Learning and Instruction 22 (2012) 47–61

Contents lists available at ScienceDirect

Learning and Instruction

journal homepage: www.elsevier.com/locate/learninstruc

Effectiveness of holistic mental model confrontation in driving conceptual change

Soniya Gadgil [a,*], Timothy J. Nokes-Malach [a], Michelene T.H. Chi [b]

[a] *Department of Psychology, Learning Research and Development Center, University of Pittsburgh, United States*
[b] *Department of Psychology, Arizona State University, United States*

ARTICLE INFO

Article history:
Received 17 July 2010
Received in revised form
5 June 2011
Accepted 8 June 2011

Keywords:
Learning
Conceptual change
Comparison
Mental models
Self-explanation

ABSTRACT

Prior research on conceptual change has identified multiple kinds of misconceptions at different levels of representational complexity including false beliefs, flawed mental models, and incorrect ontological categories. We hypothesized that conceptual change of a mental model requires change in the *system of relations* between the features of the prior model. To test this hypothesis, we compared instruction aimed at revising knowledge at the mental model level called *holistic confrontation* – in which the learner compares and contrasts a diagram of his or her flawed mental model to an expert model – to instruction aimed at revising knowledge at the false belief level – in which the learner is prompted to self-explain the expert model alone. We found evidence that participants who engaged in holistic confrontation were more likely to acquire a correct mental model, and a deeper understanding of the systems of relations in the model than those who were prompted to self-explain the expert model. The results are discussed in terms of their implications for science instruction.

1. Introduction

A central problem in research on learning and instruction is to understand how students move from having a misconception to having a correct understanding of a complex phenomenon or idea. This problem is especially pertinent in the case of science learning, where students often come into the classroom with different beliefs than those held by teachers and scientists. Past research has explored what kinds of instructional activities support the revision of prior beliefs into the to-be-learned concept – a process typically described as "conceptual change" (e.g., Duit & Treagust, 2003; Posner, Strike, Hewson, & Gertzog, 1982) and recent work has focused on understanding the underlying cognitive mechanisms and processes that facilitate such change (Chi, 2008; diSessa, Gillespie, & Esterly, 2004; Ohlsson, 2009). Although much progress has been made in understanding the nature of misconceptions and knowledge revision processes, more work is needed to develop instructional interventions based on that understanding to facilitate robust conceptual change for particular kinds of misconceptions.

In the current work, we address this goal by developing and testing a type of instruction for facilitating conceptual change when students' misconceptions are in the form of flawed mental models. To do this, we build on past work that has identified multiple types of misconceptions and articulate a framework for facilitating conceptual change for each type. Chi (2008) has characterized three kinds of student misconceptions that increase in their representational complexity from false beliefs to flawed mental models to incorrect ontological categories. This framework suggests that different kinds of cognitive processes and instruction may be differentially effective in facilitating conceptual change for a given level of representational complexity. We hypothesize that as the representational complexity of the misconception increases, so does the amount of transformation needed to rectify it. Specifically, we hypothesize that conceptual change at the mental model level requires knowledge revision to the *interrelations between the features* of the prior knowledge, which is different from revising individual false beliefs or reassignment of a concept to an ontological category. Furthermore, we propose that instruction that focuses the learner on revising systems of relations of the misconception, what we call "holistic confrontation", should be more effective in facilitating change of a flawed mental model than instruction that focuses on revising false beliefs or the type of ontological category.

We investigated these hypotheses in a laboratory experiment by giving students who had a flawed mental model of the circulatory system one of two types of instruction. Half of the students were given instruction that emphasized holistic confrontation in which they received a diagram of their flawed mental model and

* Corresponding author. 815, Learning Research and Development Center, 3939 O'Hara Street, University of Pittsburgh, Pittsburgh, PA 15260, United States. Tel.: +1 412 624 7048; fax: +1 412 624 9149.
E-mail addresses: soniya.gadgil@gmail.com, smg58@pitt.edu (S. Gadgil).

