elements was examined for differences between groups and interactions with covariates. No significant effects of Working Memory, Biology Background Knowledge, or Geoscience Background

Knowledge. No effects of condition were found; participants spent equitable proportions of relevant fixations on the question regardless of the presence or absence of signals in the display. Planned contrasts also showed no differences between type of signal present in the display.

Regression Analysis

A series of regression analyses were conducted with Comprehension of Integration score as the dependent variable and each of the remaining 22 variables as predictors. The purpose of this series of regressions was to determine which of the variables are predictive of comprehension score. The majority of the variables were not predictive as individual predictors. Eighteen of the 22 variables did not exhibit significant linear relationships with the comprehension score. There are several factors that may have contributed to these results. First the relationship between the variables could be non-linear. Examination of the bivariate scatterplots for each of the variables in the analysis indicates this is not the issue.

Second, examination of bivariate scatterplots indicated there could be a collinearity problem between some of the variables. Some of the proportion data, proportion of fixations on text and proportion of duration on text for example, were positively related to one another. These same variables were negatively related to their diagram counterparts. Thus, factor analysis to combine the offending variables is conducted in the next section to determine the structure to be used in the structural equation model.

Third, the sampling distribution for covariance variables (specifically working memory) may have been more homogeneous than expected due to the nature of the

school from which the sample was taken. Previous research has shown that working memory scores fail to predict performance measures when homogenous across the sample. The school from which these data were drawn recruits high ability students from lower socio-economic background, thus the working memory scores may have been sampled from an above average population, reducing the number of scores in the lower end of the distribution (see means Table 1).

Four of the variables included in the analysis were significant predictors of comprehension scores. Geoscience Background Knowledge was positively related to comprehension scores ($F(1,119) = 4.78, p = .03, r^2 = .04$) indicating those with higher background knowledge in geoscience performed better on the integration comprehension questions.

Number of fixations was negatively related to comprehension ($F(1,119) = 1.99, p = .05, r^2 = .04$); more fixations on the display indicated a reduction in scores for the comprehension measure. This indicates that despite the lack of statistically significant difference between groups, those who made a greater number of fixations on the data did so due to difficulty in locating, recalling, or integrating information.

The number of switches across AOI type was positively related to the comprehension of integration questions ($F(1,119) = 4.18, p = .04, r^2 = .04$). This serves as preliminary evidence that switching across AOI type, or the integration of information from different parts of the multimedia display, improves the comprehension of the materials.

Finally, the proportion of switches made between diagram and the question was positively related to the comprehension of integration questions ($F(1,119) = 7.19, p =$

.008, $r^2 = .06$). The explanation for this relationship is speculative. It could be that those who switched between the question and the diagram more were motivated to answer the question correctly by repeated re-reading of the question and examination of the diagram. Another possible explanation is those who exhibit this behavior were better able to discern their level of understanding of the material than those who did not, thus enabling them to correct themselves. Further examination of the think aloud protocols may shed some light on this relationship; however that analysis is beyond the scope of this paper and are thus not considered here.

There was one non-significant trend in the data that bears consideration for the analyses to follow. There was a marginally significant effect for proportion of relevant fixations on textual elements of the display ($F(1,119) = 2.79, p = .09, r^2 = .03$). This marginal finding is supportive of the construction integration model which states that students learning from multimedia text build their situation models first from the text, especially when background knowledge is low, as it is for Biology Background Knowledge in this sample. However the significant differences in Background Knowledge scores for the Cohort III sample suggest this effect may only be present in the Cohort II student, resulting in a non-significant regression result.

Factor Analysis

Twenty two variables were entered into a principle component factor analysis. In order to minimize the number of variables with high loadings on each factor and the number of factors needed to explain the data, an Equimax rotation was chosen. Five variables were extracted based on Eigenvalues > 1 . The factor structure was difficult to interpret due to the inclusion of both first order and second order data. Thus the first

order variables were removed from the analysis (RT, FIX, DUR, SWTYPE, WSW, and REL). The remaining 16 variables were run again and 5 factors were again extracted. The inclusion of eye tracking data related to the question (FIXQ, DURQ, RELQ, and WSWQ) made the factor structure difficult to interpret as these variables loaded equally on more than one factor. Thus these data were removed from the factor analysis, leaving 12 variables for the following analysis.

Principle Components analysis on the remaining 14 variables extracted 4 factors, based on Eigenvalues > 1 . However the background knowledge measures were split across two factors and the working memory variable had low loading on 2 factors. Thus the number of factors to extract was set to four and the analysis was run again, producing more interpretable results. Factor one was consistent with attention to diagrams and had high positive loadings for FIXD (.95), WSWD (.91), DURD (.87), RELD (.78), and SWDQ (.65). Factor two was consistent with attention to diagrams and had high positive loadings for FIXT (.66), WSWT (.65), DURT (.69), RELT (.56), and SWTQ (.76). Factor three was consistent with Background Knowledge and had positive loadings for GBK (.714) and BBK (.592). The final factor consisted of VSWM only (.867). These factors are consistent with the hypothesis of this research and thus will be entered into the path model.

Exploratory analysis of the factor structure by high/low background knowledge clusters identified in the previous analysis indicated the factor structure may be different between the groups of high and low background knowledge. In addition, factor analysis on each signaling condition independently showed the factor structure to be different as well. However, contrary to the hypothesis suggested, the factor structure for the control

group was similar to the numbered group, rather than the two signaled groups having similar structures. The hyperlink group exhibited a very different structure, with some variables loading on different factors and others loading in the opposite direction. Testing these group differences within the path model is beyond the scope of this paper, but will be considered when analyzing possible modifications to the model.

Exploratory factor analysis resulted in the exclusion of 10 variables due to inconsistent loadings and redundancy with other variables in the analysis. The remaining 12 variables produced 4 extracted factors which can be interpreted as attention to text, attention to diagrams, background knowledge and working memory. Exploratory analysis of groups separately confirmed the presence of a difference in factor structure for the high/low background knowledge clusters. Contrary to the initial hypothesis, examination of factor structure for experimental groups indicated a difference in factor structure for the hyperlink group. It was initially thought that the signaled groups would have similar factors structures different from that of the control group.

Path Model

Selection of Variables. This section discusses the procedure and results for the Coherence Formation Model of Illustrated Text Comprehension (CFITC). Analysis of the data available (N=126) and the parameters outlined in the above results section indicated the inclusion of all variables would result in an over-identified model. Thus the following changes to the model were made.

Working Memory (WM). The preliminary data analysis revealed Visuo-spatial Working Memory scores had small non-significant correlations with most of the variables under examination. In addition regression analysis showed little or no

relationship to the dependent variables in the study. Thus this variable was omitted from the analysis.

Background Knowledge (BK). The previous analyses showed differential relationships between the two measured background knowledge scores and the comprehension measure. Geoscience Background Knowledge was significantly related to Comprehension while Biology Background Knowledge was not. In addition correlations for each background knowledge variable were significant with eye tracking measures. Thus, both background knowledge variables were entered as individual predictors in the model. .

Attention to Text (ATT). The exploratory factor analysis indicated 5 factors which loaded onto the ATT latent variable. Due to the cross loading with the Attention to Diagram (ATD) latent variable, switches between text and diagram was omitted from the factor (SWTD). This left four observed scores which were used to calculate a factor score for each participant. The factor score was then entered into the path model.

Attention to Diagram (ATD). The exploratory factor analysis indicated 5 factors which loaded onto the ATD latent variable. Due to the cross loading with the Attention to Text (ATT) latent variable, switches between text and diagram (SWTD) was excluded from the factor. This left four observed scores which were used to calculate a factor score for each participant. The factor score was then entered into the path model. .

Switches Between Text and Diagrams (SWTD). This variable was not a part of the initial hypothesized model. However the factor analysis indicated that it is a variable that should be considered in the model. This variable loaded well on both attention to text and attention to diagram factors, causing some difficulty in interpretation of factor

scores. Thus the score for this variable was thus entered into the model as an endogenous mediating variable between background knowledge scores and attention to components of the display.

Comprehension. The score for this variable was comprised of the sum of the scores for three integration questions, one from each of the three trials participants received to form a composite comprehension variable. This variable was entered into the path model as an observed dependent variable.

Evaluating Model Fit. A pair of fit indices recommended by Hu and Bentler (1999) for samples of $N < 250$ (comparative fit index [CFI] $> .95$ and standardized root mean residual [SRMR] $< .09$) were used to evaluate model fit. Hu and Bentler (1999) did not recommend the use of chi-square for assessing fit as the Chi-square statistic is sensitive to many factors: It is inflated by large sample sizes, models with many variables, large correlations among variables, omission of variables, and non-normality (both skewness and kurtosis; Kenny & McCoach, 2003). This data set contains two of the causes of inflated Chi-square (large correlation among predictor or manifest variables, and nonnormality) thus the statistic was assumed to be unreliable for this data set. However, the chi-square and corresponding p-value for each model are reported for model comparisons and for completeness as is the SRMR fit index.

Series of Models.

The raw data were read into MPlus (Muthen & Muthen, 1998-2010) estimation software version 4.1. Due to the small number of participants in this study, standard errors were estimated using bootstrapping with 1000 replications of the same sample size as the observed sample. A two group model with signaled (combining numbered and

hyperlink signaled groups) and non-signaled (control) groups was tested. The first model fit (Model A) held all paths equal across groups. This model was a poor fit to the data as indicated by the fit indices (see Table 3). Modification indices indicated that the paths from Biology Background Knowledge to Switching Between Text and Diagrams, Comprehension to Attention to Text, and Comprehension on Attention to Diagrams should be allowed to vary across groups. This change was consistent with hypotheses and was thus run as Model C, which was an acceptable fit to the data with the exception of the SRMR value.

Table 3 – Table of model fit indices for series of models

Model	χ^2	<i>df</i>	χ^2 <i>p value</i>	$\Delta\chi^2$	Δdf	<i>p-value of difference test</i>	RMSEA	CFI	SRMR
A	37.718	15	0.005	--	--	--	0.137	0.86	0.140
B	19.387	12	0.07	18.331	3	< .001	0.099	0.943	0.111
C	12.205	11	0.34	7.182	1	< .05	0.042	0.991	0.096
D	10.264	10	0.41	1.941	1	ns	0.021	0.998	0.087

Examination of the path loadings in conjunction with the hypothesized effects indicated modifications to Model C. The following paths were freed to vary across groups to vary across groups; Biology Background Knowledge to Attention to Diagrams, Geoscience Background Knowledge on Attention to Diagrams, and Comprehension on Biology Background Knowledge. In addition, two paths were added to the model based on hypothesized relationships between eye tracking variables. Paths from Switching Between Text and Diagrams to Attention to Text and Attention to Diagrams were added

