EFFECTIVENESS OF HOLISTIC MENTAL MODEL CONFRONTATION IN DRIVING CONCEPTUAL CHANGE

by

Soniya Gadgil

B.Arch, University of Pune, 2000

Submitted to the Graduate Faculty of

School of Arts and Sciences in partial fulfillment

of the requirements for the degree of

Master of Science

University of Pittsburgh

2009

UNIVERSITY OF PITTSBURGH

SCHOOL OF ARTS AND SCIENCES

This thesis was presented

by

Soniya Gadgil

It was defended on

December 22, 2008

and approved by

Dr. Timothy Nokes, Assistant Professor, Department of Psychology

Dr. Christian Schunn, Associate Professor, Department of Psychology

Thesis Director: Dr. Michelene T.H.Chi, Professor, Department of Psychology (now

Professor of Psychology in Education at Arizona State University)

EFFECTIVENESS OF HOLISTIC MENTAL MODEL CONFRONTATION IN

DRIVING CONCEPTUAL CHANGE

Soniya Gadgil, M.S.

University of Pittsburgh, 2009

Students' flawed conceptions can often be corrected during learning by successive contradictions from text sentences, allowing them to revise or replace their individual false beliefs as they read a text. Although this process supports learning new knowledge at a local level (i.e., individual propositions) it is less effective in facilitating systemic conceptual change. In contrast, constructive processes such as self-explaining can facilitate conceptual change through building a mental model from cumulative revisions of individual false beliefs. In the current experiment, I investigated whether comparing and contrasting examples can achieve the same outcome. Students (n=22) in the compare group were first shown a diagram of with their own flawed mental model, and then asked to compare it with a diagram of the correct model. This condition was compared with self-explaining the correct diagram (n=22), and a control condition in which students simply read the text twice (n=20). Results showed that the compare group performed better than the other two groups on questions requiring deep inference making.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

I would like to thank Jenna Small (UVA) and Kristen Sawl for helping with the data collection and coding data respectively. I would also like to thank former members of the Chi lab, Dr. Kirsten Butcher and Dr. Marguerite Roy in all stages of executing this project. This project would have been impossible had it not been for the invaluable guidance I received from my former advisor Dr. Michelene T.H Chi and my current advisor Dr. Timothy Nokes. Finally, I am thankful for my husband Vinod who encouraged me throughout this journey, and did not lose faith in me even when the odds seemed insurmountable.

1.0   INTRODUCTION

In science learning, students often come into a learning situation with misconceived prior knowledge. These misconceptions can exist at three levels, viz. false beliefs, flawed mental models and incorrect ontological categories. Chi (2008) has argued that misconceptions at the false belief level can be rectified by direct instruction in the correct beliefs, whereas mental model revision requires a more systemic level of conceptual change. The current research examines what cognitive processes facilitate this kind of change. It is hypothesized that systemic conceptual change requires revision at the holistic relational level as opposed to the individual propositional level. To test this, I conducted an experiment that compares an instructional intervention designed to facilitate holistic confrontation (comparing and contrasting a representation of one's misconception to an expert representation) vs. self-explanation of the expert representation vs. traditional instruction (i.e., reading a text). If conceptual change is driven by system-wide knowledge revision, then instruction aimed at holistic confrontation should facilitate more robust conceptual learning than those processes aimed at individual propositions.

To situate the current work I first describe three levels of knowledge representation, the cognitive processes supporting conceptual change, and instructional interventions designed to facilitate those processes.

1

1.1   KNOWLEDGE REPRESENTATION

When knowledge is misconceived at the false beliefs level, students have certain incorrect individual beliefs about what he or she is about to learn. A *false belief* is a single incorrect idea that can usually be stated in a single proposition. For example, if a student thinks, "*all* blood vessels have valves", this is in direct contradiction to the correct proposition "only veins have valves". When false beliefs are in direct contradiction to the correct proposition, direct refutation is usually sufficient to overcome this barrier to learning (Dole, 2000).

The next level of misconceived knowledge is of flawed mental models. A *mental model* is a representation constructed by a learner that allows him or her to make inferences and reason qualitatively about a process or system (Gentner & Stevens, 1983; Johnson-Laird, 1980). It has a similar relation-structure to that of the process it represents. A mental model is more than simply a set of propositions, as it has complex inter-relationships that may not always be captured by the way of propositions.

Even when students have a mental model that is flawed, their understanding is often seemingly coherent, and the student can justify it quite successfully. For example, young children often have a flawed mental model of the earth as a hollow sphere. When asked questions like "Where do people live?", they say "On flat ground deep inside the hollow sphere." (Vosniadou, 1994). Such a flawed mental model needs extensive transformation, in order for the children to learn the new and correct model.

The third and most complex level of incorrect knowledge is at the level of ontology, when students misattribute a process to one kind of category versus another (Chi, 1997, 2005; Chi, & Roscoe, 2002). If misconceptions belong to one category and correct conceptions belong to another lateral or ontological category, then they conflict by definition of *kind* and/or

ontology. This means that conceptual change requires a shift across lateral or ontological categories. For example, students have been shown to possess the misconception that "hot objects contain heat". Thus, they misattribute heat as being a thing or an object that can be contained. Instruction to correct such misconceptions should first help students determine that heat does not belong in the category of an object or a direct causal process, but is an emergent phenomenon (Chi, 2005). Such misconceptions tend to be the most robust, and require systemwide conceptual change.

It is important to distinguish conceptual change from other kinds of learning processes such as addition of new knowledge and gap filling (Chi, 2008). *Addition of new knowledge* occurs when one has no prior knowledge about the topic whatsoever. For example, if a person has no prior knowledge of the human heart and circulatory system, all that the person learns about this topic will be represented as the addition of new knowledge components. *Gap filling* takes place when a learner has certain disjoint ideas about the material to be learned, with gaps that need to be filled in order to learn. For example, a person may know that the human heart has four chambers, but not know the names of the four chambers. In this case, learning involves simply gap filling. This kind of learning has also been termed *enrichment* (Carey, 1991) or *accretion* (Rumelhart, 1976). Gap filling and addition of new knowledge do not constitute conceptual change, as the learner does not need to transform his or her existing knowledge structure or deliberately get rid of misconceptions in order to learn. By this definition of conceptual change, only incorrect knowledge at the level of mental models or ontology is remedied by conceptual change. The first level of false beliefs is usually remedied by belief-revision. In the current work, we will focus on conceptual change only at the level of mental models.

1.2   COGNITIVE PROCESSES SUPPORTING CONCEPTUAL CHANGE

Self-explanation has been shown to be very effective in achieving conceptual change (Chi, de Leeuw, Chiu, & LaVancher, 1994). The self-explanation effect has been found to be robust across domains (Mwangi & Sweller, 1998) age-groups (Calin-Jageman, & Horn Ratner, 2005), and tasks (Pirolli & Recker, 1994; Aleven & Koedinger, 2002). The typical method used in self-explanation studies is to have students go through a text line by line, forcing them to reconcile their existing flawed mental model with the correct one, resulting in the acquisition of the correct beliefs, which are hypothesized to culminate in the construction of a correct mental model. However, self-explanation need not always be in the form of students explaining aloud from texts. Other forms of self-explanation such as explaining with diagrams (Ainsworth, 2003) and sketching explanatory diagrams (Van Meter, 2006) have also been explored and shown to be equally beneficial.