0959-4752/$ – see front matter
doi:10.1016/j.learninstruc.2011.06.002

a diagram of the expert model, and were asked to compare and contrast the models. Based on prior work on analogical comparison, we expected that by comparing and contrasting the models, students would focus on the interrelations between the features and notice discrepancies between the systems of relations of the two models. The other half of the students were given only the expert diagram and were prompted to self-explain it. Although prior work has shown that self-explanation can be a powerful learning activity (e.g., Bielaczyc, Pirolli, & Brown, 1995; Chi, de Leeuw, Chiu, & LaVancher, 1994), we expected that prompting for self-explanation would focus students on revising knowledge at the false belief level of representation. This means that conceptual change at a mental model level will take place only to the extent that the learner is able to transform the structure of his or her model based on the accumulation of revised false beliefs. If conceptual change of a mental model is best facilitated by revising systems of relations, then participants given instruction that emphasizes holistic confrontation (compare and contrast) should lead to more robust conceptual change than instruction that focuses on the accumulation of revised beliefs (prompted self-explanation).

To situate the work, we first provide a description of the three levels in which knowledge can be misconceived, followed by a discussion of the cognitive processes hypothesized to support revision of that knowledge and the instructional interventions designed to facilitate those processes. Then we report the results from our laboratory study that provide some initial support for holistic confrontation and discuss the implications of holistic confrontation for instruction.

1.1. Levels of misconceived knowledge

Knowledge can be misconceived at three levels of complexity including false beliefs, mental models, and ontological categories (Chi, 2008). The most basic level of misconceived knowledge is the false belief level. A *false belief* is a single incorrect idea that can usually be stated in a single proposition. When knowledge is misconceived at the false belief level, students have one or more incorrect individual beliefs about what they are about to learn. When we say that a belief is "incorrect" or "false" we mean that the belief differs from some scientific normative standard.[1] For example, if a student thinks, "*all* blood vessels have valves," this is in direct contradiction to the correct proposition "only veins have valves." Thus, false beliefs are relatively easy to rectify because they do not require radical restructuring of a mental model.

The next level of misconceived prior knowledge is that 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 phenomenon it represents and is often described as an "analog representation." It consists of multiple propositions and features, as well as interrelationships between those features. Some of these relations are not represented explicitly as propositions and may be inferred from the features of the model. Thus, mental models are not simply a collection of individual beliefs, but because they possess complex interrelationships between propositions, they have a cohesive integrated structure. Because of this, even though a student may possess several correct individual beliefs about a scientific phenomenon, he or she may still have a flawed mental model.

The issue of whether students' misconceptions are fragmented, individual units of knowledge such as false beliefs, or are embedded in larger theory-like structures such as mental models has generated considerable controversy in the conceptual change literature (e.g., diSessa et al., 2004). On one side of the controversy is the view that students' misconceptions can be represented as complex, cohesive knowledge structures such as mental models (e.g., Chi & Slotta,, 1993; Chi, Slotta, & de Leeuw, 1994; Vosniadou, 1994, 2008). On the other side is the view that students' prior knowledge can be characterized only as individual or fragmented fine-grained knowledge components called phenomenological primitives or "p-prims," which do not form cohesive structures or models (e.g., Smith, diSessa, & Roschelle, 1993; see Brown & Hammer, 2008 for a detailed review of these two views). A central hypothesis of the current work is that students' misconceptions are represented at multiple levels of complexity depending on prior experience and the particular conceptual domain.

The third and most complex level of misconceived prior knowledge is the ontological level. Misconceptions at this level occur when students misattribute a scientific phenomenon to one kind of category versus another (Chi, 1997, 2005). This means that if students' prior conceptions belong to one lateral or ontological category and correct conceptions belong to another category, they conflict by definition of *kind* or *ontology*, and therefore conceptual change requires a shift across lateral or ontological categories. Chi (2005) has defined two broad ontological categories for scientific processes: direct and emergent processes. Direct processes are those in which the components of a system directly cause changes in other aspects/components of the system. For example, in the circulatory system, the organization of different parts (e.g., two loops of blood flow: one between the heart and the body and the other between the heart and lungs) dictates how the blood flows in the system. In contrast, for emergent processes, the system level activity emerges from the interaction of many different components acting simultaneously and cannot be attributed to direct causal effects of any one component or part. For example, in diffusion of clear liquid and dye, single individual molecules do not dictate the direction in which either liquid moves, rather, the overall pattern of movement arises from smaller constituent level interactions between the clear liquid molecules and the dye molecules.