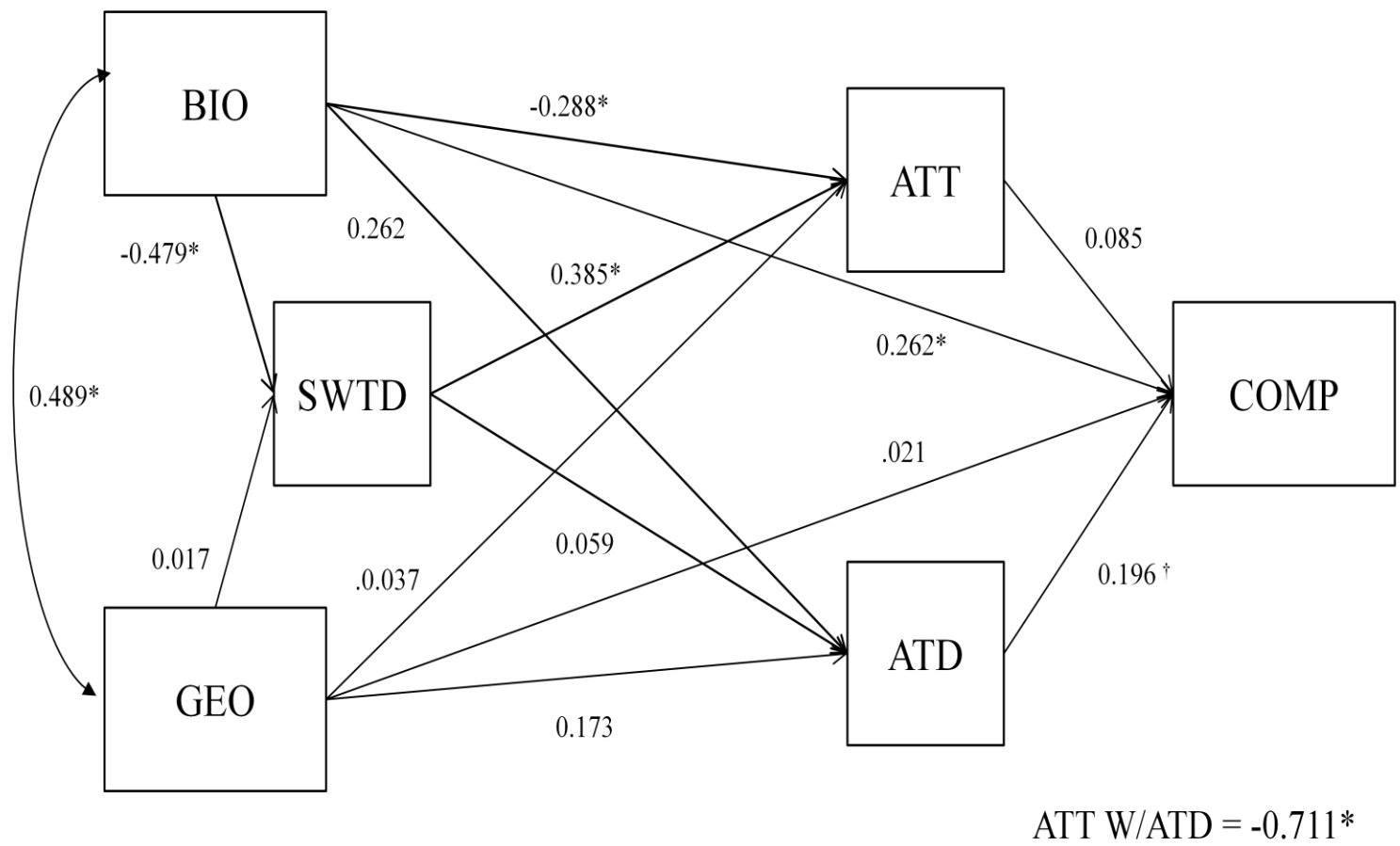


Figure 6 – Fitted Path Diagram with Standardized Estimates for Control Group

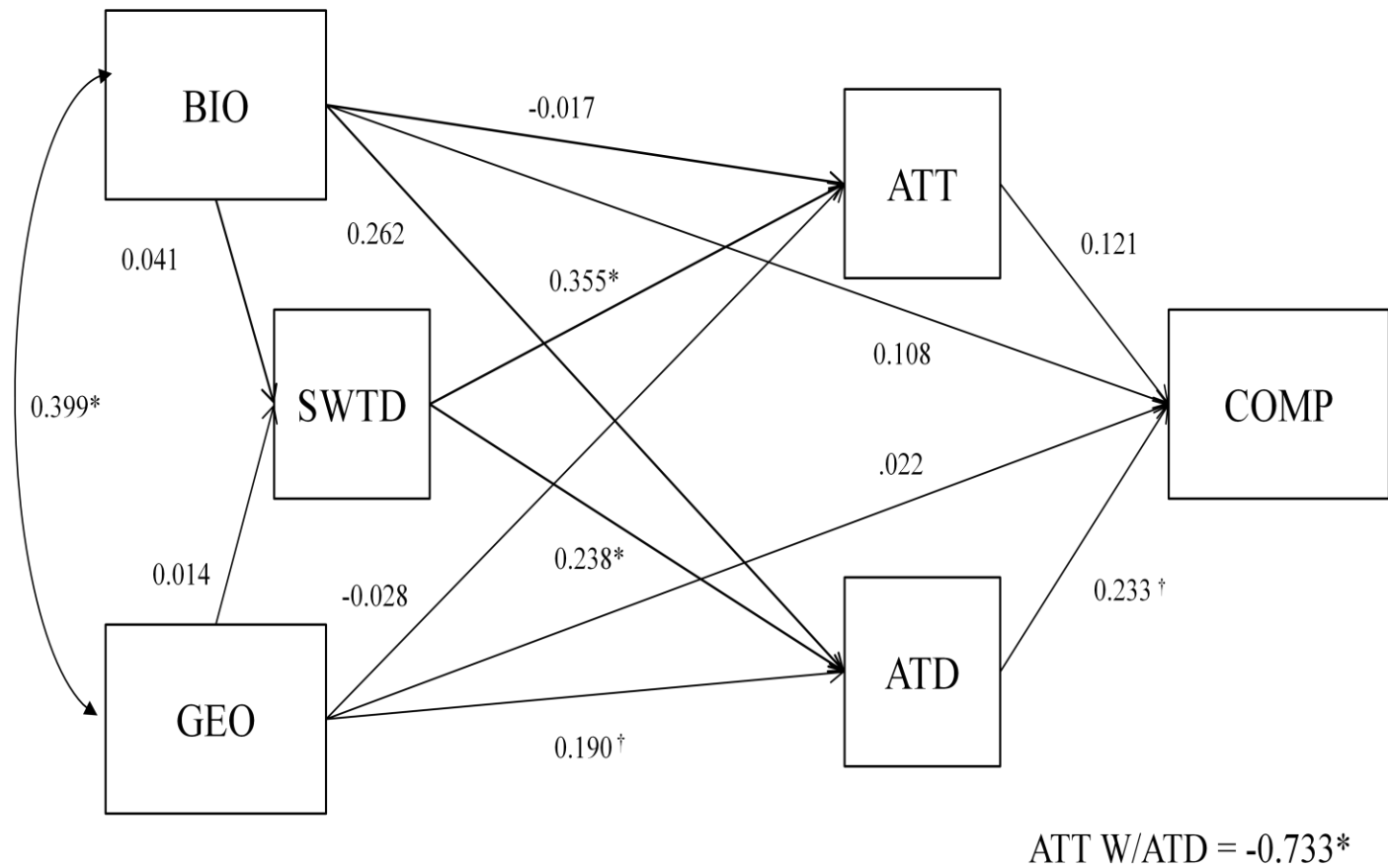


Figure 7 – Fitted Path Diagram with Standardized Estimates for Signaled Group

to the model. This model was run as model D. Model D had a similar fit to the data as Model C, but resulted in slightly better fit indices and better path estimates, resulting in the models shown in Figures 6 (Control) and Figure 7 (Signaled).

Model results for the control group showed several relationships of interest. Biology Background Knowledge had a significant positive path loading on Comprehension indicating that increased Biology Background Knowledge increased Comprehension scores for integration questions. This variable also predicted Attention to Text and Switches Between Text and Diagrams, but the relationship was negative. This indicates that increased Biology Background Knowledge resulted in less attention to the text and less switching between the text and diagram elements of the display. Switching Between Text and Diagrams had a significant positive path loading to Attention to Text which suggests that while switches were made for this group, the majority of the visual attention remained on the textual elements of the display. Finally, Attention to Diagrams had a marginally significant path loading to Comprehension which gives some preliminary support to the hypothesis that visual attention to diagrams improves Comprehension of questions designed to integrate information from both text and diagram elements.

Model results for the signaled group were different from that of the control group. Biology Background Knowledge did not have significant path loadings on Attention to Text, Switching Between Text and Diagrams or Comprehension. Thus Biology Background Knowledge was not a significant determinant of attention to the displays as

for the control group, nor was it predictive of Comprehension directly. Geoscience Background Knowledge had a significant direct path on comprehension and a marginally significant path loading to Attention to Diagrams. Like the control group, Attention to Diagrams had a marginally significant path loading on Comprehension.

Finally for the signaled group, Switching Between Text and Diagrams had positive path loadings on both Attention to Text and Attention to Diagrams. This indicates that the signaled group allocated visual attention to both the textual and diagram elements of the display. Paired with previously discussed increases in diagram attention for the signaled groups, this indicates that the presence of signals increased attention to the diagram.

While there were differences in many of the paths between the model for the control and signaled groups, the most significant in terms of multimedia comprehension is the difference in the effect of background knowledge on attention to parts of the display and comprehension directly. It would appear that in the absence of signals, Biology Background Knowledge is the primary element in both comprehension and in guiding attention to relevant display components. When signals are present the variable Switching Between Text and Diagrams, guided in part by the signaling manipulation, takes on the role of guiding visual attention. The lack of a significant path from Biology Background Knowledge to Comprehension is evidence that the presence of signals alters this relationship. For the signaled group, the relationships with respect to Biology Background Knowledge were instead found for Geoscience Background knowledge. This indicates that the successful use of signals may rely on a different type of

background knowledge. This relationship is elaborated further in the following discussion.

Table 4 – Direct and Indirect Effects of Variables on Comprehension

<i>CONTROL</i>	<i>Total</i>	<i>Direct</i>	<i>SE</i>	<i>Indirect</i>	<i>SE</i>
<i>BK on Comp</i>	0.267	0.262	0.161	0.005	0.07
<i>GEO on Comp</i>	0.052	0.021	0.109	0.031	0.036
<i>ATT on Comp</i>	0.085	0.085	0.224	--	--
<i>ATD on Comp</i>	0.196	0.196	0.214	--	--
<i>SIGNALED</i>	<i>Total</i>	<i>Direct</i>	<i>SE</i>	<i>Indirect</i>	
<i>BK on Comp</i>	0.091	0.108	0.067	-0.017	0.021
<i>GEO on Comp</i>	0.061	0.022	0.109	0.039	0.034
<i>ATT on Comp</i>	0.121	0.121	0.224	--	--
<i>ATD on Comp</i>	0.233	0.233	0.214	--	--

CHAPTER 5 – DISCUSSION

Research Question 1

Are there “groups” of learners with different profiles of Working Memory and Background Knowledge? How many of these groups are there? Does the classification of these groups capture the interactions between working memory and background knowledge? To examine this question a hierarchical cluster analysis was performed on the sample. Two groups were identified which did not differ on working memory scores, but differed on background knowledge measures. Specifically the two groups corresponded to high and low background knowledge. The number of participants in each group was slightly in favor of those with low background knowledge (N=73) with fewer participants clustering in the high background knowledge category (N=51). Examination of box plots for both types of background knowledge indicated no interactions on background knowledge scores, that is to say those high on one type were also high on the other.

There was no apparent interaction between working memory and background knowledge in this sample. This may have occurred for a variety of reasons. First, this sample was taken from a school which selects students for high intellectual ability from lower-socioeconomic backgrounds. Examination of the mean scores on the VSWM measure confirm, this selection procedure resulted in a higher proportion of above average students in the school. Working memory scores are less predictive of achievement in homogeneous samples (Unsworth, 2007a).

Second, there were statistically significant differences in the levels of background knowledge between the two samples of participants. In the sample Cohort II, geoscience background scores were higher than for the sample Cohort III, which exhibited higher overall biology knowledge scores. These differences may have influenced the cluster analysis. However this explanation seems unlikely as the distribution of participants was not cohesive with respect to the experimental groups.

Finally, only visuo-spatial working memory was assessed for this study as it was thought that diagram comprehension relied more heavily on this construct. However, previous studies on this sample have not evaluated the clustering of students based on these variables and thus the possibility of the exclusion of an important clustering variable cannot be overlooked.

These results indicate that there were not profiles of learners who differed on measures of visuo-spatial working memory but rather clusters which differed only on the amount of background knowledge pertaining to the diagram measures presented as part of the larger study.

Research Question 2

Which measures of eye movements are best for examining the effects of attention to illustrated text on comprehension? In terms of predictive analysis, regressions indicated that the overall number of fixations was negatively related to the comprehension requiring the integration of information from the text and the accompanying illustration. Thus, more fixations seem to indicate a difficulty with understanding the material or locating the appropriate areas of the display. Both of these factors could result in less accurate responses to the questions. However, the lack of

statistically significant differences between the groups in the measure of comprehension indicates this is a learner characteristic and not an effect of the materials presented.

The number of switches made across the types of AOI's in the display was positively related to the comprehension measure. This indicates that eye movements between the different areas of the display facilitate the integration of the information found in the different areas. The significant differences between groups, with signaled groups exhibiting more of this behavior than the control group, show that this behavior is affected by the inclusion of signals in the materials. However, the lack of differences in the comprehension measure between groups makes it difficult to conclude that these signals helped increase comprehension.

Finally, the proportion of fixations on areas of the display relevant for answering the integration questions was positively related to the comprehension of materials. This result replicates previous work, indicating that selection of relevant materials is an integral part of comprehension as measured by questions requiring inferences. Factors thought to influence the selection of relevant areas are background knowledge and working memory. This relationship was not explicitly tested in the regression framework, but is discussed later with respect to the ANCOVA results examining differences among groups.

Research Question 3

How does signaling effect the time spent on text and time spent on diagrams? A series of ANCOVAs were run to test the effects of signaling on the visual attention behavior of students reading biology text and produced mixed results. Overall, the control groups exhibited more fixations on the textual elements of the display, including

the number of switches within the text. This indicates that without signals, the strategy of students is to focus on the textual elements of displays. This is consistent with a number of studies, including previous results from the larger study from which the data for this analysis was obtained. For example, previous research has shown that students rarely study the diagram and often ignore them altogether (Fitzhugh et al. 2010).

With respect to the diagram, the numbered signals resulted in more fixations on the diagram than both hyperlink signals and no signals. Additionally, the numbered group made more switches within the diagram. This behavior indicates that the numbered group actively integrated the components of the diagram, possibly relating them to one another. Previous research has shown that hyperlink signals can reduce scores on a classification task designed to assess the amount of transfer from instruction to new materials (Bartholome & Bromme, 2009). While the authors speculate a possibly calibration error in what was learned stemming from the ease of the hyperlinks in highlighting specific components; the data presented here points to the hyperlinks interference of global processing of the diagram as represented by lack of switching from one diagram element to another.

Finally, the hyperlink groups on average looked more at the question than either control or numbered signal groups. There are several possible explanations for this behavior. The first is that they may have not understood fully the relationship of the question to the passage they just read. This could be due to the distraction caused by the inclusion of hyperlinks in the running text and the attention drawn to the diagram as a result of clicking on the hyperlink. Perhaps the hyperlink disrupts reading and thus interferes with the comprehension of what was read. Second, it is suggested by the

increase in switching between diagram and question for participants in the hyperlink group, that this group had difficulty with the integration question due to problems interpreting the diagram. Coupled with results from the numbered group, who made significantly more within diagram switches than any other group, and less question to diagram switching, this hypothesis seems accurate. Thus hyperlink signals represent a distraction from integration and it is recommended they not be used.

Research Question 4

Does the signaling manipulation interact with any of the predictor variables?