Chi (1994) has shown that when students were prompted to self-explain as they read a text, 5 out of 8 students could successfully transform their flawed prior model to acquire the correct model. However, the self-explanation was targeted at the belief revision level, while the misconceptions are at the mental model level. Students have been shown to have several correct beliefs, yet a flawed mental model. It is possible that the students who did not acquire the correct model could not transform their existing model by accumulation of beliefs. In this experiment, we want to see whether directing instruction at the holistic (relational) level, by means of comparing and contrasting one's flawed mental with the correct expert model, is more likely to result in a transformed mental model.

Similarly, past research has shown that refutation can be successful in effecting belief revision. In a refutation text, common misconceptions held by students are identified and

explicitly addressed. For example, Diakidoy (2003) gave sixth-grade students a standard science text about energy or another text that explicitly addressed two common misconceptions that students held about energy and refuted them. Students who read the refutation text performed better than those who read the simple expository text. However, the accretion of correct beliefs may not always add up to a correct understanding when the misconceptions are at the mental model level, and indeed some studies have reported no advantage of refutation texts over expository text (for example, Alvermann & Hynd, 1987; Maria & Blustein, 1986). A meta-analysis by Guzzetti et al (1993) showed that other strategies such as group discussion could be used to introduce cognitive conflict, and were as effective in facilitating belief revision as a refutational text. These strategies were "activation activities" such as group discussion etc. This means constructive activities rather than direct refutation are also effective in facilitating conceptual change.

A key difference between self-explanation and learning from refutational texts is that in a refutation text students' misconceptions are identified in advance and explanations are provided, whereas in self-explanation students' generate those explanations for themselves. It also provides students an opportunity to notice gaps in their own knowledge. Research in reading comprehension has shown that students are frequently not aware of their misconceptions (Kendeou & Van Den Broek, 2005). Moreover, students may recognize individual misconceptions while self-explaining, but not realize that their mental model is flawed. We hypothesize that these gaps can be made more salient by means of holistic confrontation. That is, by comparing one's incorrect mental model to a correct one provides an opportunity to notice holistic, system-wide discrepancies and their conceptual implications. One way to facilitate

noticing such systems of relations is through *analogical reasoning*, specifically by comparing and contrasting examples.

Past research in analogical comparison has shown it to be an especially powerful learning mechanism. For example, Gick and Holyoak (1983) found that when participants' schemas contained more relevant structural features, they were more likely to solve a similar problem. It seems reasonable that this can apply to conceptual learning as well. Medin, Goldstone, and Gentner (1993), found that interpretation of an ambiguous stimulus was dependent on the properties borrowed from unambiguous stimuli that it was compared to. In other words, the process of comparison encouraged students to focus on certain features over others. Therefore, we hypothesized that when students are asked to compare a flawed model to a correct model they will be more likely to focus on the relevant structural features.

Kurtz, Miao, & Gentner (2001) showed that students who were learning about the concept of heat transfer learned more when comparing examples than when studying each example separately. By comparing the commonalities between two examples, students could focus on the causal structure and improve their learning about the concept. In Chi (1994)'s work on self-explanation, induction of function or understanding the causal mechanisms of circulation was an important mediator of learning. If analogical comparison can promote focusing on the relevant features, then it could lead to more causal inference generation, compared to students seeing only the correct model.

The process of comparison can encourage inference generation by providing scaffolding for explanation. Coleman (1998) has argued that encouraging students to generate explanations should help them to monitor themselves and others for comprehension, access and use prior knowledge, gain awareness of their knowledge, and thus improve their conceptual

understanding. Students learned about photosynthesis in a collaborative situation using explanation prompts that asked students to explain and justify their answers, compare past and current beliefs, think about nature of their explanation (e.g. scientific vs. everyday), etc. In individual and collaborative post-test measures (concept mapping and problem-solving) completed after the learning unit, it was shown that prompted students, in contrast to unprompted students, acquired a more correct understanding of photosynthesis.

Similarly, Schwartz and Bransford (1998) conducted an experiment in which undergraduate students studying cognitive psychology analyzed contrasting cases for one set of memory concepts, or simply read about cases for another set of concepts. In the contrasting cases condition, students graphed the data themselves, whereas for in the read-only condition, they were provided with the graphs instead of the raw data. All students later attended a lecture on the same topic. Students who analyzed and contrasted cases illustrating memory phenomena showed better performance on a transfer task compared to controls who simply read about the cases. Thus, analogical comparison led to better preparation for future learning from the lecture, by facilitating learning of critical aspects of the domain and noticing critical patterns that they could use to organize the concepts given in the lecture.

There is also some evidence in the problem-solving literature that students are able to generate inferences by contrasting cases. For example, Chang (2006) showed that having students compare statistics problems that differed either on the surface level or structural level based on their prior knowledge, it increased their performance on problem-solving tasks. Similarly, Rittle-Johnson and Star (2007) gave seventh grade algebra students two analogous algebra problem solutions to compare and contrast, with the control group receiving just two

worked examples sequentially. Students in the compare and contrast condition outperformed the control group on both procedural as well as conceptual knowledge.

These results suggest that analogical comparison can promote the cognitive processes such as encoding of critical features, inference generation, and schema extraction. The key difference between past research on contrasting cases and the current work is that previous work focused on giving students two correct examples to compare whereas in the current work students compared their *own flawed model* to the expert model. We hypothesized that this kind of comparison would facilitate inference generation, noticing of critical conceptual (relational) features, extraction of a schema, and provide an opportunity for system-wide change at the mental model level (noticing and correcting errors). We compare this instructional intervention to self-explaining the expert diagram and a control condition involving only reading an instructional text.

2.0   METHOD

2.1   PARTICIPANTS

Eighty-one undergraduate and graduate students from University of Pittsburgh were recruited for the study. Sixty-four of them qualified to participate, based on a test of their prior knowledge. The mean age of participants was 21.67 y, $sd = 2.07y$. Students' self-reported their GPAs; $M = 3.29$, $sd = 0.41$. All students were native or fluent English speakers who had not studied the heart and the circulatory system before. They had no college-level courses in Biology. Participants were either paid \$15 for the 2h study, or were part of a Psychology subject pool, wherein they participated in experiments for course credit. The paid and for-credit students were equally distributed across the three conditions.

2.2   DESIGN

The design was a between subjects design with the participants being randomly assigned to one of three instructional conditions (compare ($n = 22$), explain-diagram (n = 22), read-twice ($n = 20$)). There were no differences in mean age ($F(2, 61) = 1.60$, ns) or GPA ($F(2, 61) = 0.76$, ns) across conditions. Students who participated for credit and paid participants were equally divided across conditions.

2.3 MATERIALS AND MEASURES

The materials consisted of a flawed (single-loop) and correct (double-loop) diagram of the circulatory system, and an expository text that described the structure, function, and path of blood flow. The flawed model was selected on the basis of prior research, which has identified several flawed models of the circulatory system that students tend to have (Chi, 2001). Initial piloting showed that college students are most likely to have a single-loop model of the circulatory system. Students in the compare condition were shown a diagram of their own flawed mental model (see Appendix A) and a diagram of correct double-loop model (see Appendix B). Students in the explain-diagram condition were shown only the correct double-loop model. Students in the read-twice condition were not shown the diagram.