Misconceptions at the level of ontological category tend to be the most robust, and require conceptual change of the category. Given that the ontological category "emergent processes" is not intuitive, students are often not aware that it exists, and therefore, instruction to correct such misconceptions needs to first make students aware of the ontological category, and then help them assign the target concept to the appropriate category. For the "diffusion" example, the instruction should help students determine that diffusion does not belong in the category of a direct or causal process, but is instead an emergent phenomenon (Chi, 2005). Many concepts in the domain of physical and biological processes such as heat, diffusion, and evolution require ontological shift.

To sum up, each level of misconceived knowledge has a particular representation and structure. False beliefs consist of individual propositions, mental models consist of systems of propositions and features (interrelations), and ontology is a fundamental (root) category for a concept that impacts the interrelations of the underlying tree structure (i.e., hierarchical inheritance). Each level increases in its complexity from false beliefs to mental models to the ontological category, and each level is hypothesized to require increasing amounts of transformation to facilitate conceptual change.

[1] False beliefs are only incorrect in how they relate to some normative conception. We note that all science concepts and theories are tentative, and in a sense are always under revision – this is the nature of science. However, we adopt the present terminology to provide an operational definition to describe how conceptions change in relation to a normative standard (they are not incorrect or correct in a absolute sense).

In the domain we chose for our current investigation - the circulatory system, the identified misconceptions are hypothesized to be at the mental model level. Past work provides empirical evidence consistent with the view that students hold various types of mental models of this concept. For example, if students have a flawed, single-loop model, they believe that the heart takes in blood, oxygenates it, and sends it to the rest of the body in a continuous single loop. In contrast, a scientifically correct double-loop model consists of two distinct loops, one loop in which blood comes into the heart deoxygenated, is sent to the lungs to receive oxygen, and returns to the heart oxygenated, and a second loop in which oxygenated blood is sent to the rest of the body from the heart through arteries, and returns deoxygenated through veins. Other kinds of flawed models of the circulatory system have also been identified, for example, the no-loop model, the ebb-and-flow model, and several models that fall in between the single-loop and double-loop models (Chi, de Leeuw, et al., 1994; Chi, Slotta, et al., 1994). Because these models consist of multiple features and relations, they require more transformation and different kinds of cognitive processes than those required to change individual false beliefs. Importantly, these models are all members of the same ontological category of direct processes. Therefore, conceptual change for this concept is not expected to require an ontological shift, but a transformation of the mental model. In the next section, we describe the cognitive processes that have been hypothesized to facilitate change at the different representational levels.

1.2. Cognitive processes that facilitate belief revision and conceptual change

It is important to first distinguish conceptual change from other kinds of learning processes such as the addition of new knowledge and gap filling (Chi, 2008). *Addition of new knowledge* occurs when one has no prior knowledge about the topic. For example, if a person has no prior knowledge of the human circulatory system, all that the person learns about this topic will be represented as the addition of new knowledge to long-term memory. Similarly, *gap filling* takes place when a learner has some incomplete prior knowledge about the material to be learned, with gaps that need to be filled in. For example, a learner may know that the human heart has four chambers, but might not know the names of the four chambers. In this case, learning involves simply "filling in" the missing information or elaborating the prior knowledge. This kind of learning has also been termed *enrichment* (Carey, 1991) or *accretion* (Rumelhart & Norman, 1978). Gap filling and the addition of new knowledge do not constitute conceptual change, as the learner does not need to change his or her existing knowledge or deliberately get rid of prior misconceptions. We define conceptual change as the transformation of prior knowledge that conflicts with the to-be-learned concepts (for alternative views of conceptual change, see diSessa et al., 2004; Ohlsson, 2009).