Results of the ANCOVA showed no interactions between working memory, background knowledge and the presence or absence of signals. However, the Geoscience Background Knowledge showed a non significant trend for the proportion of duration on diagrams. While not significant, this may reflect an influence of geoscience background knowledge on the amount of time spent in the diagram. Specifically, those with higher background knowledge scores spend less time on the text and more time evaluating the diagram. This is consistent with prior research showing that those with high background knowledge may benefit from diagram only instruction (Kalyuga et al., 2004) and are better able to interpret diagrams in their area of expertise (Hmelo-Silver & Pfeffer, 2004). However it must also be mentioned that the trend may be a reflection of the difference in geoscience background knowledge between the two cohorts of students. Still, this may show that the signals compensated for a lack of sufficient background knowledge in Cohort III as scores on the comprehension measure were comparable across groups.

Research Question 5

Using a new sample, does the Coherence Formation Model of Illustrated Text Comprehension (CF Model) fit the data well? Using a sample of students, from two different cohorts from the same school resulted in good model fit for the CF model of Illustrated Text. The addition of a signaling manipulation for one of the groups resulted in a two group model with eight of the 12 paths constrained to be equal across the two groups. Thus while the groups differed slightly, the relationships between most of the variables in the model were invariant across groups. This is to be expected given the characteristics of the sample and lack of invariance would be suspect.

The lack of any significant relationship between Working Memory in this study and subsequent exclusion from the model was disappointing. The lack of significant results in this study could be due to several factors. First, strategy use has been shown to mediate the relationship between working memory and performance measures (Bacon, Handley, Dennis & Newstead, 2008). Anecdotally, students exhibited one strategy regularly which was to read the question prior to reading the entire passage. This strategy was instructed by the teacher participating in the intervention, both prior to the intervention and during. This strategy may remove the need for students to maintain information in working memory and instead rely on a search strategy for answering questions. This strategy is apparent in the think aloud protocols collected during the comprehension measure. Second, the stimulus materials, including the question, were available during the entire question answering period. This characteristic of the paradigm may have reduced or eliminated the demand for working memory resources.

Third, the sample was relatively high scoring on state mandated assessments compared to the city and state averages. Thus homogeneity of scores within the sample could have reduced the predictive power of the working memory assessment scores. The means for the experimental groups are high compared to previous samples from the same school and thus lend support to this hypothesis.

Research Question 6

Which predictors have the largest effect on attention to text and diagrams as measured with eye movements? Is the relationship the same between the groups identified? The control group showed significant path loadings for Biology Background Knowledge on Attention to Text and the proportion of Switches Between Text and Diagrams. This relationship was negative indicating that increased Biology Background Knowledge resulted in less switching, less attention to the text elements of the display and vice versa. Consistent with Kintch's theory of comprehension and Multimedia Theory, those with higher background knowledge spend less time evaluating the textual elements. Higher background knowledge enables a reader to construct an accurate representation of the text with greater efficacy than those with lower background knowledge. In addition, background knowledge directs attention to the relevant parts of the display, eliminating the need for searching and rereading behaviors. In addition, Biology Background Knowledge had a direct relationship with Comprehension scores, indicating that this knowledge may have contributed to the successful formation of a situation model as well.

Contrary to previous research, this effect was not found for Attention to Diagrams. While higher background knowledge has been shown to increase the use of

diagrams, many of these studies examine experts in the domain under investigation. However, the participants of this study cannot be classified as experts, as they are in an introductory biology course, which may account for the lack of diagram investigation. Previous research on similar participants has indicated many do not examine the diagram at all while reading, and very little when answering comprehension questions (Fitzhugh et al., 2010). In addition, the lack of diagram inspection may reflect reading strategies either taught or spontaneously adopted by the participants of the study.

For the signaled groups, the previously described relationship between Biology Background Knowledge and Attention to Text and switching behavior was not found. Thus the inclusion of signals changes this relationship. Consistent with prior research, signals enable students to circumvent the use of domain specific background knowledge (biology) or lack thereof, by directing their attention to the relevant parts of the display. The signals used in this study, specifically the numbered prompts, were designed to signal which text information is related to which diagram portions. Thus integration is facilitated by exogenous cues rather than those internal to the reader.

With the addition of signals, Geoscience Background Knowledge became a significant predictor of visual attention to the text and has a marginal effect on the attention to the diagram portions of the display. Recent research by an affiliate of the author has demonstrated that geoscience proficiency is related to the ability to use diagrams and text materials successfully (Shiple, personal communication). This is of interest for the signaling manipulation used in this study. Results of the preliminary analysis for these two groups showed differences in the allocation of attention to the diagram and text, with the numbered signals increasing both proportion of fixations on

the diagram, and the number of within diagram relationships examined. Successful use of the numbered prompts may be aided by the ability of readers to integrate diagram and textual information as reflected by higher geoscience scores. This relationship may also account for the direct relationship of Geoscience Background Knowledge to Comprehension on the signaled group not found in the control group.

Research Question 7

What is the relationship between attention to differing parts of a multimedia display and comprehension of textbook materials? Overall, attention to the text and diagram portions of the display did not directly account for the comprehension of materials as specified in the model. However in both groups, Attention to Diagrams was marginally related to such comprehension while Attention to Text was seemingly unrelated. There are several factors that could have contributed to these effects. First is the strategies adopted by participants of this study for answering comprehension questions. A large proportion of students read the question before reading the text, adopting a type of search strategy, as opposed to a reading and comprehension strategy. Second, much of the text presented to students (an entire scanned textbook page) was not related to the questions asked and thus could account for the lack of relationship.

For the signaled group, the relationship between Attention to Text increased in magnitude, while remaining non-significant. The attention to the relevant parts of the display increased for this group as a result of the inclusion of signals. This indicates that perhaps attention to relevant parts of the display is the primary predictor of comprehension of materials. In progress research by research team of the larger study, with the same population of subjects as examined here, will attempt to answer this

question by eliminating extraneous text information from the displays. Further research should take into account the amount of extraneous information when designing research stimuli.

Limitations

The main limitation of this study is the small sample size. The small sample size, while adequate for tests such as ANCOVA and Regression, made the modeling of the relationships difficult at best. The model had to be modified in several ways to acquire enough degrees of freedom for convergence. A second limitation is due to time and resource constraints, a small number of covariate variables were collected. Of special note is the lack of verbal working memory scores for this sample, the inclusion of which may have bolstered the relationship between working memory and comprehension as well as attention to text.

The availability of the text and diagram during the question answering period may have reduced or eliminated the need to rely on working memory resources in this experiment. Future research examining the effects of working memory on elements of comprehension should be wary of such issues and design paradigms which allow for the use of working memory.

Implications

The results of this study have implications for those attempting to understand how to increase student comprehension of illustrated science text, specifically in the area of biology. Most interesting is the lack of differences in comprehension scores between the experimental groups. This indicates the existence of multiple paths to the same outcome. Students in this sample either relied heavily on their Biology Background Knowledge or

their Geoscience Background knowledge for the construction of their situation models. For the control group, which received the text exactly as presented in the textbook, their Biology Background knowledge directed their attention to parts of the display and directly affected comprehension scores. For the signaled group, this relationship was found for Geoscience Background Knowledge.

These results point out two major considerations when designing interventions in the classroom. The first is that the lack of domain specific background knowledge may be compensated for by the addition of a signal indicating which elements of the text relate to which elements of the diagram. The second is that the successful use of such prompts may require additional or alternative background knowledge for the successful integration of signal elements as indicated by the effects of Geoscience Background Knowledge in the signaled group.

The differences in attention to components to varying areas of the illustrated textbook so often used in classrooms indicates that numbered signals which link elements of the running text to corresponding elements of the diagram has the largest effect. Given the general lack of attention devoted to diagrams in student reading, increased attention to the diagram is a desirable result. Although this did not result in higher comprehension scores, the increased switching between elements indicates at the very least, a step in the right direction. Eye movements consistent with solutions have been shown to increase solution accuracy and decreased reaction time on problem solving tasks. Unfortunately, severe non-normality in the reaction time data made comparisons of solution time impossible for this sample. Further research should focus on more controlled

presentation of questions, perhaps by experimenter control rather than student control to examine this possibility.

Theoretical Contributions

For Future Research. With respect to future research within the larger diagrammatic reasoning study, there are several directions which may help elucidate some of the findings of this study. First of which is the examination of the think aloud data collected with the comprehension and eye tracking data. This additional process data, when temporally aligned with the eye tracking data can help answer questions brought up by the findings. For example, the increased switching between questions and diagrams for the hyperlink group. It is hypothesized here that this was due to an interference of global processing of the diagram due to the segmentation the hyperlink stimuli produce. Think aloud protocols may contain verbalizations that confirm or deny this hypothesis. Inferences made in the verbalizations can also help elucidate whether greater integration of material for the numbered group exists despite a lack of differences in overall comprehension scores.

Methodological Contributions. This study contributed to the eye tracking literature by examining eye movements over the course of extended reading and question answering. The inclusion of ecologically valid stimuli demonstrates the usefulness of eye tracking in understanding processes involved with comprehension as it exists in a non-laboratory paradigm. Finally the student population extends predominately undergraduate research to a younger population. This aids in the understanding of how to improve STEM understanding in students prior to their self selection to institutions of higher learning.

Practical Contributions. This study demonstrates that a relatively simple addition of numbered circles in the text can reduce the lack of dependence on background knowledge for understanding illustrated textbook materials. However, the use of these prompts may require additional background knowledge in a different domain in order for successful integration to occur. While the existence of the numbered and hyperlink signals was made aware to students, no specific instructions on their use were given. Thus students engaged in the use of the prompts without explicit instruction. A few minutes of additional scaffolding in their use may provide yet a stronger facilitative effect on the comprehension of illustrated science text and may reduce the reliance on additional background knowledge as indicated by the geoscience relationships described. The fact that the numbered prompts resulted in more of the desired eye movement behavior means these changes can be incorporated without the need for expensive technology in the classroom as would be needed for a hyperlink signaled text.

Conclusion

The study presented here examined factors that contribute to the comprehension of illustrated biology texts in a classroom setting, using ecologically valid stimulus materials. The results show that the addition of simple signals in form of numbered circles can change the distribution of fixations on portions of the display as compared to standard text materials with no prompts. However this was not shown to be directly related to the comprehension of the materials in the path models. In regression analysis, the switches made between the text and diagram elements significantly predicted scores. This indicates that although the model fit is good, there may be an element that is absent, which accounts for this lack of significant paths.

Most interesting was the change in relationship between two types of background knowledge and attention to components in a path model found for the two groups. The changes in the relationship between domain specific background knowledge (biology), and a related form of background knowledge (geoscience), points out the complications with including signals in textbooks. While students can use prompts to compensate for lack of domain specific knowledge in the guidance of visual attention, it appears that the successful use of such prompts may depend on background knowledge in a related domain. The solution to this issue may be the inclusion of procedural knowledge on how the signals should be used, which was lacking in this experiment. Further research in this area should take into account these findings when designing materials for student learning.

REFERENCES

- Abrams, R. A., Meyer, D. E., & Kornblum, S. (1989). Speed and accuracy of saccadic eye movements: Characteristics of impulse variability in the oculomotor system. *Journal of Experimental Psychology: Human Perception and Performance*, 15(3), 529-543.
- Ainsworth, S. (2006). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, 16(3), 183-198.
- Alloway, T. P., Gathercole, S. E., Kirkwood, H. J., & Elliott, J. E. (2008). Evaluating the validity of the Automated Working Memory Assessment. *Educational Psychology*, 7, 725–734.
- Bacon, A.M., Handley, S.J., Dennis, I., & Newstead, S.E. (2007) Reasoning strategies: the role of working memory and verbal-spatial ability. *European Journal of Cognitive Psychology*, 20(6), 1065-1086
- Baldi, s., Jin, Y., Skemer,M., Green, P., Herget, D., and Xie, H. (2007) Highlights from PISA 2006: Performance of U.S. 15-year-old students in Science and Mathematics Literacy in an International Context. *US Department of Education, NCES 2008-016*
- Bartholome, T., & Bromme, R. (2009). Coherence formation when learning from text and pictures: What kind of support for whom? *Journal of Educational Psychology*, 101(2), 282-293.
- Bentler, P. (1995). EQS structural equations program manual. Encino, CA: Multivariate Software.

- Britton, B. K., Van Dusen, L., Gülgöz, S., Glynn, S. M., & Sharp, L. (1991). Accuracy of learnability judgments for instructional texts. *Journal of Educational Psychology*, 83(1), 43-47. doi:10.1037/0022-0663.83.1.43
- Boedemer, D. & Faust, U. (2006). External and mental referencing of multiple representations. *Computers in Human Behavior*, 22, 27-42
- Boucheix, J., & Guignard, H. (2005). What animated illustrations conditions can improve technical document comprehension in young students? Format, signaling and control of the presentation. *European Journal of Psychology of Education*, 20(4), 369-388.
- Bowen, G. M., Roth, W., & McGinn, M. K. (1999). Interpretations of graphs by university biology students and practicing scientists: Toward a social practice view of scientific representation practices. *Journal of Research in Science Teaching*, 36(9), 1020-1043.
- Butcher, K. R. (2006). Learning from text with diagrams: Promoting mental model development and inference generation. *Journal of Educational Psychology*, 98(1), 182-197.
- Canham, M., & Hegarty, M. (2010). Effects of knowledge and display design on comprehension of complex graphics. *Learning and Instruction*. DOI: 10.1016/j.learninstruc.2009.02.014.
- Carney, R. N., & Levin, J. R. (2002). Pictorial illustrations still improve students' learning from text. *Educational Psychology Review*, 14(1), 5-26.