The text about the heart and circulatory system was common to all three conditions and was taken from a popular college-level Biology textbook (Shier, Butler, & Lewis, 2006; see Appendix C). The text was 72-lines long, and the diagrams were not available to the students when they read the text.

2.3.1 Screening Test

The *mental model questions* in the screening test were designed such that it would be possible to gauge the student's current mental model. These questions consisted of ten questions; three questions were multiple-choice for one point each, and seven were short answer questions ranging from one to four points (see Table 1 a for an example each type). The short answer

11

questions required students to describe their current mental model. The total possible score was 25 points.

If a student were found to have a flawed mental model other than the single-loop model, he or she would be excluded from the experiment. The inclusion criterion was that a student had to score lower than 10 on the screening test in order to participate in the experiment. The screening test was question based instead of drawing a blood path diagram, because we wanted to minimize the chance that the students in the explain-diagram condition would covertly compare their model to the correct model by making their current model explicit. The inter-item reliability co-efficient for this test (Cronbach's alpha) was 0.59.

2.3.2 Definitions Test

The definitions test consisted of *declarative questions* in which students were asked to define terms related to the circulatory system (see Table 1b for examples). This assessment was used to assess their declarative understanding of these terms. To help students define terms students were given generic prompts such as "Where is it located?", "What is it's function etc." This test was administered on a computer using simple word-processing software. It was administered at both pre and post-test to provide an assessment of declarative knowledge gains from the instructional intervention. There were twelve items on this test and the total possible score on this test was 45 points. The inter-item reliability co-efficient for this test (Cronbach's alpha) was 0.87.

2.3.3   Knowledge-Inference Test

The *knowledge inference questions* required deep reasoning on part of the students. The questions on this test were not directly addressed in the text or diagrams, and required students to integrate knowledge across text sentences, and also make connections with their prior knowledge in order to answer them correctly. For example, the text did not address function information about all parts related to the circulatory system, and students needed to make causal inferences in order to impute their function. These were administered only at post-test to avoid influencing the students' thinking during learning.

The knowledge inference questions were divided in two types of questions including integration inference and system-level questions (see Table 1c for the two types of questions). The *integration inference questions* were those that required inferences and integration across text sentences. These questions correspond to Category 2[1] questions from the Chi (2001) study. Out of a total of eighteen inference questions, eight were integration inference questions. The *system-level questions* required students to infer important new knowledge that could only be answered correctly if the students had a correct mental model. These questions corresponded to Category 3 and Category 4 questions from the Chi (2001) study. There were ten system-level questions. The total possible score on this test was 25 points.

The inter-item reliability co-efficient for this test (Cronbach's alpha) was 0.77.

---

[1] Category 1 questions from the Chi (2001) study were not included because the definition and mental-model questions tap the same information as do the Category 1 questions in the Chi (2001) study.

Table 1. Examples of Learning Measures

Question Type	Example
a.) Mental-model based questions	Describe the path of blood from the heart to the various parts of the body.
b.) Declarative Questions	Definitions of terms such as aorta, ventricle, septum.

c.) Knowledge Inference Questions:
    • Integration Questions
 	
    • System-level Questions
	"Why are arteries more flexible and strong compared to veins? Explain the structure of arteries that facilitates this."
	"Why don't we have valves in pulmonary veins?"

2.4   PROCEDURE

The procedure consisted of the screening test, followed by the pre-test, the instructional intervention, reading the text, and a post-test. Each participant first took the screening test. There were two purposes for conducting the screening test. The first was, to confirm that students had low prior knowledge about the heart and circulatory system. The second purpose was to gauge the student's current mental model. A pilot study revealed that most college students have a single-loop model of the heart. In a single-loop model, students incorrectly believe that blood gets oxygenated in the heart, and flows out to different parts of the body to distribute oxygen, as opposed to the correct double-loop model, in which blood flows in two distinct loops: heart-lungs-heart, and heart-body-heart.

The experimenter then scored the screening test. Only students who had a single-loop model and scored fewer than 10 out of 25 points went on to participate in the experiment. After the screening test, students took a pretest on declarative knowledge described under Materials.

Students in the *compare condition* were shown a diagram of a flawed model of the circulatory system and were instructed to compare it to the expert diagram. The experimenter first explained the flawed diagram to the participant, and asked whether he or she agreed with it. Once they agreed with it, they were told that this was not the correct model, and were shown a diagram of the correct model. They were then asked to make comparisons between the two diagrams. The experimenter guided the comparison process using prompts such as *"Can you trace the path of blood flow in each of the diagrams?" "What are the important parts of the circulatory system based on each of the diagrams?"* Students were instructed to talk aloud during the entire process (see Appendix D for a complete list of prompts).

Students in the *explain-diagram condition* were only shown the correct diagram and asked to self-explain it. Again, they were given similar prompts, parallel to those in the compare condition, for example, *"Can you trace the path of blood flow in the diagram?"; "What are the important components of the circulatory system based on this diagram?"* etc. (see Appendix D for a complete list of prompts).

Audio protocols were recorded for both conditions. The time spent self-explaining / comparing was limited to approximately 5-7 minutes in both conditions.

Students in both conditions then read the circulatory-system text presented one line at a time on the computer screen. Diagrams were not available to the students as they read. Students in the read-twice condition simply read the text twice and did not compare or explain diagrams.

This was followed by a post-test that consisted of drawing the blood path diagrams, and

answering questions. The questions at post-test consisted of the same mental model questions that they had taken at screening, declarative questions they had answered at pretest, and two additional parts that were not present in the pretest and screening test: drawing the blood path diagram, and answering knowledge inference questions (described under Materials). Students were asked to draw a diagram showing their current understanding of the heart and circulatory system on a provided template. They were asked to explain aloud as they drew, and audio protocols were collected.

2.5   PREDICTIONS

On the mental model based questions, the compare condition was expected to outperform the explain-diagram condition. The process of comparing and contrasting the students' own model with the correct model would make the differences between the two more salient. When the student notices the discrepancy between her own model and the correct model, she is more likely to revise her existing model. On the other hand, while self-explaining the diagram of the correct model, a student may not notice the discrepancies, making it less likely for her to revise her existing model. Both the compare and explain-diagram group were expected to outperform the read-twice group.

Both groups will be exposed to the same declarative knowledge in the form of text that they read in all three conditions. Therefore, on the declarative knowledge, no differences between groups were expected.

On the knowledge inference questions, again compare group was expected to outperform the explain-diagram group. Studying the correct diagram in relation to the flawed one will invite inference generation.

3.0   RESULTS

The results are presented in two major sections, viz. learning outcomes and protocol analyses. Under learning outcomes, students' performance on the different tests and the drawing task is discussed. The protocol analyses were conducted to give an insight into the kinds of cognitive processes the students engaged in as they learned.

Alpha was set to .05 for all main effects and interactions, and Bonferroni corrections were used for all planned comparisons setting alpha to .01 (Keppel, 1991). Effect sizes were calculated for all main effects, interactions, and main comparisons. Cohen (1988) has suggested that effects be regarded as small when 0.2, as medium when 0.5, and large when 0.8.

3.1   LEARNING OUTCOMES

Pre-test performance. The graphs (Fig. 1 and Fig. 2) below show the learning outcomes on all the learning measures at pretest and post-test. Pretests measured the students' declarative understanding and mental model understanding. The post-test included a test on knowledge inference questions in addition to the declarative and mental model tests given at pretest.