Cognitive processes associated with conceptual change include inference generation and knowledge revision at either the propositional, mental model, or ontological level (Chi, 2000, 2008). Inference generation is a cognitive process in which the learner constructs new knowledge from prior knowledge. It is hypothesized to be a core part of many higher-order cognitive processes including: reasoning, problem solving, decision-making, self-explanation, and text comprehension among others. Knowledge revision requires metacognitive monitoring in which the learner detects misunderstandings or gaps in his or her current understanding and then sets the goal of revising that knowledge to remove the contradiction or fill in the missing knowledge (Chi, 2000; Winne & Hadwin, 1998; Zimmerman & Campillo, 2003). Conceptual change requires the revising of prior knowledge to accommodate the new, conflicting information (e.g., data, evidence, argument, instruction). The central difference across the levels is whether it is a single proposition (belief revision), an interrelated set of propositions and features (mental model revision), or the underlying category (ontological revision) that is being revised.

Analogical reasoning processes may also facilitate conceptual change. Much prior work has shown that analogical processes that align and map systems of features and relations from one mental representation to another can facilitate learning, problem solving, and transfer (Gentner, 1983; Gentner, Loewenstein, & Thompson, 2003; Gick & Holyoak, 1983; Nokes & Ross, 2007). One mechanism through which analogy is hypothesized to work is *intersection discovery*, in which the objects and relations that overlap between the two representations are preserved and the non-overlapping objects and relations are discarded (Hummel & Holyoak, 1997). The resulting knowledge representation is described as an abstract schema. Schema abstraction highlights the structural features of the concept and makes transfer to new contexts easier (Gick & Holyoak, 1983). The comparison and mapping process not only highlights the similarities of the representations being compared but also highlights the differences and provides students an opportunity to engage in inference generation and knowledge revision. A few studies have explored the role of analogical reasoning in conceptual change process, however, most of these studies as we will review in the following section have targeted change at the individual belief level rather than at the mental model level.

In the current work, we hypothesize that conceptual change of a mental model will be facilitated if systems of relations within a student's own misconceived model are contrasted with the correct model. Analogical reasoning processes should highlight not only the individual propositions and features that differ between the students' prior knowledge and the expert model but also the interrelations between those features at a system level. Highlighting these features and relations of the mental model should facilitate systemic conceptual change. In the next section, we review the research on instructional interventions designed to facilitate inference generation, knowledge revision, and analogical reasoning processes.

1.3. Instruction supporting belief revision and conceptual change

One of the primary instructional techniques that have been employed to bring about conceptual change is cognitive conflict, particularly instantiated in the form of refutational texts (see Tippett, 2010 for a recent review). In a refutational text, common misconceptions held by students are identified and explicitly addressed. There is some evidence that refutation can be successful in helping students revise their incorrect beliefs. For example, Diakidoy, Kendeou, and Ioannides (2003) gave sixth-grade students either a standard, expository science text about energy or another text that explicitly addressed two common misconceptions that students held about energy and refuted them. After reading the text, they were tested on questions related to the misconceptions, such as those that required students to differentiate between energy, force, and matter. Results showed that students who read the refutation text performed better than those who read the simple expository text. Other studies have shown similar results for facilitating conceptual change using refutation texts at the belief level (Hynd & Alvermann, 1986; Kendeou & van den Broek, 2007).