- Chang, K.-T., Lenzen, T., & Antes, J. (1985). The Effect of Experience on Reading Topographic Relief Information: Analyses of Performance and Eye Movements. *The Cartographic Journal*, 22(2), 88-94.
- Cook, M. P., Krajcik, J., & Varelas, M. (2006). Visual representations in science education: The influence of prior knowledge and cognitive load theory on instructional design principles. *Science Education*, 90(6), 1073-1091.
- Cooney, J. B., & Swanson, H. L. (1987). Memory and learning disabilities: An overview. In H. L. Swanson (Ed.), *Memory and learning disabilities: Advances in learning and behavioral disabilities (Supplement 2, pp. 1-40)*. Stamford, CT: JAI Press.
- Cromley, J.G., Perez, A.C., Fitzhugh, S.L., Tanaka, J., Newcombe, N.S., Shipley, T.F. & Wills, T.W. (2010) Improving students' diagrammatic Reasoning: a classroom intervention study with eye tracking data, *manuscript submitted February 2010 to Journal of Educational Psychology (paper to be presented at the AERA conference April 30 – May 4, 2010)- received revise and resubmit - resubmitted November 2010*
- Cromley, J. G., Snyder-Hogan, L. E., & Luciw-Dubas, U. A. (2009). Cognitive Activities in Complex Science Text and Diagrams. *Contemporary Educational Psychology*,
- Cromley, J. G., Wills, T. W., Stephens, C. R., Dumas, D., Herring, M. H., Luciw-Dubas, U. A., Snyder-Hogan, L. E., Burton, D., & Mendelsohn, T. (2009). A content analysis of images in biology and geoscience textbooks. *Paper presented at the 2010 annual meeting of the National Association for Research in Science Teaching*, Philadelphia, PA, March 20-24, 2010.

- Ekstrom, R.B., French, J.W., Harmon, H.H. and Demen, D. (1976). Kit of Factor Referenced Cognitive Tests. *Educational Testing Service*
- Erison, K.A., & Kintsch, W. (1995) Long-Term Working Memory. *Paper from Institute of Cognitive Science, University of Colorado, Boulder, ICS Technical Report #94-01*
- Fitzhugh, S.L., Cromley, J.G, Perez, A., Newcombe, N.S and Wills, T (2010). High school students' comprehension of text and diagrams: Testing a model with eye tracking data. *Paper presented at the annual meeting of AERA April-May 2010.*
- Fitzhugh, S.L., Bergey, B., Wills, T., Cromley, J.G. (2011) The effect of diagrammatic reasoning intervention on students attention to text and diagrams; an examination of eye movements. *Poster to be presented at the AERA annual meeting, April 2011*
- Gyselinck, V., Jamet, E., & Dubois, V. (2008). The role of working memory components in multimedia comprehension. *Applied Cognitive Psychology*, 22(3), 353-374.
- Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction*, 15(4), 313-331.
- Ginns, P. (2006). Integrating information: A meta-analysis of the spatial contiguity and temporal contiguity effects. *Learning and Instruction*, 16(6), 511-525.
- Goldberg, H. J., & Kotval, X. P. (1999). Computer interface evaluation using eye movements: Methods and constructs. *International Journal of Industrial Ergonomics*, 24, 631-645.

- Grasser A.C., Kassler, M.A., Kreuz, R.J., & McLain-Allen, B. (1998) Verifications of statements about story words that deviate from normal conceptions of time: What is true about Einstein's dreams? *Cognitive Psychology*, 35(3) pp 246-301
- Graesser, A. C., Lu, S., Olde, B. A., Cooper-Pye, E., & Whitten, S. (2005). Question asking and eye tracking during cognitive disequilibrium: Comprehending illustrated texts on devices when the devices break down. *Memory & Cognition*, 33(7), 1235-1247.
- Grant, E. R., & Spivey, M. J. (2003). Eye movements and problem solving: Guiding attention guides thought. *Psychological Science*, 14(5), 462-466.
- Hannus, M., & Hyona, J. (1999). Utilization of illustrations during learning of science textbook passages among low- and high-ability children. *Contemporary Educational Psychology*, 24(2), 95-123.
- Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations: On the distinction between emotional interest and cognitive interest. *Journal of Educational Psychology*, 89(1), 92-102.
- Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage: A theory of cognitive interest in science learning. *Journal of Educational Psychology*, 90(3), 414-434.
- Hegarty, M. (1992). Mental animation: Inferring motion from static displays of mechanical systems. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 18(5), 1084-1102.
- Hegarty, M., & Just, M. A. (1993). Constructing mental models of machines from text and diagrams. *Journal of Memory and Language*, 32(6), 717-742.

- Hegarty, M., & Sims, V. K. (1994). Individual differences in mental animation during mechanical reasoning. *Memory & Cognition*, 22(4), 411-430.
- Hegarty, M., Kriz, S., & Cate, C. (2003). The roles of mental animations and external animations in understanding mechanical systems. *Cognition & Instruction*, 21(4), 325-360.
- Hegarty, M., Quilici, J., Narayanan, N. H., Holmquist, S., & Moreno, R. (1999). Multimedia instruction: Lessons from evaluation of a theory-based design. *Journal of Educational Multimedia and Hypermedia*, 8(2), 119-150.
- Heiser, J., & Tversky, B. (2006). Arrows in comprehending and producing mechanical diagrams. *Cognitive Science*, 30(3), 581-592.
- Hmelo-Silver, C. E., & Pfeffer, M. G. (2004). Comparing expert and novice understanding of a complex system from the perspective of structures, behaviors, and functions. *Cognitive Science*, 28(1), 127-138.
- Houts, P. S., Doak, C. C., Doak, L. G., & Loscalzo, M. J. (2006). The role of pictures in improving health communication: A review of research on attention, comprehension, recall, and adherence. *Patient Education and Counseling*, 61(2), 173-190.
- Huk, T., & Steinke, M. (2007). Learning cell biology with close-up views or connecting lines: Evidence for the structure mapping effect. *Computers in Human Behavior*, 23(3), 1089-1104.
- Irwin, D. E. (1992b). Visual memory within and across fixations. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 146-165). New York: Springer-Verlag.

- Irwin, D. E. (1998). Lexical processing during saccadic eye movements. *Cognitive Psychology*, 36(1), 1-27.
- Inhoff, A.W., & Radach, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29-54). Oxford, England: Elsevier.
- Ikeda, M., & Saida, S. (1978). Span of recognition in reading. *Vision Research*, 18, 83-88.
- Jacobson, J.Z., & Dodwell, P.C. (1979). Saccadic eye movements during reading. *Brain and Language*, 8, 303
- Just, M., & Carpenter, P. (1985). Cognitive Coordinate Systems: Accounts of Mental Rotation and Individual Differences in Spatial Ability. *Psychological Review*, 92(2), 137-172.
- Just, M., & Carpenter, P. (1976). Eye Fixations and Cognitive Processes. *Cognitive Psychology*, 8, 441-480.
- Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, 92(1), 126-136.
- Kane, M. J., Hambrick, D. Z., Tuholski, S. W., Wilhelm, O., Payne, T. W., & Engle, R. W. (2004). The generality of working memory capacity: A latent-variable approach to verbal and visuospatial memory span and reasoning. *Journal of Experimental Psychology: General*, 133(2), 189-217.

- Kenny, D. A., & McCoach, D. B. (2003). Effect of the number of variables on measures of fit in structural equation modeling. *Structural Equation Modeling, 10*, 333-351.
- Kintsch, E. (1990). Macroprocesses and microprocesses in the development of summarization skill. *Cognition and Instruction, 7*(3), 161-195.
- Kintsch, W. (1988). The role of knowledge in discourse comprehension: A construction-integration model. *Psychological Review, 95*(2), 163-182.
- Kintsch, W. (1994). Text comprehension, memory, and learning. *American Psychologist, 49*(4), 294-303.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, England: Cambridge University Press
- Kintsch, W., Britton, B. K., Fletcher, C. R., Kintsch, E., Mannes, S. M., & Nathan, M. J. (1993). A comprehension-based approach to learning and understanding. In D. L. Medin (Ed.), *The Psychology of Learning and Motivation: Vol. 30* (pp. 165-214). New York: Academic Press.
- Kintsch, W., & van Dijk, T. A. (1978). Towards a model of text comprehension and production. *Psychological Review, 85*, 363-394.
- Kline, R. (2005) Principles and practice of structural equation modeling. Second Edition. *The Guilford Press, New York, NY*
- Kosslyn, S. M. (1994). *Image and brain*. Cambridge, MA: MIT Press.
- Kozma, R., Chin, E., Russell, J., & Marx, N. (2000). The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning. *Journal of the Learning Sciences, 9*(2), 105-143.

- Larkin, J. H., & Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science*, 11(1), 65-100.
- Lemke, J. L. (1998). Multiplying meaning: Visual and verbal semiotics in scientific text. In J. R. Martin, & R. Veel (Eds.), *Reading Science* (pp. 87-113). London: Routledge.
- Levie, W., & Lentz, R. (1982). Effects of text illustrations: A review of research. *Educational Communication & Technology Journal*, 30(4), 195-232.
- Mackworth, N. (1974) The line of sight approach to children's reading and comprehension. In S. Wanat, H. Singer, & M. Kling (Eds.), *Extracting Meaning from Written Language*. Newark, DE: International Reading Association, 1974.
- Mannes, S.M., & Kintsch, W. (1987) Knowledge organization and Text Organization. *Cognition and Instruction*, 4(2), 91-115
- McNamara, D. (2001) Reading both high-coherence and Low-coherence Texts: effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, 55(1), pp 52-62
- Mastropieri, M. A., & Scruggs, T. E. (1989). Constructing more meaningful relationships: Mnemonic instruction for special populations. *Educational Psychology Review*, 1, 83-111.
- Mayer, R. E. (1997). Multimedia learning: Are we asking the right questions?. *Educational Psychologist*, 32(1), 1-19.

- Mayer, R. E., (2005). Introduction to multimedia learning. In R.E. Mayer's [Ed] *The Cambridge Handbook of Multimedia Learning*, Cambridge University Press, New York, NY
- Mayer, R. E., Bove, W., Bryman, A., Mars, R., & Tapangco, L. (1996). When less is more: Meaningful learning from visual and verbal summaries of science textbook lessons. *Journal of Educational Psychology*, 88(1), 64-73.
- Mayer, R. E., Hegarty, M., Mayer, S., & Campbell, J. (2005). When static media promote active learning: Annotated illustrations versus narrated animations in multimedia instruction. *Journal of Experimental Psychology: Applied*, 11(4), 256-265.
- Mayer, R. E., Heiser, J., & Lonn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, 93(1), 187-198.
- Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology*, 90(2), 312-320.
- McNamara, D. (2001) Reading both high-coherence and Low-coherence Texts: effects of text sequence and prior knowledge. *Canadian Journal of Experimental Psychology*, 55(1), pp 52-62.
- McNamara, D. Kintsch, E, Songer, N.B., Kintsch, W (1996). Are Good Texts Always Better? Interactions of Text Coherence, Background Knowledge, and Levels of Understanding in Learning from Text. *Cognition and Instruction* Vol. 14 (1)
- Muthén, L.K. and Muthén, B.O. (1998-2010). Mplus User's Guide. Sixth Edition.

Los Angeles, CA: Muthén & Muthén

- Munzer, S., Seufert, T., & Brunken, R. (2009) Learning from multimedia presentations: facilitation function of animations and spatial ability. *Learning and Individual Differences*, Vol 19(4), pp 481-485
- Nolen, S. B. (2003). Learning environment, motivation, and achievement in high school science. *Journal of Research in Science Teaching*, 40(4), 347-368.
- Novick, L. R. (2006). Understanding spatial diagram structure: An analysis of hierarchies, matrices, and networks. *Quarterly Journal of Experimental Psychology*, 59(10), 1826-1856.
- OECD. 2007 Executive Summary PISA 2006: Science Competencies for Tomorrow's World. Retrieve from www.PISA.oecd.org. Accessed Jan 2011.
- Otero, J., & Campanario, J.M. (1990). Comprehension evaluation and regulation in learning from science texts. *Journal of Research in Science Teaching*, 27, 447-460.
- Otero, J & Campanario, J.M., & Hopkins, K. (1992) The relationship between academic achievement and meta-cognitive comprehension monitoring ability of Spanish secondary students. *Educational and Psychological Measurement*, 52, 419 -428
- Otero, J., & Kintsch, W. (1992) Failures to detect contradictions in text: what readers believe versus what they read. *Psychological Science*, 3, 229-235
- Paas, F., Renkl, A., & Sweller, J. (2003). Cognitive load theory and instructional design: Recent developments. *Educational Psychologist*, 38(1), 1-4.
- Paivio, A. (1971). Imagery and verbal processes. New York: Holt, Rinehart, and Winston.