An Analysis of Variance (ANOVA) was conducted on participants' pre-test scores across the three groups. The three groups were not significantly different on the declarative or mental questions, $F(2, 61) = 0.37$, *ns* and $F(2, 61) = 1.32$, *ns,* respectively. This shows that these (pre-

screened) participants came into the experiment with extremely limited knowledge of the domain and there were no significant differences across the groups in prior knowledge. Therefore, any differences observed on the post-test can be attributed to the experimental manipulation.
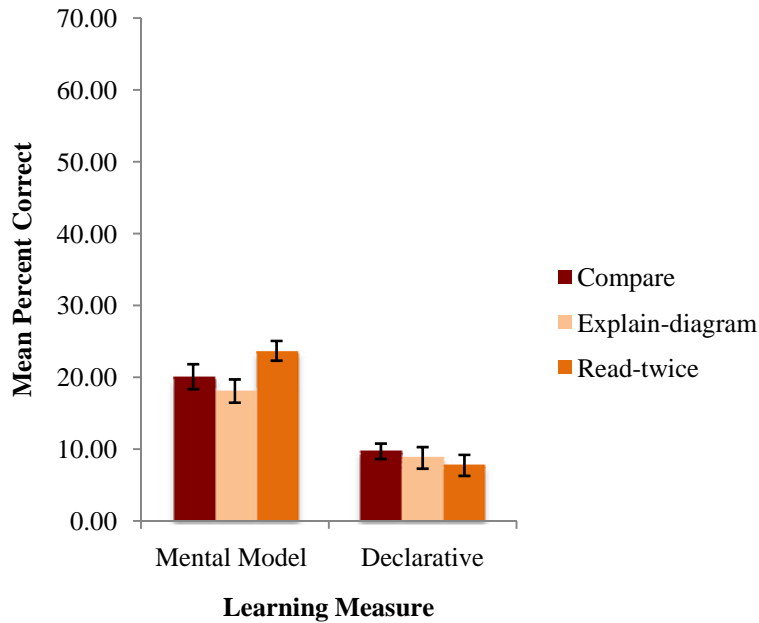

Figure 1. Average Pretest Scores on Declarative and Mental Model Questions
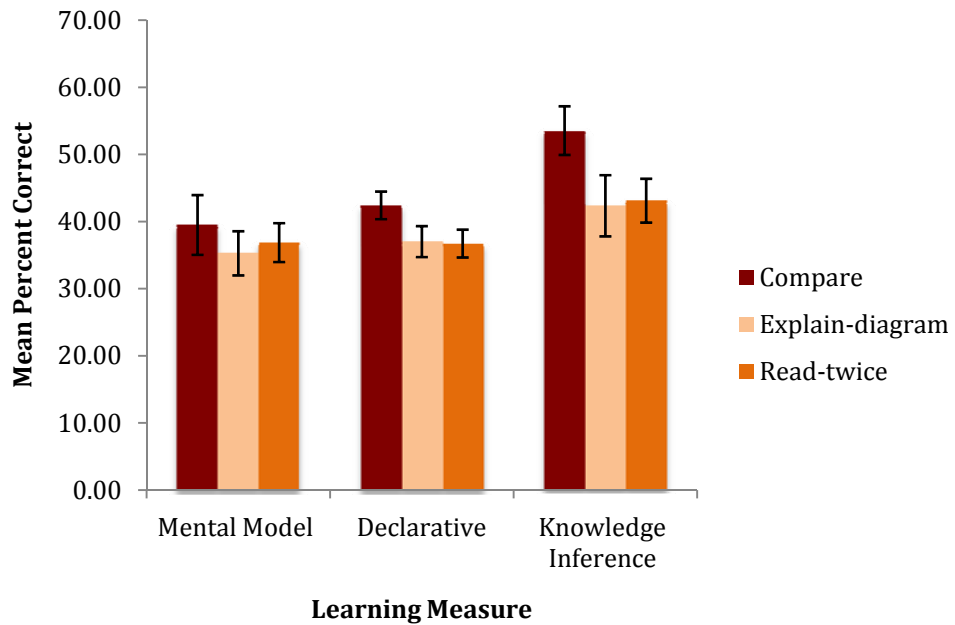

Figure 2. Average Posttest Scores on Declarative, Mental Model, and Inference Questions

Post-test performance. All three conditions showed significant learning gains from pretest to posttest on the declarative and mental model questions; $t\,(63) = 18.19$, $p < 0.05$, $t\,(63) = 20.51$, $p < 0.05$. However, the three conditions were not significantly different from each other in their learning on declarative ($F(2,\,61) = 0.76$, *ns*) and mental model questions ($F(2,\,61) = 0.99$, *ns*).

A one-way ANOVA was conducted on scores on the knowledge inference questions. The three groups were marginally different at $F(2,\,60)=2.62$, $p = 0.08$. To test our a-priori hypothesis that the compare condition will outperform the other two conditions, contrasts were performed. The compare condition significantly outperformed the explain-diagram condition; $t\,(61) = 2.061$, $p < 0.05$, $d = 0.52$; and marginally outperformed the read-twice condition; $t\,(61)= 1.87$, $p < 0.1$, $d = 0.61$. The read-only and explain correct diagram condition did not differ significantly from each other; $t\,(61) = -0.132$, *ns*.

3.2   ANALYSIS OF INFERENCE QUESTIONS BY QUESTION-TYPE

As described in the Materials section, there were two kinds of The knowledge inference questions, viz. *integration inference questions* that required inferences and integration across text sentences and *system-level questions* that required students to infer important new knowledge and could only be answered correctly if the students had a correct mental model. Below, I report a breakdown of students' performance by question type.

3.2.1 Integration Inference Questions

A one-way ANOVA was conducted to examine the effect of the intervention on integration inference questions. There was no significant difference between the three groups, $F(2, 61)=2.06$, *ns*.

3.2.2 System-level Questions

The compare group was expected to acquire a more robust model allowing students to perform especially better on system-level knowledge inference questions. A one-way ANOVA was conducted to examine the effect of the intervention on system-level questions. The three groups were marginally different, $F(2,61)=2.4$, $p < 0.1$. A contrast revealed that the compare group significantly outperformed the read-only group $t(61) = 2.12$, $p < 0.05$, $d = 0.42$. The explain-diagram and read-twice groups were not significantly different from each other, $t(61)=-0.62$, *ns.* The compare group and the explain-diagram group were also not significantly different, $t(61)= 1.53$, *ns* (see Figure 3).
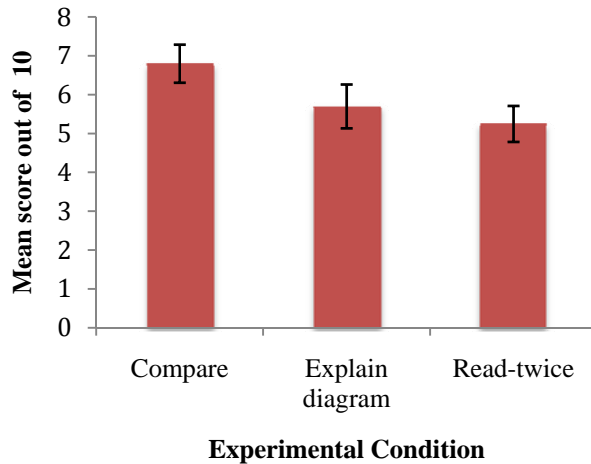

Figure 3. Performance on System-Level Knowledge Inference Questions by Condition

3.3 FINE-GRAINED ANALYSIS OF DECLARATIVE QUESTIONS

The three conditions did not differ from each other on declarative questions ($F(2, 61) = 0.761$, *ns*). However, each definition asked for different kinds of information, such as location, structure, function, and other related information. Therefore, each definition was broken down into these four components, and the students' score on each component was assessed (see Table 2)

Table 2. Examples of Fine-Grained Analysis of Declarative Questions

Attribute	Description
Location	Leaves the left ventricle
Structure	Large artery
Function	Takes oxygenated blood out of left ventricle
Other	Branches off into arteries and further arterioles; Composed of three layers; At its base is an aortic valve that prevents blood from flowing back to the heart.