Because refutational texts typically refute individual propositions in a text in a serial fashion, refutation as an instructional technique may work well for the false belief level of misconceived knowledge, but may be less effective when the misconception resides at the level of the mental model. For conceptual change to take place at the mental model level, a learner must revise both the individual features of the model as well as the *interrelationships* between those features. When reading a refutational text, a student is likely to read it line by line, and focus on revising individual false beliefs explicitly refuted by the text rather than on the interrelationships between those beliefs, therefore making change at a mental model level less likely. For example, in a study conducted by Hynd, McWhorter, Phares, and Suttles (1994), high-school students who had a nonscientific view of motion based on the impetus theory, were given either a refutational text or a standard physics text. The students' misconceptions in this case were at a mental model level, because an impetus theory is not a single incorrect belief or collection of beliefs, but involves several interrelated propositions that can lead to incorrect inference generation. Although reading a refutational text led to conceptual change in 52% of students in this study, compared to only 22% of students who read the standard text, this was clearly not a ceiling effect. A possible explanation is that while approximately half of the students were able to transform their mental models based on accumulation of individual correct propositions, others were not able to do so, because they were not focused on the interrelationships between the features of the model.

Another type of instruction that has been studied extensively in the context of conceptual change is prompting students to self-explain as they read a passage or solve a problem. Self-explanation prompts have been shown to facilitate learning and transfer in a number of tasks (Chi, 2000; Chi, de Leeuw, et al., 1994; Chi, Slotta, et al., 1994; Nokes, Hausmann, VanLehn, & Gershman, in press). Self-explanation is hypothesized to facilitate conceptual change because it forces students to reconcile their existing knowledge with the correct scientific knowledge, resulting in the acquisition of the correct beliefs through inference generation and knowledge revision. In a typical self-explanation study, students are prompted to explain a text line by line by giving them content-free prompts such as "what does that make you think of?" or "can you elaborate on that?" Chi, de Leeuw, et al. (1994) and Chi, Slotta, et al. (1994) showed that when students were prompted to self-explain as they read a text about the circulatory system, approximately two-thirds could successfully transform their flawed model to the correct model, presumably through the accumulation of several correct beliefs while self-explaining. However, a third of the students still retained a flawed mental model and appeared to simply integrate the newly learned propositions into their flawed model. Why were these students unable to revise their prior model?

One reason for this result may be that prompting for self-explanations targeted the propositional level of conceptual change instead of focusing on the interrelations between features at the mental model level. Had the students been focused on the relational structures at the mental model level, they may have been more likely to transform the system of relations. Therefore, we hypothesize that holistic confrontation will be more effective than self-explanation in driving conceptual change at the mental model level because it targets systems of relations rather than individual beliefs. Comparing one's flawed mental model to a correct model provides an opportunity to notice holistic, relational discrepancies and their conceptual implications. One way to facilitate noticing such systems of relations is through *analogical comparison* by comparing and contrasting models (Gentner et al., 2003; Kurtz, Miao, & Gentner, 2001).

Past research in analogical comparison has shown it to be an especially powerful learning activity (Gentner et al., 2003; Kurtz, Miao, & Gentner. 2001; Nokes & VanLehn, 2008). The process of comparing can serve to highlight particular aspects of the situation, object, or example that is being compared that would not typically be salient (i.e., outside of the context of comparison). For example, in an experiment conducted by Medin, Goldstone, and Gentner (1993), participants were shown an ambiguous stimulus along with either of two unambiguous stimuli (see Fig. 1). Participants shown B with A attributed properties of A to B (three prongs), whereas those shown B with C attributed properties of C to B (e.g., four prongs). In other words, the process of comparison facilitates the perception of particular features over others as well as focusing on how those representations differ (e.g., A and B have different shapes versus B and C differ in that B has a warped prong). Thus, during holistic confrontation, when learners are asked to compare a flawed model to a correct model, they will be more likely to focus on specific features and relations that they would not focus on when shown each model individually.

Analogical comparison has been used as an instructional intervention in a variety of contexts including science learning (e.g., Schwartz & Bransford, 1998; Nokes & VanLehn, 2008) and mathematics problem solving (e.g., Rittle-Johnson & Star, 2007). Schwartz and Bransford (1998) examined the benefits of comparison instruction in an experiment in which undergraduate students either compared cases (data) illustrating memory concepts or read summaries about the cases. In the contrasting cases condition, students compared cases and looked for patterns in the data, whereas in the read-only condition, they were provided with text summaries of the data. All students later attended a lecture in which they were taught about the theories that explained the cases. Students in the contrasting cases condition showed better performance on a transfer task compared to those in the summary condition. Schwartz and Bransford hypothesize that the students who contrasted cases noticed critical features and relations of the domain concepts that prepared them to learn more deeply from the lecture that later explained those concepts.