- Paivio, A. (1991). Dual Coding theory: retrospect and current status. *Canadian Journal of Psychology*, 45(3), 255-287
- Petre, M. (1995). Why looking isn't always seeing: Readership skills and graphical programming. *Communications of the ACM*, 38(6), 33-44.
- Pollatsek, A., Bolozky, S., Well, A. D., & Rayner K. (1981). Asymmetries in the perceptual span for Israeli readers. *Brain and Language*, 14, 174-180.
- Pollatsek, A., & Rayner, K. (1982). Eye movement control in reading: The role of word boundaries. *Journal of Experimental Psychology: Human Perception and Performance*, 8(6), 817-833.
- Pozzer, L. L., & Roth, W. (2003). Prevalence, function, and structure of photographs in high school biology textbooks. *Journal of Research in Science Teaching*, 40(10), 1089-1114.
- Pozzer-Ardenghi, L., & Roth, W. (2005). Making sense of photographs. *Science Education*, 89(2), 219-241.
- Pozzer-Ardenghi, L., & Roth, W. (2007). On performing concepts during science lectures. *Science Education*, 91(1), 96-114.
- Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, 124(3), 372-422.
- Roth, W., Bowen, G. M., & McGinn, M. K. (1999). Differences in graph-related practices between high school biology textbooks and scientific ecology journals. *Journal of Research in Science Teaching*, 36(9), 977-1019.
- Sanchez, C. A., & Wiley, J. (2006). An examination of the seductive details effect in terms of working memory capacity. *Memory & Cognition*, 34(2), 344-355.

- Satorra, A. & P.M. Bentler, (2001). A scaled difference chi-square test statistic for moment structure analysis, *Psychometrika*, 66, 507-514
- Satorra, A. & P M. Bentler, (2010) Ensuring Positiveness of the Scaled Difference Chi-square Test Statistic. *Psychometrika*, 75 (2), 243-248.
- Schmalhofer, F., McDaniel, M. A., & Keefe, D. (2002). A unified model for predictive and bridging inferences. *Discourse Processes*, 33(2), 105-132.
- Schmidt-Weigand, F., Kohnert, A., & Glowalla, U. (2010). A closer look at split visual attention in system- and self-paced instruction in multimedia learning. *Learning and Instruction*. Vol 20(2), pp 100-110 DOI: 10.1016/j.learninstruc.2009.02.011.
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13(2), 141.
- Schnotz, W., (2005) An Integrated Model of Text and Picture Comprehension. In R.E. Mayer's [Ed] *The Cambridge Handbook of Multimedia Learning*, Cambridge University Press, New York, NY
- Schwonke, R., Berthold, K., & Renkl, A. (2010). How multiple external representations are used and how they can be made more useful. *Applied Cognitive Psychology*. DOI: 10.1002/acp.1526.
- Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction*, 13(2), 227-237.
- Seufert, T., & Brünken, R. (2006). Cognitive load and the format of instructional aids for coherence formation. *Applied Cognitive Psychology*, 20(3), 321-331.
- Sims, V. K., & Hegarty, M. (1997). Mental animation in the visuospatial sketchpad: Evidence from dual-task studies. *Memory & Cognition*, 25(3), 321-332.

- Singer, M., & Halldorson, M., (1996) Constructing and validating motive bridging inferences. *Cognitive Psychology*, 30(1),1-38
- Singer, M., & Kintsch, W. (2001) Text retrieval: A theoretical exploration. *Discourse Processes*, 31(1), 27-59
- St Clair-Thompson, H. L., & Gathercole, S. E. (2006). Executive functions and achievements in school: Shifting, updating, inhibition, and working memory. *Quarterly Journal of Experimental Psychology*, 59(4), 745-759.
- Sweller, J. (2005). Implication of cognitive load theory for multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19-30). Cambridge: Cambridge University Press.
- Thompson, V. A., & Paivio, A. (1994). Memory for pictures and sounds: Independence of auditory and visual codes. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale*, 48(3), 380-398.
- Underwood, G., Jebbett, L., & Roberts, K. (2004). Inspecting pictures for information to verify a sentence: Eye movements in general encoding and in focused search. *The Quarterly Journal of Experimental Psychology A: Human Experimental Psychology*, 57(1), 165-182.
- Unsworth, N, Heitz, R., Schrock, J., and Engle, R. (2005) An Automated Version of the operation span task. *Behavioral Research Methods*, 37 (3) 498-505
- Vandenberg, R. & Lance, C., (2000) A Review and Synthesis of the Measurement Invariance Literature: Suggestions, Practices, and Recommendations for Organizational Research. *Organizational Research Methods*, Vol. 3 No. 1, January 2000 4-70

Wiley, J. & Voss, J. F. (1999) Constructing arguments from multiple sources: Tasks that promote understanding and not just memory for text. *Journal of Educational Psychology, 91*, 1-11.

Appendix A: IRB Approval to use data set



TEMPLE
UNIVERSITY®

Office for Human Subjects Protections
Institutional Review Board
Medical Intervention Committees A1 & A2
Social and Behavioral Committee B

3400 North Broad Street
Philadelphia, Pennsylvania 19140
Phone: 215.707.3300 Fax: 215.707.8387
e-mail: richard.throm@temple.edu

MEMORANDUM

To: **SHIPLEY, THOMAS**
LIB ARTS-PSYCHOLOGY (1811)

From: Richard C. Throm
Director, Office for Human Subjects Protection
Institutional Review Board Coordinator

Date: 19-Jan-2011

Re: Exempt Request Status for IRB Protocol:
13636: A Model of the Effects of Background Knowledge and Working Memory on the Comprehension of
Multimedia Biology Text through their Effects on Eye Movements

It has been determined by Expedited Review that this study qualifies for exemption status as follows:

45 CFR 46 Protection of Human Subjects

Section 101 (b): Unless otherwise required by department or agency heads, research activities in which the only involvement of human subjects will be in one or more of the following categories are exempt from this policy:

Exemption 4: Collection or Study of Existing Data. Research involving the collection or study of existing data, documents, records, pathological specimens, or diagnostic specimens, if these sources are publicly available or if the information is recorded by the investigator in such a manner that subjects cannot be identified, directly or through identifiers linked to the subject.

Nothing further is required from you at this time; however, if anything in your research design should change, you must notify the Institutional Review Board immediately.

If you should have any questions, please feel free to contact me at 215-707-8757.

Thank you for keeping the IRB informed of your clinical research.

Appendix B: Letter to parents describing larger study



Department of Psychological Studies in Education (PSE) phone 215-204-5513
1301 W. Cecil B. Moore Avenue fax 215-204-6013
(003-09/10), 2nd Floor Ritter Annex web www.temple.edu/education/pse
Philadelphia, PA 19122-6091

September 21, 2010

Dear Parent/Guardian:

With your permission, we would like to include your child in a study of a method for teaching students to better understand the illustrations in their standard biology textbooks. The study, which will be conducted at Memorial High School, has been approved by Temple University Institutional Review Board (IRB). We expect the study to be completed by the end of June, 2010.

The purpose of this study is to see whether a teaching method is effective in increasing high school students' understanding of textbook illustrations. We will ask your child to complete worksheets during their regular science class using the illustrations from their standard textbook. We will also ask students to answer a few questions about themselves, and complete surveys of knowledge about a biology topic, understanding objects in 3-D (spatial ability), 4) memory for objects in 3-D (spatial working memory), 5) memory for words (verbal working memory), motivation for biology, and by asking some children to read parts of the textbook on a computer which can track what parts of the page they are looking at (eye movements). Rest assured that no part of the "eye tracking" computer equipment will be attached to your child, and no videotaping is involved.

There will be one session of approximately 40 minutes for all students and a second session of 30 minutes for the one-third of the students who participate in the "eye tracking" portion of the study. The study will be conducted at the school, either during class time or after school.

We will take a number of steps to ensure your child's privacy and anonymity. For example, no individuals' names will be used in any written or oral presentations resulting from the study and only a limited number of professional researchers will have access to the original in-class tests. Each student's participation will be fully voluntary. We will describe the general purpose of the study to each student so that each student can decide whether to participate and s/he may withdraw from the study at any time.

We hope that you will consent to your child's participation in this study that may one day contribute to improved learning for many students. If you will read, sign, and date the attached form and have your child return it as soon as possible to his/her teacher, you will be helping us to begin our efforts. All students who return forms will be entered in a drawing to win a Temple t-shirt or baseball cap. If you have any questions, please feel free to contact me using the phone number or email address below.

Thank you for your attention.

Sincerely,

Jennifer Cromley, Assistant Professor
jcromley@temple.edu, (215) 204-8094

**TEMPLE UNIVERSITY
IRB (COMMITTEE B) APPROVAL**

APR 19 2010

**VALID FOR NO MORE
THAN ONE YEAR**

Appendix C: Consent form for Parents



Department of Psychological Studies in Education (PSE) *phone* 215-204-5513
1301 W. Cecil B. Moore Avenue *fax* 215-204-6013
(003-09/10), 2nd Floor Ritter Annex *web* www.temple.edu/education/pse
Philadelphia, PA 19122-6091

Title: Teaching Effective Use of Diagrammatic Reasoning in Biology
Investigator Name: Dr. Jennifer Cromley Participant ID: 5-HS-S-
Department: Psychological Studies in Education, College of Education, Ritter Annex 201
Phone Number: (215) 204-8094

We are currently engaged in a study measuring the effectiveness of a teaching method for diagrams in high school biology textbooks. To help us gain further insights into this area we are asking for your permission to have your child participate in the study. All students in the class will participate in regular class instruction. If you consent, we will ask your child to complete a form about their background and the following six (6) short paper-and-pencil surveys: 1) using illustrations to learn science (diagrammatic reasoning), 2) confidence for learning (self-efficacy), 3) understanding objects in 3-D (spatial ability), 4) memory for objects in 3-D (spatial working memory), 5) memory for words (verbal working memory), and 6) prior knowledge about a biology topic. In addition, about one-third of the children will be selected by the flip of a coin to read parts of the textbook on a computer which can track what parts of the page they are looking at (eye movements).

We do not know of any physical or emotional risk, discomfort, or inconvenience that this study may expose your child to.

The responses your child will provide will be collected using an identifying number, and the responses will only be available to researchers working on this project (i.e., Dr. Cromley and other members of the research team). Your child's participation and anything he/she says during the sessions will be held in the strictest confidence.

We welcome questions about the research at any time. Your child's participation in this study is on a voluntary basis, and you and/or the child may refuse to participate at any time without negative effects.

Questions about your child's rights as a research subject may be directed to Mr. Richard Thom, Office of the Vice President for Research, Institutional Review Board, Temple University, 3400 N. Broad Street, Philadelphia, PA, 19140, (215) 707-8757.

Signing your name below indicates that you have read and understand the contents of this Consent Form and that you consent to your child's taking part in this study.

Yes, I do consent

Date: _____

No thank you, I don't want my child to participate

Parent/guardian signature: _____ Your relation to the student: _____

Parent/guardian printed name: _____

Investigator signature: _____

TEMPLE UNIVERSITY
IRB (COMMITTEE B) APPROVAL

APR 19 2010

VALID FOR NO MORE
THAN ONE YEAR

Appendix D: Biology Background Knowledge Measure

- 1) Scientists use the word bacterium (bacteria) to mean
 - a) All disease causing organisms or germs
 - b) Any harmful substance
 - c) Microorganisms that can only live in animal bodies
 - d) Single celled creatures that cannot be seen without a microscope

- 2) Often biologists will speak of two species as being related or being cousins. When they say this they mean
 - a) They are related the same way you are related to your own cousins
 - b) They share great-great grandparents
 - c) If we go back tens of thousands of generations these two species were once the same species
 - d) They live in the same environment so they act in similar ways

- 3) The scientific name for human beings is Homo sapiens. This name combines the genus and species that we belong to. The next level for grouping animals is an order followed by class, phylum, and kingdom. The reason why biologists group plants and animals like this is to
 - a) Organize information to make it easier to find
 - b) Give every organism a unique name so that they will not be confused
 - c) Show how different organisms are related evolutionarily
 - d) Show how different organisms are related by their shape, color, or other features

- 4) A cell is
 - a) a very small man-made container
 - b) the smallest basic unit of a plant or animal
 - c) an organ, such as the heart, lungs, or brain
 - d) a building block like a protein, fat, or carbohydrate

- 5) A tissue is
 - a) a group of cells working together to do the same job in an organism
 - b) part of animal muscles
 - c) part of organs in plants and animals
 - d) a thin layer of cells in the body that separates one part from another part

- 6) Staining is a tool that biologists use to
 - a) Make tissues look attractive for illustrations in articles and books
 - b) Enhance the true color of tissues to make them easier to see
 - c) To make parts different colors so they look different under a microscope
 - d) To preserve the tissues on the slide so they don't rot

- 7) The main purpose of the heart in an animal is to
 - a) Move blood around the body
 - b) Keep the body warm
 - c) Help it breathe
 - d) Control emotions or behavior

Appendix D (continued)

- 8) Biologists use agar plates to
- Grow microorganisms
 - Hold their instruments
 - Heat chemicals
 - Mix chemicals

- 9) The main purposes of blood in an animal's body are

I	II	III
To keep the body hydrated	To deliver oxygen and take away carbon dioxide	To deliver nutrients, hormones, and other chemical signals, and take away wastes

- I, II, and III
 - I and II only
 - II and III only
 - I and III only
- 10) Which of the following are seeds?