Students' performance on any of these components was not found to be significantly different by condition.

3.4 ANALYSIS OF LEARNING TIME

The time spent on learning was differed by condition ($F(2, 61) = 6.59$, $p < .05$). Students in the compare condition spent an average of 16.62m studying the text, those in the explain-diagrams condition spent 17.14 m studying the text, and those in the read-twice condition spent 24.19m

reading the text. Students in the read-twice condition were told they could spent between 30 and 35 m reading the text twice, but ended up spending less time than that.

The total time spent by students for the entire experiment was also significantly different by condition. ($F(2, 61) = 5.129$, $p < 0.05$). Thus, students in the compare and explain-diagram conditions showed a significant efficiency in learning, compared to students in the read-twice condition.

3.5   DIAGRAM ANALYSIS

Only those students who had a single-loop diagram to begin with, were included the experiment. The final diagrams that students generated after the learning session were coded for correctness of mental models. They were classified according to the following categories:

1. **Single Loop:** Blood is primarily contained in blood vessels and pumped from the heart to the body. It returns to the heart by the way of same blood vessels.

2. **Single loop with lungs:** Blood is primarily contained in blood vessels; heart pumps blood from body to lungs; blood returns to heart from body or lungs; heart oxygenates blood.

3. **DL (partially correct):** Blood is primarily contained in blood vessels; heart pumps blood to body and lungs; lungs oxygenize the blood; oxygenated blood returns to body from lungs without returning to the heart; and comes back to the heart deoxygenated.

4. **Correct Double Loop:** Blood is primarily contained in blood vessels; heart pumps blood to lungs; lungs oxygenate blood; oxygenated blood returns to heart from lungs to be sent

to rest of the body; when blood gets deoxygenated, it returns to heart; heart divided into four chambers with septum separating the heart down the middle

5. **Other (unclassifiable)**

These categories were adapted from categories of mental models in Chi (1994). No students were found to have a purely single loop model after instruction. The following table shows the outcomes of diagram analysis:

Table 3. Performance of Students on Diagram-Drawing Task

	Compare	Explain- Diagram	Read-twice
SL	0	0	0
SL with lungs	0	1	1
DL (partially correct)	2	5	6
Correct double loop	19	15	13
Unclassifiable	1	1	0

Two raters blindly coded the diagrams for correctness of mental models. Inter-rater reliability was K=0.94. Eighty-six percent of students in the compare condition generated a correct diagram, as opposed to 68% in the explain-diagram condition and 65 % in the read-twice condition. A chi-square test showed no significant differences between conditions; $\chi^2(6, N = 64) = 0.532$, *ns.*

3.6 PROTOCOL ANALYSIS

Protocols generated by students in the compare and explain-diagram conditions were transcribed. Each protocol was between 35-50 lines. Protocols were analyzed to get an insight into the learning processes students engaged in, as they conducted these activities. Each protocol was be segmented at the grain size of idea unit.

Self-explaining has been shown to be a powerful learning activity. Examining the protocols was undertaken to help us understand whether the comparison group generated more or better explanations.

3.6.1 Total Number of Explanations

Self-explanations were operationalized as any statement that went beyond simply paraphrasing the information on the diagram. For example, statements in which students simply traced the path of blood were not counted as a self-explanation, because that information was explicit on the diagram. However, if students made inferences about the functions of parts (for example, "*looks like the septum keeps the oxygenated and deoxygenated blood from mixing*"), or elaborated something on the diagram by making connections to prior knowledge (for example, "*probably something flows through it, because, the word reminds me like septic tanks, simulates like a filtering or cleaning process*"), those statements were coded as self-explanations. Statements students made to monitor their understanding were also coded as self-explanations, and so were comparisons made by students in the compare condition. Comparison statements were coded as such when students explicitly compared the flawed and the correct diagrams.

Students generated an average of 10.07 self-explanations across conditions. The number of self-explanations was significantly different by condition. Students in the compare condition generated an average of 12.59 self-explanation statements per protocol, whereas students in the explain-diagram condition generated 7.55 statements. This difference was statistically significant $(F (1, 42) = 17.59, p < .05)$.

Students in only the compare condition were prompted to make explicit comparisons with their initial mental models. Students generated an average of 5.14 ($sd = 2.59$) comparison

statements per protocol. Protocols of students in the explain-diagram condition were also examined to see whether students make explicit comparisons even without prompting. However, no such evidence of comparison was found in the protocols. Students did make statements such as "I got this wrong on the test" which suggests some comparison process going on, however, the frequency of such comparisons was low (occurred only three times in all of the protocols).

3.6.2 Fine-Grained Coding of Verbal Protocols

A more fine-grained analysis of students' verbal utterances was undertaken using a coding scheme that was adapted from Renkl (1997) and adapted to suit the format of the experiment. For example, it did not include anticipative reasoning and principle-based explanation as categories, because they are more relevant to a problem-solving situation. A category of Comparison statements was added to address the comparison activities that students engaged in, in the experiment. The following categories were used to code the verbal protocols:

i. Function-related explanation: Students explanations were coded as functrion-related explanations if they made inferences about the function of a component while explaining. For example, *"the septum divides the two sides of the heart, so that the two don't ever mix."*

ii. Elaboration of situation: When students went beyond merely describing the flow of blood in the diagram, including metaphors and analogies. Renkl (1997) has described these statements as those indicating construction of a situation model. For example, *"probably something flows through it, because, the word reminds me like septic tanks, simulates like a filtering or cleaning process"*

iii. Revision statement: This category was modified from the category "Noticing coherence" from the Renkl (1997) study. This indicated when students explicitly changed a statement they had previously made, as they studied or compared diagrams.

iv. Monitoring negative: These were coded when students made statements to indicate that they were not understanding the material. Example of monitoring negative: *"I am not sure what deoxygenated means exactly."*

v. Monitoring positive:

These were coded when students made statements to indicate that they were understanding the material. Example of monitoring positive: *" I had no idea what the septum referred to, so now I know that it kind of divides"*

vi. Comparison statements: Statements were coded as comparisons statements if students were explicitly comparing the flawed diagram with the correct one. For example, *"Well this one has the different chambers in it instead of just one unified thing."*

Table 4. Results of Protocol Analyses

	Compare		Explain diagram	
	M	**SD**	**M**	**SD**
Function-related explanation	4.91	2.07	3.82	1.87
Elaboration of situation	1.05	1.17	1.32	1.21
Revision statement	0.45	0.67	1.14	1.16
Monitoring negative	0.68	0.89	0.82	1.00
Monitoring positive	0.36	0.72	0.45	0.73
Comparison statement	5.14	2.59	0.00	0.00

The first author coded all the protocols, and 25 % of the protocols were recoded by an independent coder, and reliability between the codings assessed. Reliable agreement was obtained between coders. ($K = 0.87$).