Further evidence comes from work by Kurtz et al. (2001) in which participants who were asked to compare examples while learning about the concept of heat transfer learned more than those who studied each example separately. By comparing the commonalities between two examples, students could focus on the causal structure and extract an abstract schema leading to better learning. Similarly, Rittle-Johnson and Star (2007) gave one group of seventh grade algebra students two analogous algebra problem solutions to compare and contrast, while the control group received just two worked examples sequentially. Students in the compare and contrast condition outperformed the control group on both procedural as well as conceptual assessments of learning.

Analogical comparison has also been studied in the context of conceptual change (see Dagher, 1994 for a review). For example, Clement (1993) sought to address students' prior beliefs in physics, that static objects are barriers that cannot exert forces. For example, if a book is placed on a table, students believe that the table does not exert an upward force on the book. In our framework, this misconception would be at the level of false beliefs. To address this false belief, Clement gave students series of what he calls "bridging analogies", which made the link between the to-be-learned concept, and the analog. Those students who received these bridging analogies showed greater gains on a conceptual understanding posttest. Blanchette and Dunbar (2002) also found that analogical reasoning can alter the representations of target information, in this case, information about a socio-political issue. Participants who read the

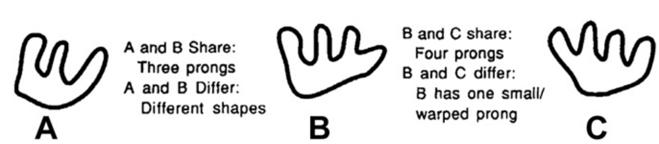


Fig. 1. Example of how the contrasting case (A or C) impacts how the perception of B. From Medin, Goldstone, and Gentner (1993).

informational text about the socio-political issue ending with an analog were found to change their initial representations of the issue on subsequent test by erroneously incorporating the analogical inferences. These findings show that, although there has been some success with using analogies in conceptual change, it has largely been in the context of revising individual beliefs, and has not been demonstrated in the realm of mental model change. The key difference between past research on analogical comparison and our present 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, in the context of the human circulatory system.

1.4. Research hypotheses

Past research suggests that analogical comparison can promote cognitive processes such as the encoding of critical features, inference generation, and schema extraction, which help bring about conceptual change at both the false belief and the mental model level. We hypothesized that comparing a flawed diagram to an expert diagram would facilitate inference generation and noticing of critical relations, and will provide an opportunity for change at the mental model level (noticing and correcting errors), compared to simply explaining the expert diagram. More specifically, we expected students in compare condition to perform better than those in the explain-diagram condition, on questions tested their revised model, and on questions that required inference generation. On the other hand, on questions that required learning of simple declarative facts (e.g., defining parts of the circulatory system), we expected the two conditions to perform similarly.

Our four primary dependent measures included a diagram test on which students were asked to draw a diagram of their current understanding of the circulatory system, a declarative knowledge test which tested their knowledge of terms relating to the circulatory system, a mental model test which tested their mental models of the circulatory system, and an inference test which tested them on knowledge not directly conveyed in the learning materials, but required them to generate inferences in order to answer correctly. These dependent measures will be described in more detail in Section 2. Additionally, we also analyzed data from verbal protocols that we collected as students compared or explained the diagrams.

Hypothesis 1. The diagram test was the most stringent assessment of mental model change because participants would be able to draw the correct path of blood in the circulatory system only if they had the correct mental model after learning. If participants in the compare condition noticed the discrepancies between their model and the correct model, they would have an opportunity to revise their existing model accordingly. In contrast, we expected that those participants explaining the correct diagram would focus on explaining features of the expert model and less on the discrepancies between their model and the expert model. Therefore, we expected a higher proportion of the participants in the compare condition to draw the correct double-loop model at posttest when compared to the explain-diagram condition.