I	II	III	IV
bean	Dots on the outside of a strawberry	Nut	Peach pit

- I and II only
 - II and III only
 - III and IV only
 - I and IV only
- 11) Amphibians are
- Animals like frogs and snakes that are cold blooded
 - Cold blooded animals that must live part of the time in water
 - Animals that have a tail early in their life but lose it later
 - Animals that live on land and lay eggs
- 12) A male peacock uses its tail feathers for
- Building a nest
 - Flight
 - Attracting a mate
 - Distracting its prey while hunting

Appendix D (continued)

- 13) Different parts of animals such as skin, claws, and different colored body parts are adapted for specific purposes. Which of the following is not one of these purposes?
 - a) Camouflage
 - b) Defense against predators
 - c) To look beautiful
 - d) To protect the animal's insides
- 14) When body parts have lots of folds, pockets, or finger-like projections the main advantage of this is
 - a) To catch prey
 - b) To keep away dust or other harmful substances
 - c) To increase surface area
 - d) To store food or nutrients
- 15) Three of the following things are true for all mammals. Which one is **not** true for all mammals?
 - a) All mammals feed milk to their young
 - b) All mammals give birth to live young including the duck-bill platypus
 - c) All mammals are warm blooded
 - d) All mammals have hair
- 16) The body has many enzymes that catalyze different chemical reactions. The reason why the body needs this is in order to
 - a) Speed up chemical reactions
 - b) Slow down chemical reactions that are going too fast
 - c) Break down food or other substances
 - d) Block unwanted processes from happening
- 17) Embryos are
 - a) Young animals
 - b) Cells that fuse together and eventually grow into a young animal
 - c) An animal in its very early stage of development
 - d) An unfertilized egg
- 18) What does fixing a chemical mean?
 - a) To temporarily store a chemical by turning it into another chemical
 - b) To repair a chemical that has been damaged
 - c) To use a chemical to hold something in place
 - d) To permanently bind one chemical to another
- 19) Stomata are
 - a) Cells on the root of a plant that absorb water
 - b) Pores on plant leaves that keep water inside the plant
 - c) Cells on plant leaves that make plants green
 - d) Pores on plant leaves that release food waste

Appendix D (continued)

- 20) A vacuole is
- A cover to protect the root tip
 - A sac (open space) in the cell
 - A cell in the plant root
 - A membrane on the plant leaf
- 21) Cytoplasm is
- The nucleus of a cell
 - The tissue inside of a cell
 - The jelly-like like substance inside a cell
 - The open space inside a cell
- 22) All of the following molecules are found in the body. Arrange them in the order from the smallest to the largest molecule.
- Water (H_2O), glucose (sugar), collagen (a protein), RNA
 - RNA, glucose (sugar), collagen (a protein), Water (H_2O)
 - Water (H_2O), collagen (a protein), glucose (sugar), RNA
 - RNA, collagen (a protein), glucose (sugar), Water (H_2O)
- 23) Osmosis is
- When a cell combines with another cell
 - When something is absorbed with no effort or energy
 - Movement of a chemical across a semi-permeable membrane from an area of higher concentration to an area of lower concentration
 - A step in the body's defense system
- 24) When an animal hatches from an egg or is born
- All animals look like a small version of the adult animal
 - All animals have the same body parts as an adult but some parts will be relatively smaller or larger
 - Some animals go through a stage where they look very different but all of them quickly grow to look like the adult animal
 - Some animals go through many stages where they look different from the adult animal

Appendix E: Geoscience Background Knowledge Measure

GEO Prior Knowledge Measure

- 26) The inside of the Earth is:
- hollow.
 - completely solid.
 - made of solid and liquid layers.
 - made of gas, liquid, and solid layers.
- 27) A volcano is:
- melted rock that pushes out from under the crust.
 - an underground explosion.
 - the result of an earthquake that blows a hole in the crust of the Earth.
 - due to chemical reactions under the crust of the Earth.
- 28) The bottom of the sea:
- has been the same since the Earth was first formed.
 - has been the same since the age of the dinosaurs.
 - only changes when an underwater volcano erupts.
 - is constantly changing.
- 29) When a hotter rock pushes into a cooler rock:
- the two rocks immediately mix until they reach the same temperature.
 - the cooler rock sinks.
 - the hotter rock covers up the cooler rock.
 - the two rocks stop each other from moving.
- 30) Earthquakes can be detected because they:
- make a lot of noise.
 - release energy in seismic waves.
 - cause volcanoes, and the volcanoes are a sign that an earthquake has happened.
 - cause huge waves called tsunamis, and the tsunamis are the sign that an earthquake has happened.
- 31) Water vapor moves into the air by _____ and moves out of the air by _____
- boiling, rain
 - wind, rain
 - evaporation, condensation
 - splitting in a chemical reaction, re-forming in a chemical reaction

Appendix E (continued)

GEO Prior Knowledge Measure

- 32) Geologists would describe the surface of the Earth as
- made up of large sections of rock called plates, which make up the crust.
 - a very thin layer of dirt, with molten rock below.
 - mostly water with islands called continents floating on top of the water.
 - all solid rock (bedrock) throughout the planet, with a thin layer of dirt or water over top.
- 33) Air is made up almost entirely of:
- oxygen.
 - oxygen and nitrogen.
 - oxygen and carbon dioxide.
 - oxygen, water vapor, and rare gases.
- 34) A greenhouse (a building built mostly out of clear glass windows) helps to grow plants because
- it keeps water inside.
 - the energy from the sun heats the air inside.
 - the glass filters out harmful sun rays.
 - helpful gases build up inside the building.
- 35) Which of the following are all forms of energy?
- Force, light, and fire
 - Heat, light, sound, and motion
 - Gravity, heat, chemical reactions, and light
 - Food, light, and motion

Appendix F: Example control group stimuli

Whenever the cell divides, the provirus also divides, resulting in two infected host cells. In this cycle, called the **lysogenic** (*lie soh JEHN ihk*) cycle, the viral genome replicates without destroying the host cell. This cycle is shown in Figure 4. In some lysogenic viruses, a change in the environment can cause the provirus to begin the lytic cycle. This results in the destruction of the host cell.

In animal cells, viruses can replicate slowly so that the host cell is not destroyed by the virus. For example, the virus that causes cold sores in humans hides deep in the nerves of the face. When the conditions in the body become favorable for the virus, such as when a person is under stress, the virus then begins to cause tissue damage that is seen as a cold sore or fever blister.

Host Cell Specificity

Viruses are often restricted to certain kinds of cells. For example, TMV infects tobacco and related plants, but does not infect animals. Scientists hypothesize that this specificity may be due to the viruses' origin. Viruses may have originated when fragments of host genes escaped or were expelled from cells. The hypothesis that viruses originated from a variety of host cells may explain why there are so many different kinds of viruses. Biologists think there are at least as many kinds of viruses as there are kinds of organisms.

Structure of HIV—an Enveloped Virus

Many viruses that infect only animals, such as the influenza virus shown in Figure 2, have an exterior viral envelope. **Figure 5** shows human immunodeficiency virus (HIV), the virus that causes acquired immune deficiency syndrome (AIDS). Figure 5 illustrates the envelope and other features common to several animal viruses. In many cases, the viral envelope is composed of a lipid bilayer derived from the membrane of the host cell. On the surface of the virus, glycoproteins are embedded within the envelope.

Within the envelope lies the capsid, which in turn encloses the virus's genetic material. In the case of HIV, the genetic material is composed of two molecules of single-stranded RNA. The approximately 9,000 nucleotides of HIV make up nine genes. Three of these genes are common to many different viruses.

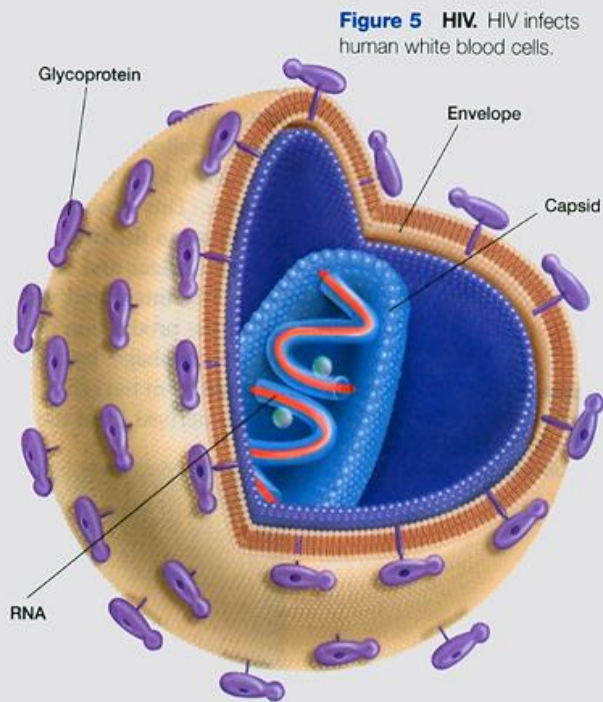


Figure 5 HIV. HIV infects human white blood cells.

Appendix G: Example numbered group stimuli

Host Cell Specificity

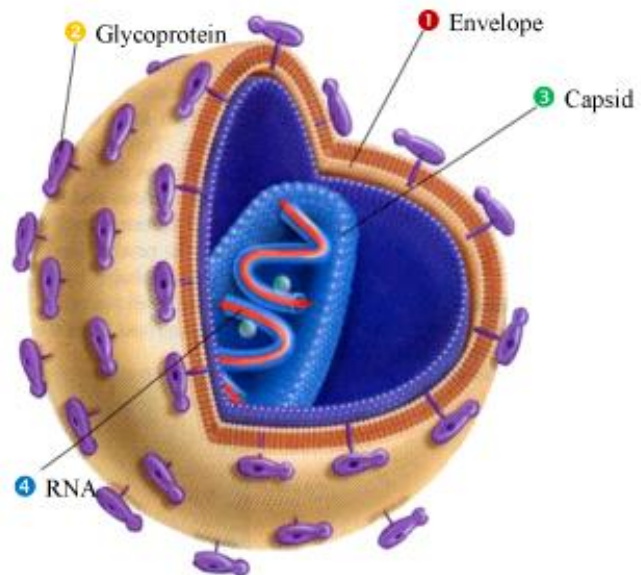
Viruses are often restricted to certain kinds of cells. For example, TMV infects tobacco and related plants, but does not infect animals. Scientists hypothesize that this specificity may be due to the viruses' origin. Viruses may have originated when fragments of host genes escaped or were expelled from cells. The hypothesis that viruses originated from a variety of host cells may explain why there are so many different kinds of viruses. Biologists think there are at least as many kinds of viruses as there are kinds of organisms.

Structure of HIV – an Enveloped Virus

Many viruses that infect only animals, such as the influenza virus shown in Figure 2, have an exterior viral envelope. Figure 5 shows human immunodeficiency virus (HIV), the virus that causes acquired immune deficiency syndrome (AIDS). Figure 5 illustrates the envelope 1 and other features common to several animal viruses. In many cases, the viral envelope is composed of a lipid bilayer derived from the membrane of the host cell. On the surface of the virus, glycoproteins 2 are embedded within the envelope.

Within the envelope lies the capsid 3, which in turn encloses the viruses' genetic material. In the case of HIV, the genetic material is composed of two molecules of single-stranded RNA 4. The approximately 9,000 nucleotides of HIV make up 9 genes. Three of these genes are common to many different viruses.

Figure 5 HIV. HIV infects human white blood cells.



Host Cell Specificity

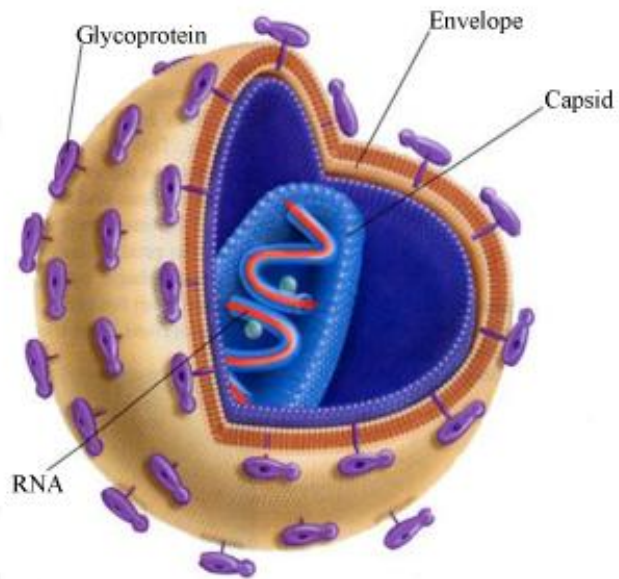
Viruses are often restricted to certain kinds of cells. For example, TMV infects tobacco and related plants, but does not infect animals. Scientists hypothesize that this specificity may be due to the viruses' origin. Viruses may have originated when fragments of host genes escaped or were expelled from cells. The hypothesis that viruses originated from a variety of host cells may explain why there are so many different kinds of viruses. Biologists think there are at least as many kinds of viruses as there are kinds of organisms.

Structure of HIV – an Enveloped Virus

Many viruses that infect only animals, such as the influenza virus shown in Figure 2, have an exterior viral envelope. Figure 5 shows human immunodeficiency virus (HIV), the virus that causes acquired immune deficiency syndrome (AIDS). Figure 5 illustrates the [envelope](#) and other features common to several animal viruses. In many cases, the viral envelope is composed of a lipid bilayer derived from the membrane of the host cell. On the surface of the virus, [glycoproteins](#) are embedded within the envelope.

Within the envelope lies the [capsid](#), which in turn encloses the viruses' genetic material. In the case of HIV, the genetic material is composed of two molecules of single-stranded [RNA](#). The approximately 9,000 nucleotides of HIV make up 9 genes. Three of these genes are common to many different viruses.

Figure 5 HIV. HIV infects human white blood cells.