An ANOVA showed that only function-related explanations were marginally different by condition, with the compare condition generating more function-related statements than the explain-diagram *condition* $F(1,42) = 3.37$, $p < 0.1$. There was also a difference in the revision statements, with the explain diagram condition generating more number of revision statements than the compare condition $F(1,42) = 5.64$, $p < 0.05$.

Comparison statements. A significant correlation was found with the number of comparison statements made by students in the compare group and their scores on system-level knowledge inference questions; $r = 0.430$, $p < 0.05$.

3.6.3   Features of Expert Diagram Noticed

Features of the expert diagram that students were expected to explain as they studied the diagrams were identified, and the protocols were coded to determine how many features the students mentioned while explaining/comparing diagrams. This was to see whether students in the compare condition noticed more features than students in the explain-diagram condition. Students were able to state 7.9 statements on an average out of 12 for both groups.

There was no difference between conditions in terms of number of diagram features noticed by students.

4.0   DISCUSSION

Both, the compare condition and the explain-diagrams condition performed better than the read-twice condition, on knowledge inference questions. However, all three groups performed similarly on declarative questions. This was in line with our initial hypotheses that the comparison of the flawed diagram to the correct one will encourage inference-generation, which will in turn will improve performance on knowledge inference questions.

Compared to previous research, mental model transformation as measured by analysis of diagrams generated by students was more frequent in the compare condition, although the difference between conditions was not statistically significant. Note that the percentage of students who changed from single-loop to double-loop was 65% and 68 % in the read-twice and explain diagram conditions respectively. In previous research, (Chi, 1994) self-explanation was shown to be successful 66 % of the times, so the findings are comparable. However, students in the read-twice condition also performed at comparable levels, and therefore, we don't see a large self-explanation effect at the mental model level. One possible explanation for this could be that the topic was not sufficiently challenging for college students. The Chi (1994) study was conducted with eighth graders. Thus, further research may need to take into account the task demands, and test the paradigm with more challenging materials. Another potential explanation of high performance by read-twice group is that they may be doing covert self-explanation.

However, there was no way to assess this hypothesis because students were not asked to talk aloud as they studied the texts.

On the system-level knowledge inference questions, students in the compare condition performed significantly better than those in the read-twice condition. These were questions that required generation of new knowledge, and were not directly addressed in the text. Most system-level questions required students to induce the function of a component if only its structure or location was specified in the text. We saw that the compare condition produced more goal-directed statements, in which made inferences about the function of a component while explaining. Moreover, the number of comparisons generated while learning correlated with performance on knowledge inference questions. Thus, the process of comparison facilitated making goal-directed inferences, which improved performance on system-level knowledge inference questions.

Students in the read-twice condition did not see the diagrams as they learned, and performed worse than the conditions that saw a diagram. This replicates the results by Butcher (2006) that visual scaffolds aid learning in this domain. However, the comparison process adds something over and above the diagram effect. In Butcher, the students did poorly on inference questions. In the present study, although all groups did better on inference questions compared to students in Butcher (2006), the students in the compare group performed significantly better on the inference questions compared to students in the other two conditions.

Previous work on explaining with diagrams by Ainsworth and Louizou (2003) has shown that self-explaining a diagram led to better learning outcomes compared to self-explaining a text with the same content. The present experiment goes beyond this finding, and describes how learning from diagrams can be optimized based on students' mental models.

The results of this research provide evidence that comparing an expert model to one's own flawed model can be as effective as self-explanation in driving conceptual change. Further research should explore whether the effects of comparison are sustained even when the read-twice condition is given a diagram. That way, we can tease apart how much of the effect can be attributed to the visual representation itself, and how much to the constructive process of comparing and contrasting. Future research can also explore whether this process of comparing one's own flawed model to a correct model is helpful in the learning of problem-solving skills.

Comparing and contrasting can have immense applications in the classroom. Misconceptions are common in science learning, and comparison and contrasting can be an effective learning technique in achieving conceptual change. Future work also needs to address whether the process of comparing and contrasting is especially beneficial for low-ability explainers. This will help us better understand how to tailor instruction according to individual differences.
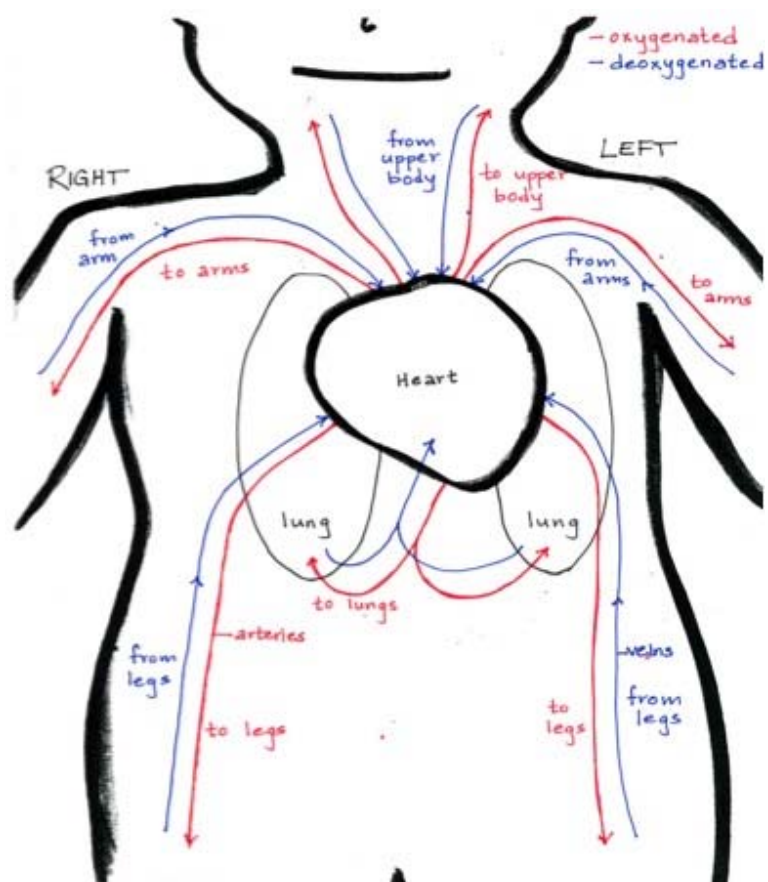
APPENDIX A

SINGLE-LOOP MODEL (FLAWED)

Figure 4. Single-loop model (flawed)