Hypothesis 2a. Comparing diagrams should increase the likelihood of noticing the critical features and relations of the model, and therefore, we expected the compare condition to outperform the explain-diagram condition on the mental model test.

Hypothesis 2b. By noticing critical features and relations of the model, participants would be more likely to generate inferences, and therefore, we expected the compare condition to outperform the explain-diagram condition on the inference test.

Hypothesis 3. We did not expect differences across conditions on the declarative knowledge test since both groups were exposed to the same declarative knowledge through the instructional text that they read after either comparing diagrams or explaining the correct diagram.

Hypothesis 4. With regards to the participants' verbal protocols, we predicted that participants in the compare condition would make more constructive statements than those in the explain-diagram condition. We defined constructive statements very broadly to include any statement that went beyond paraphrasing the diagram, which included function-related inferences, elaborations, monitoring statements, and comparisons (see Section 2 for the coding rubric). We expected that the overall advantage on constructive statements would be driven by the number of comparisons generated by each group. Specifically, we expected that the compare group to generate more statements of comparison between the flawed model and the expert model than the explain-diagram group. Note that the explain-diagram group could also generate such comparisons by comparing the expert diagram to their internal mental representation, although it was not given to them in the form of a diagram.

Hypothesis 5. We predicted that participants in the compare condition would generate more function-related inference statements than the explain-diagram group because comparing the diagrams would facilitate noticing multiple features and provide an opportunity to integrate those features. Critically, function-related inferences bring together multiple features and relations of the model. They describe how a feature relates to other features and parts of the model.

Hypothesis 6. Finally, we predicted the explain-diagram group to make more revision statements at the false belief level than the compare group. These statements are captured by revisions to their prior knowledge at the propositional level.

2. Method

2.1. Participants

We screened 84 graduate and undergraduate students from the University of Pittsburgh and 58 of them (27 women, 31 men, Mage = 21.2 years, age range: 18–29 years) qualified to participate based on a test of their prior knowledge (described in Section 2.4). All students were native or fluent English speakers. Participants were either paid $15 for the 2-h study or received partial course credit.

2.2. Design

We used a between-subjects design with participants randomly assigned to one of two instructional conditions: compare ($n = 30$) or explain-diagram ($n = 28$). The distribution of participants who participated for credit and for pay was equal across conditions.[2]

[2] As an afterthought, we thought we needed a control condition and therefore a new condition was run after completing the data collection for the two experimental conditions. The new condition asked the learners to read the text twice without having seen any diagrams beforehand. In retrospect, not providing any diagrams changes the task demands of this read-twice condition considerably so that it does not serve the function of a control condition. Therefore, we will not report the results of this new read-twice condition since it does not make sense to compare the results of this new condition to the two conditions reported in this study.

Table 1
Description of materials.

		Materials	Description	Example
Pretest (about 25 min)		• Mental model test (Score: 0–20)	6 short-answer questions	Describe the path of blood from the heart to the various parts of the body. (See Appendix A for complete list)
		• Declarative knowledge test (Score: 0–22)	12 definition questions	Aorta: What does it mean? Where is it located? What is its function?
Instructional intervention (about 25 min)	Compare Condition (Diagram comparison: 5–7 min)	• Flawed single-loop diagram	Diagram	See Appendix B
		• Correct double-loop diagram	Diagram	See Appendix C
		(Talk aloud while comparing diagrams; verbalizations recorded)	List of prompts	See Appendix E
		• Expository text describing human circulatory system; 72 sentences, presented on computer screen line by line.	Text	See Appendix D
	Explain-Diagram Condition (Studying correct diagram: 5–7 min)	• Correct double-loop diagram	Diagram	See Appendix C
		(Talk aloud while explaining diagram; verbalizations recorded)	List of prompts	See Appendix E
		• Expository text describing human circulatory system; 72 sentences, presented on computer screen line by line.	Text	See Appendix D
Posttest (about 55 min)		• Blank diagram to be completed and explained while drawing; audio protocols were collected.	Diagram	
		• Knowledge Inference test (Score:0–25)	18 short-answer questions	Why don't we have valves in pulmonary veins?
		• Mental model test (Score: 0–20); Same as pretest	6 short-answer questions	Same as pretest; (See Appendix A for complete list)
		• Declarative knowledge test (Score: 0–22); Same as pretest	12 definition questions	Same as pretest