Host Cell Specificity

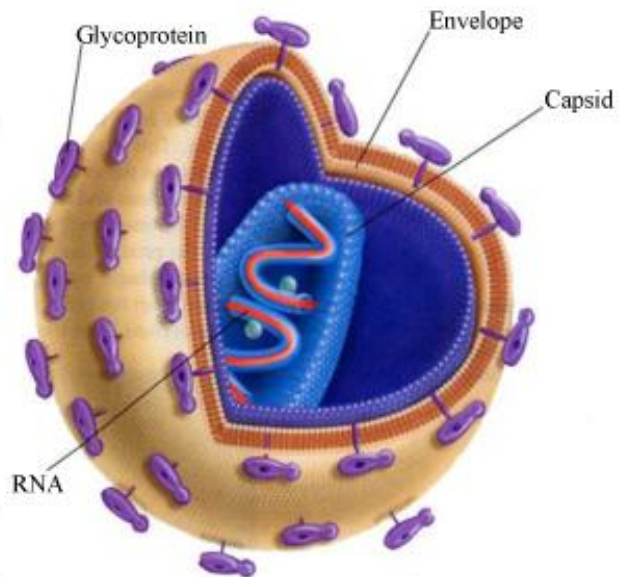
Viruses are often restricted to certain kinds of cells. For example, TMV infects tobacco and related plants, but does not infect animals. Scientists hypothesize that this specificity may be due to the viruses' origin. Viruses may have originated when fragments of host genes escaped or were expelled from cells. The hypothesis that viruses originated from a variety of host cells may explain why there are so many different kinds of viruses. Biologists think there are at least as many kinds of viruses as there are kinds of organisms.

Structure of HIV – an Enveloped Virus

Many viruses that infect only animals, such as the influenza virus shown in Figure 2, have an exterior viral envelope. Figure 5 shows human immunodeficiency virus (HIV), the virus that causes acquired immune deficiency syndrome (AIDS). Figure 5 illustrates the [envelope](#) and other features common to several animal viruses. In many cases, the viral envelope is composed of a lipid bilayer derived from the membrane of the host cell. On the surface of the virus, [glycoproteins](#) are embedded within the envelope.

Within the envelope lies the [capsid](#), which in turn encloses the viruses' genetic material. In the case of HIV, the genetic material is composed of two molecules of single-stranded [RNA](#). The approximately 9,000 nucleotides of HIV make up 9 genes. Three of these genes are common to many different viruses.

Figure 5 HIV. HIV infects human white blood cells.



Appendix H: Example second screen hyperlink stimuli

Host Cell Specificity

Viruses are often restricted to certain kinds of cells. For example, TMV infects tobacco and related plants, but does not infect animals. Scientists hypothesize that this specificity may be due to the viruses' origin. Viruses may have originated when fragments of host genes escaped or were expelled from cells. The hypothesis that viruses originated from a variety of host cells may explain why there are so many different kinds of viruses. Biologists think there are at least as many kinds of viruses as there are kinds of organisms.

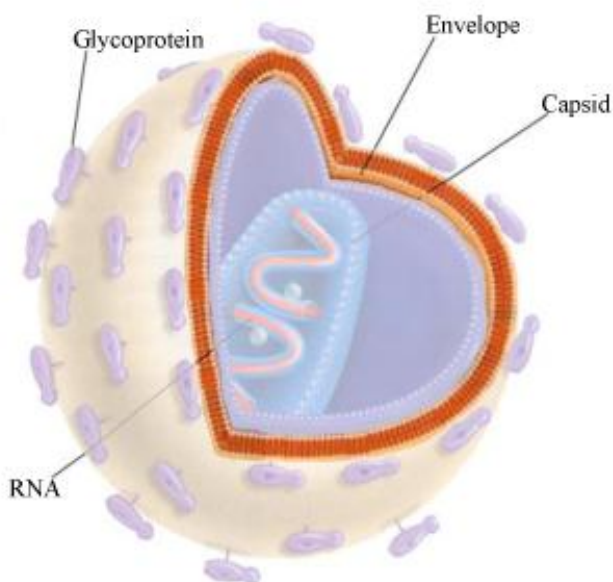
Structure of HIV – an Enveloped Virus

Many viruses that infect only animals, such as the influenza virus shown in Figure 2, have an exterior viral envelope. Figure 5 shows human immun-

odeficiency virus (HIV), the virus that causes acquired immune deficiency syndrome (AIDS). Figure 5 illustrates the [envelope](#) and other features common to several animal viruses. In many cases, the viral envelope is composed of a lipid bilayer derived from the membrane of the host cell. On the surface of the virus, [glycoproteins](#) are embedded within the envelope.

Within the envelope lies the [capsid](#), which in turn encloses the viruses' genetic material. In the case of HIV, the genetic material is composed of two molecules of single-stranded [RNA](#). The approximately 9,000 nucleotides of HIV make up 9 genes. Three of these genes are common to many different viruses.

Figure 5 HIV. HIV infects human white blood cells.



Host Cell Specificity

Viruses are often restricted to certain kinds of cells. For example, TMV infects tobacco and related plants, but does not infect animals. Scientists hypothesize that this specificity may be due to the viruses' origin. Viruses may have originated when fragments of host genes escaped or were expelled from cells. The hypothesis that viruses originated from a variety of host cells may explain why there are so many different kinds of viruses. Biologists think there are at least as many kinds of viruses as there are kinds of organisms.

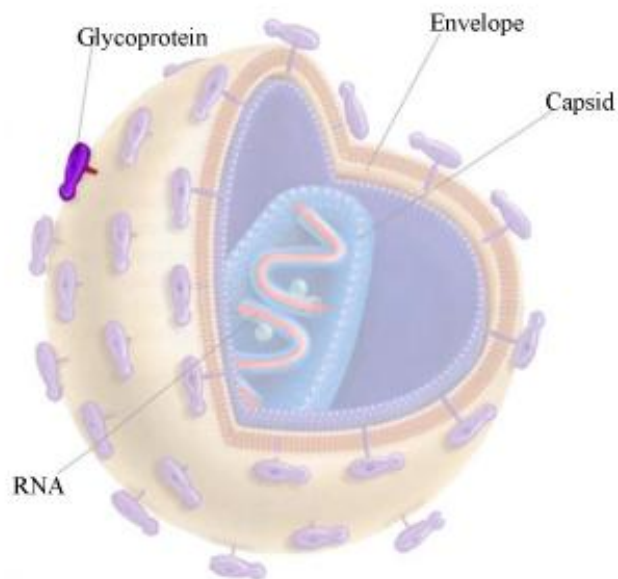
Structure of HIV – an Enveloped Virus

Many viruses that infect only animals, such as the influenza virus shown in Figure 2, have an exterior viral envelope. Figure 5 shows human immun-

odeficiency virus (HIV), the virus that causes acquired immune deficiency syndrome (AIDS). Figure 5 illustrates the [envelope](#) and other features common to several animal viruses. In many cases, the viral envelope is composed of a lipid bilayer derived from the membrane of the host cell. On the surface of the virus, [glycoproteins](#) are embedded within the envelope.

Within the envelope lies the [capsid](#), which in turn encloses the viruses' genetic material. In the case of HIV, the genetic material is composed of two molecules of single-stranded [RNA](#). The approximately 9,000 nucleotides of HIV make up 9 genes. Three of these genes are common to many different viruses.

Figure 5 HIV. HIV infects human white blood cells.



Appendix H: Example third screen hyperlink stimuli

Host Cell Specificity

Viruses are often restricted to certain kinds of cells. For example, TMV infects tobacco and related plants, but does not infect animals. Scientists hypothesize that this specificity may be due to the viruses' origin. Viruses may have originated when fragments of host genes escaped or were expelled from cells. The hypothesis that viruses originated from a variety of host cells may explain why there are so many different kinds of viruses. Biologists think there are at least as many kinds of viruses as there are kinds of organisms.

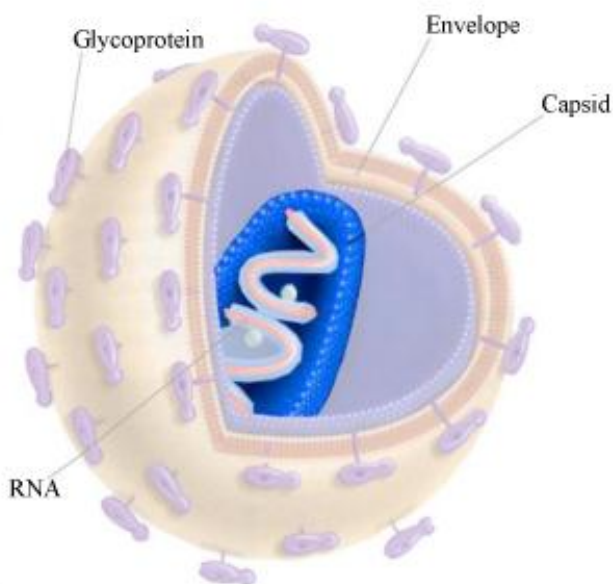
Structure of HIV – an Enveloped Virus

Many viruses that infect only animals, such as the influenza virus shown in Figure 2, have an exterior viral envelope. Figure 5 shows human immun-

odeficiency virus (HIV), the virus that causes acquired immune deficiency syndrome (AIDS). Figure 5 illustrates the [envelope](#) and other features common to several animal viruses. In many cases, the viral envelope is composed of a lipid bilayer derived from the membrane of the host cell. On the surface of the virus, [glycoproteins](#) are embedded within the envelope.

Within the envelope lies the [capsid](#), which in turn encloses the viruses' genetic material. In the case of HIV, the genetic material is composed of two molecules of single-stranded [RNA](#). The approximately 9,000 nucleotides of HIV make up 9 genes. Three of these genes are common to many different viruses.

Figure 5 HIV. HIV infects human white blood cells.



Appendix H: Example fifth screen hyperlink stimuli

Host Cell Specificity

Viruses are often restricted to certain kinds of cells. For example, TMV infects tobacco and related plants, but does not infect animals. Scientists hypothesize that this specificity may be due to the viruses' origin. Viruses may have originated when fragments of host genes escaped or were expelled from cells. The hypothesis that viruses originated from a variety of host cells may explain why there are so many different kinds of viruses. Biologists think there are at least as many kinds of viruses as there are kinds of organisms.

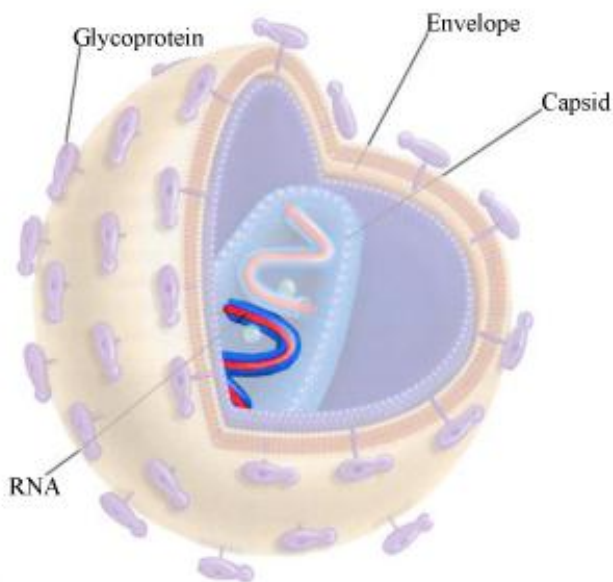
Structure of HIV – an Enveloped Virus

Many viruses that infect only animals, such as the influenza virus shown in Figure 2, have an exterior viral envelope. Figure 5 shows human immun-

odeficiency virus (HIV), the virus that causes acquired immune deficiency syndrome (AIDS). Figure 5 illustrates the [envelope](#) and other features common to several animal viruses. In many cases, the viral envelope is composed of a lipid bilayer derived from the membrane of the host cell. On the surface of the virus, [glycoproteins](#) are embedded within the envelope.

Within the envelope lies the [capsid](#), which in turn encloses the viruses' genetic material. In the case of HIV, the genetic material is composed of two molecules of single-stranded [RNA](#). The approximately 9,000 nucleotides of HIV make up 9 genes. Three of these genes are common to many different viruses.

Figure 5 HIV. HIV infects human white blood cells.