APPENDIX B

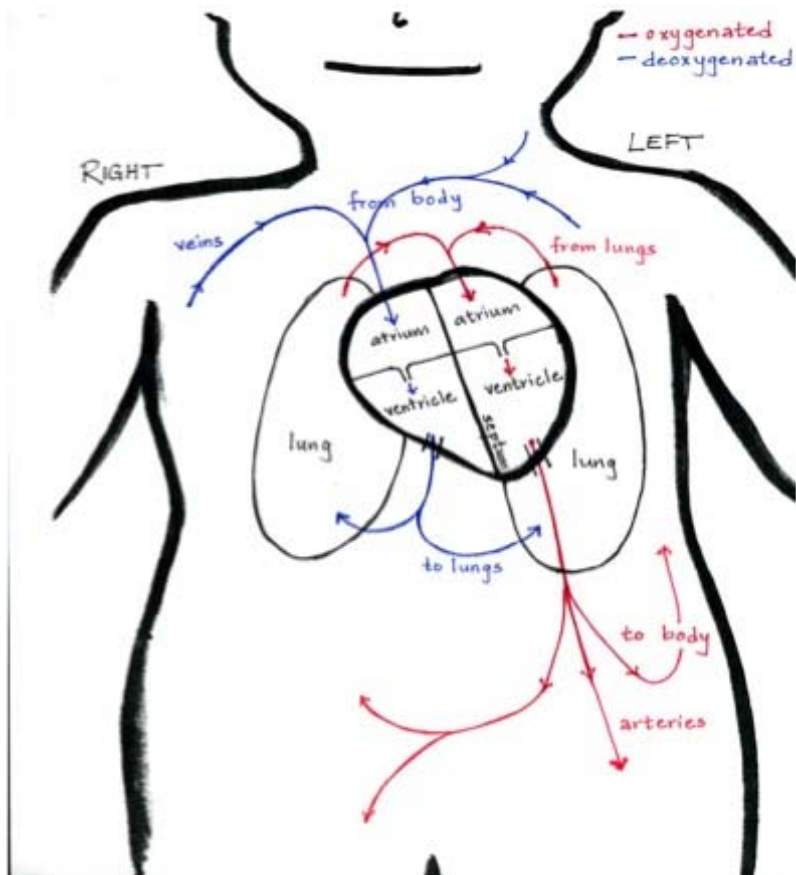## DOUBLE-LOOP MODEL (CORRECT)

Figure 5. Double-loop model (Correct)

APPENDIX C

EXPERIMENTAL TEXT (SHIER ET. AL., 2006)

1. The heart is a hollow, cone-shaped, muscular pump.
2. The heart pumps 7,000 liters of blood through the body each day, contracting some 2.5 billion times in an average lifetime.
3. An average adult's heart is about 14 cm long and 9cm wide.
4. It lies within the thoracic cavity and rests on the diaphragm.
5. The pericardium encircles the heart
6. Between the layers of pericardium is a space, the pericardial cavity that contains a small volume of serous fluid.
7. This fluid reduces friction between the pericardial membranes as the heart moves within them.
8. Internally, the heart is divided into four hollow chambers- two on the left and two on the right.
9. The upper chambers, called atria, have thin walls and receive blood returning to the heart.
10. The lower chambers, the ventricles, receive blood from the atria and contract to force blood out of the heart into arteries.
11. A solid wall like septum separates the atrium and ventricle on the right side from their counterparts on the left.
12. The right atrium receives blood from two large veins, the superior vena cava and the inferior vena cava.
13. The large tricuspid valve, which has three tapered projections called cusps, lies between the right atrium and the right ventricle.
14. The valve permits blood to move from the right atrium into the right ventricle and prevents backflow.
15. When the muscular wall of the right ventricle contracts, the blood inside its chamber is put under increasing pressure and the tricuspid valve closes passively.
16. As a result, the only exit for the blood is through the pulmonary trunk, which divides to form the left and right pulmonary arteries that lead to the lungs.
17. At the base of this trunk is a pulmonary valve with three cusps that allows blood to leave the right ventricle and prevents backflow.

18. The left atrium receives blood from the lungs through four pulmonary veins, two from the right lung and two from the left lung.
19. Blood passes from the left atrium into the left ventricle through the bicuspid valve.
20. When the left ventricle contracts the bicuspid valve closes passively and the only exit is through a large artery, the aorta.
21. At the base of the aorta is the aortic valve, which opens and allows blood to leave the left ventricle.
22. The bicuspid and tricuspid valves are called atrioventricular valves because they are between the atria and ventricles.
23. Blood that is low in oxygen and high in carbon dioxide enters the right atrium.
24. As the right atrial wall contracts, the blood passes through the tricuspid valve and enters the chamber of the right ventricle.
25. When the right ventricular wall contracts, the tricuspid valve closes, and blood moves through the pulmonary valve and into the pulmonary trunk and pulmonary arteries.
26. From the pulmonary arteries, blood enters the capillaries associated with the microscopic air sacs of the lungs (alveoli)
27. Gas exchanges occur between the blood in the capillaries and the air in the alveoli.
28. The freshly oxygenated blood returns to the heart through the pulmonary veins that lead to the left atrium
29. The left atrial wall contracts and blood moves through the bicuspid valve and into the chamber of the left ventricle.
30. When the left ventricular wall contracts, the bicuspid valve closes and blood moves through the aortic valve and into the aorta and its branches.
31. A heartbeat heard through a stethoscope sounds like "lubb-dupp".
32. The first part of a heard sound (lubb) occurs during ventricular contraction, when the atrioventricular valves are closing.
33. The second part (dupp) occurs during ventricular relaxation, when the pulmonary and aortic valves are closing.
34. The blood vessels form a closed circuit of tubes that carries blood from the heart to the cells, and back again.
35. These vessels include arteries, arterioles, capillaries, venules, and veins.
36. Arteries are strong elastic vessels that are adapted for carrying blood away from the heart under high pressure.
37. These vessels subdivide into progressively thinner tubes and eventually give rise to finer, branched arterioles.
38. The wall of an artery consists of three distinct layers.
39. The innermost layer is composed of a simple squamous epithelium, called endothelium, which rests on a connective tissue membrane that is rich in elastic and collagenous fibers.
40. The middle layer makes up the bulk of the arterial wall.
41. It includes smooth muscle fibers, which encircle the tube and a thick layer of elastic connective tissue.
42. The outer layer is relatively thin and chiefly consists of connective tissue with irregularly organized elastic and collagenous fibers.
43. This layer attaches the artery to the surrounding tissue.
44. Capillaries, the smallest diameter blood vessels, connect the smallest arterioles and the smallest venules.

45. Capillaries are extensions of the inner linings of arterioles in that their walls are composed of endothelium.
46. Theses thin walls form the semipermeable layer through which substances in the blood are exchanged for substances in the tissue fluid surrounding body cells.
47. The substances exchanged move through capillary walls through diffusion, filtration, and osmosis.
48. Venules are the microscopic vessels that continue from the capillaries and merge to form the veins.
49. The veins, which carry blood back to the atria, follow pathways that roughly parallel those of the arteries.
50. Blood pressure decreases as blood moves through the arterial system and into the capillary networks, so little pressure remains at the venular ends of capillaries.
51. Instead, blood flow through the venous system is only partly the direct result of heart action and depends on other factors, such as skeletal muscle contraction and breathing movements.
52. Contracting skeletal muscles press on nearby vessels, squeezing the blood inside.
53. As skeletal muscles press on veins with valves, some blood moves from one valve section to another.
54. Respiratory movements also move venous blood
55. During inspiration, the pressure on the thoracic cavity is reduced as the diaphragm contracts and the rib cage moves upward and outward.
56. At the same time, the pressure within the abdominal cavity is increased as the diaphragm presses down on the abdominal viscera.
57. Consequently, blood is squeezed out of abdominal veins into thoracic veins.
58. During exercise, these respiratory movements act with skeletal muscle contractions to increase the return of venous blood to the heart.
59. Blood vessels can be divided into two major pathways.
60. The pulmonary circuit consists of vessels that carry blood from the heart to the lungs and back again.
61. The systemic circuit carries blood from the heart to all other parts of the body and back again.
62. Blood enters the pulmonary circuit as it leaves the right ventricle through the pulmonary trunk
63. The pulmonary trunk extends upward from the heart.
64. The pulmonary trunk divides into the right and left pulmonary arteries, which penetrate the right and left lung.
65. After repeated divisions, the pulmonary arteries give rise to arterioles that continue into the capillary networks associated with the walls of the alveoli, where gas is exchanged between blood and air.
66. From the pulmonary capillaries, blood enters the venules, which merge to form small veins, which merge to form larger veins.
67. Four pulmonary veins, two form each lung, return blood to the left atrium, which completes the pulmonary look.
68. Freshly oxygenated blood moves from the left atrium to the left ventricle.

69. Contraction of the left ventricle forces the blood into the systemic circuit, which includes the aorta and its branches that lead to all the body tissues, as well as the companion system of veins that returns blood to the right atrium.
70. Blood signifies life, and for good reason, it has many vital functions.
71. This complex mix of cells, cell fragments, and dissolved biochemicals transports nutrients, wastes, oxygen, and hormones; helps maintain the stability of interstitial fluids, and distributes heat.
72. The blood, heart, and blood vessels form the cardiovascular system and link the body's internal and external environment.

APPENDIX D

List of Prompts for Compare Condition

1. "What similarities and differences can you see in the two diagrams?"

2. "Can you trace the path of the blood as it travels through the body?"

3. "Can you explain to me why this one will not work?"

4. "Take a look at the different parts shown on the correct diagram and the incorrect diagram. Can you say something about the functions of the different parts?"

5. "What is the most important thing that strikes you that is different about this diagram?"

6. "Based on both these diagrams, what would you say are the major components of the circulatory system?"

List of Prompts for Explain-Diagram Condition

1. "What do you see happening in this diagram?"

2. "Can you trace the path of the blood as it travels through the body?"

3. "Can you explain to me how this works?"

4. "Can you say something about the functions of the different parts?"

5. "What is the most important thing that strikes about this diagram?"

6. "Based on this diagram, what would you say are the major components of the circulatory system?"

BIBLIOGRAPHY

Ainsworth, S., & Loizou, A. T. (2003). The effects of self-explaining when learning with text or diagrams. *Cognitive Science, 27*, 669-681.

Aleven, V., & Koedinger, K. R. (2002). An effective metacognitive strategy: Learning by doing and explaining with a computer-based Cognitive Tutor. *Cognitive Science, 26*(2), 147-179.

Butcher, K. R. (2006). Learning From Text With Diagrams: Promoting Mental Model Development and Inference Generation. *Journal of Educational Psychology, 98*(1), 182-197.

Bielaczyc, K., Pirolli, P. L., & Brown, A. L. (1995). Training in self-explanation and self-regulation strategies: Investigating the effects of knowledge acquisition activities on problem solving. *Cognition & Instruction, 13*(2), 221-252.

Calin-Jageman, R. J., & Horn Ratner, H. (2005). The role of encoding in the self-explanation effect. *Cognition & Instruction, 23*(4), 523-543.

Chang, N (2006). Learning to Discriminate and Generalize through Problem Comparisons. (Doctoral dissertation, Carnegie Mellon university, 2006).

Chi, M.T.H., de Leeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*(3), 439-477.

Chi, M.T.H. (1997). Creativity: Shifting across ontological categories flexibly. In T.B. Ward, S.M. Smith, & J. Vaid (Eds.), *Conceptual Structures and processes: Emergence, Discovery and Change.* (Pp. 209-234). Washington, D.C: American Psychological Association.

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science 25*(4), 471-533.

Chi, M. T. H. (2005). Commonsense Conceptions of Emergent Processes: Why Some Misconceptions Are Robust. *Journal of the Learning Sciences, 14*(2), 161-199.

Chi, M.T.H (2008) Three Kinds Of Conceptual Change: Belief Revision, Mental Model Transformation, And Ontological Shift. *In S., Vosniadou, A., Baltas & X., Vamvakoussi, (Eds.), Re-Framing the Conceptual Change Approach in Learning and Instruction. Advances in Learning and Instruction Series, Elsevier Press.*

Coleman, E. B. (1998). Using Explanatory Knowledge During Collaborative Problem Solving in Science. *Journal of the Learning Sciences, 7*(3/4), 387.

Diakidoy, I.A.N., Kendeou, P., & Ioannides, C. (2003). Reading about energy: The effects of text structure in science learning and conceptual change. *Contemporary Educational Psychology, 28*(3), 335-356.

Dole, J. A. (2000). Readers,Texts, and Conceprual Change Learning. *Reading and Writing Quarterly, 16*(2), 99-118.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7*(2), 155-170.

Gerjets, P., Scheiter, K., & Catrambone, R. (2006). Can learning from molar and modular worked examples be enhanced by providing instructional explanations and prompting self-explanations? *Learning and Instruction 16*, 104-121.

Gick, M., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*, 1.

Grosse, C. S., & Renkl, A. (2006). Effects of multiple solution methods in mathematics learning. *Learning and Instruction, 16*(2), 122-138.

Johnson-Laird, P.N. (1980) Mental models in cognitive science. *Cognitive Science*, 4, 71, 115.

Keppel, G. (1991). *Design and analysis: A researcher's handbook . Engelwood Cliffs*.

Kurtz, K. J., Miao, C.-H., & Gentner, D. (2001). Learning by Analogical Bootstrapping. *The Journal of the Learning Sciences, 10*(4), 417-446.

Medin, D.L., Goldstone, R.L., & Gentner, D. (1993). Respects for similarity. Psychological Review, 100, 254-278.

Mwangi, W., & Sweller, J. (1998). Learning to Solve Compare Word Problems: The Effect of Example Format and Generating Self-Explanations. *Cognition & Instruction, 16*(2), 173-199.

Pirolli, P., & Recker, M. (1994). Learning strategies and transfer in the domain of programming. *Cognition and Instruction, 12(3)*, 235.

Renkl, A. (1997). Learning from worked-out examples: A study on individual differences. *Cognitive Science, 21*(1), 1-29.

Rittle-Johnson, B., Star, J.R. (2007). Does Comparing Solution Methods Facilitate Conceptual and Procedural Knowledge? An Experimental Study on Learning to Solve Equations. *Journal of Educational Psychology 99(3)* 561-574

Rumelhart, D. E., & Norman, D. A. (1978). Accretion, tuning and restructuring: three modes of learning. In J. W. Cotton & R. Klatzky (Eds.), *Semantic factors in cognition*.

Schwartz, D. L., & Bransford, J. D. (1998). A Time For Telling. *Cognition & Instruction, 16*(4), 475-522.

Shier, D., Butler, J., & Lewis, R. (2006). *Hole's Essentials of Human Anatomy and Physiology*. Boston: McGraw Hill.

Vosniadou, S., & Brewer W. F. (1994). Mental models of the day/night cycle, *Cognitive Science 18(1),* 123-183.

Van Meter, P., Aleksic, M., Schwartz, A., & Garner, J. (2006). Learner-generated drawing as a strategy for learning from content area text. *Contemporary Educational Psychology, 31*, 142-166.