Appendix I: Scoring rubric for transcribed comprehension answers

Page	Q1	Q2	Q3	Answer_1	Answer_2	Answer_3: Answers must show evidence of integrating text and diagrams.
437	20	4	24	<p>2= Viruses may have originated when genes from a host cell were expelled or escaped.</p> <p>1= some reference to viruses attacking specific cells or organisms OR incomplete (e.g., "originate from the host")</p>	<p>2 = They are different colors OR the glycoprotein penetrates the envelope OR cross section shows a piece of the glycoprotein sticking into the envelope OR references the text (e.g., "in the text, it says the glycoprotein is embedded")</p> <p>1 = no possible codes</p> <p>0 = Restates question (e.g., "you can tell because it's attached," "cause it's sealed")</p>	<p>2 – The envelope is the label that indicates the lipid bilayer – part 2: should reference that the text said the envelope was made of a lipid bilayer and/or that the diagram shows two layers in the envelope</p> <p>1 – Envelope (by itself with no other explanation)</p>
623	20	4	24	<p>2 = role of the nematocytes sting the prey so the cnidarians can push it with their tentacles into the mouth.</p> <p>1 = no possibility</p>	<p>2 = Nerve cells are found in the Endoderm</p> <p>1 = no possibility</p>	<p>2 = The mouth is a small opening in the center of the tentacles. I can tell b/c the diagram shows the opening there and the text states that fingerlike tentacles surround the opening to the gastro vascular cavity; any mention of the tentacles' role in feeding</p> <p>1 = If just refers to diagram, such as "it looks like it's on top" or "where the hole is"</p>
624	20	4	24	<p>2 = either or solitary polyp, no medusa stage or mention of difference in habitats (fresh water vs. salt water)</p> <p>1 = no possible code</p>	<p>2 = Basal Dick</p> <p>1= no possible</p>	<p>2 = White (cannot get this answer without reading the text and looking at the diagram)</p> <p>1 = no possible</p> <p>0 = blue, green</p>
694	20	4	24	<p>2 = Central ring of the nervous system</p> <p>1 = mention of the arms acting independently of one another without mention of the central ring.</p>	<p>2= spines or ossicles (or the definition of ossicles as spint protrusions)</p> <p>1= no possible code</p>	<p>2 = Yes AND one of the following additional statements: the paragraph says most have 5 arms but the number of arms can vary OR radiate from a central point OR divisible by five OR all sea stars are radially semetrical</p> <p>1 = Yes only; OR Yes AND it has 10 arms and can divide evenly, if you fold it in half it will match</p>

Appendix I: Scoring rubric for transcribed comprehension answers (continued)

Page	Q1	Q2	Q3	Answer_1	Answer_2	Answer_3: Answers must show evidence of integrating text and diagrams.
725	20	4	24	<p>2 = Two characteristics need: 1) No breastbone to anchor its flight muscles; 2) no hollow bones</p> <p>1 = either one correct characteristic and/or "long reptilian tail."</p>	<p>2 = Coloring is artificial as indicated by the caption which refers to the drawing and the artist's rendition OR false b/c animal is extinct and feathers are not preserved in the fossil record</p> <p>1 = no code possible</p> <p>Note: Don't need to explain how they arrived at their answer.</p>	<p>2 = Fused collarbone OR wishbone</p> <p>1 = no possible</p>
726	20	4	24	<p>2 = they use the DNA similarity of current birds to infer the evolution of birds</p> <p>1 = Only one correct answer</p>	<p>2 = High metabolism OR eating frequently</p> <p>1 = naming a physical appearance (wings, feathers, beaks) or functional (they fly, they eat seeds)</p>	<p>2= The American Kestrel might also be a predator because it is similar in appearance to a hawk OR beak and/or talons and/or its eyes are for hunting, are sharp, etc.</p> <p>1 = American Kestrel with no rationale; OR predatory arguments for other birds (e.g., Ostrich b/c it can run fast)</p>
826	20	4	24	<p>2 = genetically pre-disposed to the behavior; OR <i>without being taught</i>, action occurs in the same way; OR no variation between parent and offspring; OR from generation to generation; OR instinct</p> <p>1 = JUST "b/c the action always occurs the same way"</p>	<p>2 = What happens when two animals with distinct innate behaviors are mated - the innate behavior of the offspring shows signs of inheritance from both parents (can give partial credit for nest building behavior)</p> <p>1 = nest building behavior OR innate behavior</p>	<p>2 = The single long nesting material used by Fischer's lovebirds. The diagram pictures several shorter pieces in the beak.</p> <p>1= Vague/incomplete answer (e.g., "possibly the length")</p>

Appendix J: Correlation Matrix of all variables

Correlation Matrix of Raw Scores										
	1	2	3	4	5	6	7	8	9	10
1. WM	1									
2. BBK	.031	1.000								
3. GBK	.031	.177*	1.000							
4. COMP	.066	.087	.309**	1.000						
5. RT	.069	-.067	-.375**	-.187*	1.000					
6. RTT _{xt}	-.105	-.226*	.417**	.345**	-.311**	1.000				
7. RTD _{ia}	-.084	-.224*	.405**	.382**	-.211*	.750**	1.000			
8. RTInt	.009	-.109	.239**	.268**	.083	.332**	.378**	1.000		
9. COMPT	-.052	-.104	.138	.337**	-.082	.436**	.335**	.194*	1.000	
10. COMP	.092	-.057	.176	.137	-.222*	.370**	.263**	.247**	.402**	1.000
11. COMP	.045	.007	.296**	.919**	-.197*	.348**	.390**	.262**	.360**	.224*
12. FIX	.074	.134	-.268**	-.213*	.236*	-.298**	-.285**	-.099	-.152	-.138
13. FIXT	.037	-.155	-.068	-.114	.008	-.049	-.097	.083	-.066	.035
14. FIXD	-.044	.145	.042	.157	.032	-.049	-.049	-.182	.165	-.154
15. FIXS	-.012	.176	-.412**	-.323**	.366**	-.454**	-.448**	-.028	-.206*	-.255**
16. FIXQ	-.008	.014	-.004	-.034	-.119	.121	.174	.018	-.112	.116
17. DUR	0.026	-.148	-.142	-.020	.034	-.024	-.062	.025	.032	-.005
18. DURT	0.079	-.135	-.135	-.170	.132	-.127	-.212*	.131	-.103	-.027
19. DURD	-0.021	.021	.104	.191*	-.046	.107	.117	-.125	.280**	.089
20. DURS	0.038	.099	-.350**	-.211*	.309**	-.331**	-.331**	-.002	-.161	-.108
21. DURQ	-0.003	.188*	.027	.014	-.153	-.091	.006	-.174	-.119	-.058
22. SWTYPE	0.075	.166	-.326**	-.257**	.363**	-.426**	-.432**	-.154	-.224*	-.311**
23. SWTD	0.087	-.090	-.121	-.035	.087	-.146	-.148	-.121	.059	-.112
24. SWTS	0.012	.253**	-.402**	-.343**	.310**	-.435**	-.434**	-.093	-.226*	-.250**
25. SWTQ	-0.086	-.067	-.088	-.136	-.127	-.087	-.119	.108	-.286**	-.003
26. SWDS	-0.022	.080	-.385**	-.195*	.365**	-.354**	-.329**	-.022	-.146	-.246**
27. SWDQ	0.022	.041	.245**	.351**	-.192*	.264**	.237*	.065	.257**	.115
28. SWSQ	-0.050	-.003	-.266**	-.274**	.258**	-.293**	-.315**	.019	-.139	-.156
29. WSW	0.058	.080	-.167	-.132	.071	-.122	-.096	-.031	-.055	.048
30. WSWT	0.034	-.177	.018	-.028	-.050	.072	.015	.110	-.003	.066
31. WSWD	-0.066	.192*	.006	.074	.092	-.141	-.130	-.191*	.079	-.215*
32. WSWs	0.034	-.042	-.288**	-.109	.268**	-.243**	-.228*	.089	-.040	.011
33. WSWQ	-0.002	.009	-.058	-.078	-.065	.066	.120	.011	-.102	.135
34. REL	-0.016	-.219*	.234*	.125	-.215*	.410**	.353**	.179	.120	.178
35. RELT	0.037	-.218*	.022	-.056	-.169	.146	-.007	.248**	.046	.106
36. RELD	-0.053	-.034	.156	.178	.043	.168	.185	.033	.134	-.027
37. RELS	0.028	.092	-.398**	-.257**	.375**	-.381**	-.356**	.010	-.163	-.212*
38. RELQ	-0.015	-.026	.212*	.109	-.184	.276**	.363**	-.084	.006	.173

Appendix J: continued

Correlation Matrix of Raw Scores										
	11	12	13	14	15	16	17	18	19	20
1. WM										
2. BBK										
3. GBK										
4. COMP										
5. RT										
6. RTTxt										
7. RTDia										
8. RTInt										
9. COMPT										
10. COMPD										
11. COMPI	1.000									
12. FIX	-.255**	1.000								
13. FIXT	-.078	-.219*	1.000							
14. FIXD	.057	.419**	-.629**	1.000						
15. FIXS	-.335**	.458**	-.061	-.011	1.000					
16. FIXQ	.015	-.211*	-.492**	-.241**	-.004	1.000				
17. DUR	.037	-.616**	.031	-.323**	-.127	.289**	1.000			
18. DURT	-.136	.022	.794**	-.440**	.161	-.444**	-.088	1.000		
19. DURD	.138	.276**	-.592**	.812**	-.125	-.067	-.200*	-.573**	1.000	
20. DURS	-.222*	.488**	-.051	.018	.854**	-.037	-.191*	.124	-.043	1.000
21. DURQ	.005	-.177	-.502**	-.012	-.054	.717**	.223*	-.689**	.015	-.114
22. SWTYPE	-.315**	.896**	-.217*	.380**	.619**	-.165	-.470**	.062	.188*	.572**
23. SWTD	-.089	.258**	.338**	.237*	-.002	-.601**	-.228*	.441**	.099	.054
24. SWTS	-.362**	.529**	-.026	.032	.929**	-.071	-.216*	.223*	-.109	.762**
25. SWTQ	-.094	-.302**	.453**	-.616**	-.093	.194*	.262**	.358**	-.623**	-.126
26. SWDS	-.185*	.299**	-.056	-.042	.741**	.063	.009	.046	-.156	.632**
27. SWDQ	.303**	.074	-.682**	.594**	-.189*	.263**	-.112	-.629**	.631**	-.132
28. SWSQ	-.294**	.157	-.205*	-.042	.491**	.215*	.019	-.049	-.060	.380**
29. WSW	-.150	.909**	-.177	.376**	.220*	-.216*	-.636**	-.019	.307**	.316**
30. WSWT	.001	-.263**	.971**	-.590**	-.131	-.484**	-.003	.758**	-.539**	-.099
31. WSWD	-.019	.455**	-.617**	.962**	.072	-.247**	-.329**	-.418**	.741**	.062
32. WSWs	-.094	.221*	-.058	-.073	.700**	.063	.008	.039	-.059	.808**
33. WSWQ	-.014	-.222*	-.420**	-.327**	.070	.960**	.361**	-.397**	-.118	.026
34. REL	.160	-.200*	-.023	.021	-.265**	.077	.027	-.011	.119	-.207*
35. RELT	-.022	-.142	.649**	-.471**	-.116	-.312**	.047	.579**	-.377**	-.082
36. RELD	.159	.255**	-.473**	.697**	-.065	-.150	-.280**	-.356**	.606**	-.015
37. RELS	-.265**	.406**	-.002	-.024	.892**	-.054	-.142	.194*	-.111	.841**
38. RELQ	.131	-.408**	-.353**	-.054	-.314**	.611**	.265**	-.396**	.066	-.313**

Appendix J: (continued)

Correlation Matrix of Raw Scores										
	21	22	23	24	25	26	27	28	29	30
1. WM										
2. BBK										
3. GBK										
4. COMP										
5. RT										
6. RTT _{xt}										
7. RTD _{ia}										
8. RTI _{nt}										
9. COMPT										
10. COMPD										
11. COMPI										
12. FIX										
13. FIXT										
14. FIXD										
15. FIXS										
16. FIXQ										
17. DUR										
18. DURT										
19. DURD										
20. DURS										
21. DURQ	1.000									
22. SWTYPE	-.134	1.000								
23. SWTD	-.551**	.245**	1.000							
24. SWTS	-.130	.646**	.043	1.000						
25. SWTQ	.086	-.268**	-.258**	-.085	1.000					
26. SWDS	.118	.426**	-.040	.578**	-.062	1.000				
27. SWDQ	.333**	-.010	-.345**	-.204*	-.442**	-.107	1.000			
28. SWSQ	.118	.241**	-.217*	.320**	.002	.303**	-.080	1.000		
29. WSW	-.185*	.629**	.224*	.319**	-.275**	.123	.136	.048	1.000	
30. WSWT	-.510**	-.279**	.310**	-.103	.354**	-.095	-.563**	-.252**	-.199*	1.000
31. WSWD	.000	.457**	.158	.127	-.502**	-.010	.496**	.033	.368**	-.622**
32. WSWs	-.025	.285**	.019	.487**	-.105	.554**	-.112	.392**	.119	-.086
33. WSWQ	.678**	-.176	-.512**	-.011	.141	.134	.128	.241**	-.224*	-.435**
34. REL	-.107	-.184*	.098	-.258**	-.148	-.259**	.130	-.189*	-.176	.046
35. RELT	-.478**	-.168	.368**	-.078	.162	-.147	-.427**	-.212*	-.091	.641**
36. RELD	-.097	.218*	.126	-.053	-.491**	-.042	.506**	-.098	.241**	-.415**
37. RELS	-.127	.538**	.050	.842**	-.100	.580**	-.194*	.413**	.205*	-.054
38. RELQ	.521**	-.345**	-.419**	-.350**	.078	-.239*	.248**	.012	-.387**	-.295**

Appendix J: continued

Correlation Matrix of Raw Scores								
	31	32	33	34	35	36	37	38
1. WM								
2. BBK								
3. GBK								
4. COMP								
5. RT								
6. RTT _{xt}								
7. RTD _{ia}								
8. RTI _{nt}								
9. COMPT								
10. COMPD								
11. COMPI								
12. FIX								
13. FIXT								
14. FIXD								
15. FIXS								
16. FIXQ								
17. DUR								
18. DURT								
19. DURD								
20. DURS								
21. DURQ								
22. SWTYPE								
23. SWTD								
24. SWTS								
25. SWTQ								
26. SWDS								
27. SWDQ								
28. SWSQ								
29. WSW								
30. WSWT								
31. WSWD	1.000							
32. WSWs	-.088	1.000						
33. WSWQ	-.348**	.160	1.000					
34. REL	-.045	-.172	.057	1.000				
35. RELT	-.492**	-.083	-.241**	.432**	1.000			
36. RELD	.640**	-.073	-.206*	.424**	-.321**	1.000		
37. RELS	.036	.680**	.015	-.158	-.038	-.064	1.000	
38. RELQ	-.078	-.185*	.541**	.487**	-.281**	.041	-.280**	1.000