2.3. Materials and measures

Participants learned about the human heart and circulatory system. Materials and measures were adapted from Chi, Siler, Jeong, Yamauchi, and Hausmann (2001) and consisted of a pretest, learning materials, and a posttest. All learning and test activities except for a drawing test were administered on a computer. See Table 1 for a summary of all pretest, posttest and learning materials.

2.3.1. Pretest materials

The pretest consisted of two parts: a mental model and a declarative knowledge test. The mental model test was used to determine participants initial mental model of the circulatory system. It consisted of six short-answer questions that required students to describe their current mental model (see Table 1 for an example question; see Appendix A for a full list of questions). The following coding scheme adapted from Chi, de Leeuw, et al. (1994) and Chi, Slotta, et al. (1994) was used to classify students' mental models:

1. *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.
2. *Partially correct double loop:* 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. Or all the above elements are present, but path is incorrectly described.
3. *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.
4. *Single loop:* Blood is primarily contained in blood vessels and pumped from the heart to the body and returns to the heart via different blood vessels.
5. *Other (unclassifiable):* Any description that could not be classified under one of the above categories was coded as "Other."

The six mental model questions were designed to assess students' existing mental models of the circulatory system. The maximum possible score on the mental model test was 20 points. Question 1 asked them to describe the path of the blood in the circulatory system. It was worth four points, one point for each correct component of path. If a student correctly described the double-loop model, he or she was dropped from further analyses. If the student gave a partially correct or incomplete answer such as "blood goes from heart to the body and back to the heart," without mentioning the lungs, then Question 2 was examined. Question 2 was worth five points, one point for mentioning each of the following: arteries, veins, arterioles, venules, and capillaries. If they correctly mentioned the arteries and veins in Question 2, it was clear that they had some idea that different kinds of blood vessels carry different kinds of blood. This type of answer suggests that they had either a single-loop or a single-loop with lungs model. Question 3 was worth five points, one point was given for describing each of the following components: the four chambers, atria, ventricles, valves, and septum. On Question 3, if students correctly described the parts of the heart, they were considered to have either a correct double-loop or single-loop with lungs model and hence were dropped form further analysis. Question 4 was worth three points and participants were given one point each for mentioning the heart, lungs, and blood vessels. On Question 4, if they mentioned the lungs and correctly described their function of oxygenating blood, they were dropped form further analysis. For Question 5, they were given one point for supplying oxygen, and one point for any secondary function, such as removing waste or maintaining body temperature. If on Question 5, if they correctly mentioned carrying oxygen to cells from lungs and carrying CO_2 away from cells, as one of the functions of blood, they were considered to have the correct mental model and dropped from further analysis. Finally, on Question 6 if they stated the function of the heart as oxygenating blood, they were considered to have a single-loop model, and were selected to receive further instruction. Consistent with our expectations, we did not find evidence for the ebb-and-flow or no-loop models in students' answers. The inter-item reliability coefficient for these questions was moderate with Cronbach's alpha = 0.60.

The *declarative knowledge test* consisted of questions in which participants were asked to define terms related to the circulatory system (see Table 1 for example). This measure was used to assess their declarative understanding of these terms. Participants were given generic prompts, such as "Where is it located?" and "What is its function?" etc. This test had 12 items, with a maximum possible score of 22 points, and the inter-item reliability was moderate with Cronbach's alpha = 0.70 at pretest. Ten of these had a maximum possible score of two points, with one point awarded for the correctly describing what the term means, and one point for the correct description of its function. For example, for the term "atrium", if a student wrote, "an atrium is a chamber of the heart and receives blood from the body or the lungs", they would get full credit, but if they just wrote, "atrium is a chamber of the heart", they would get only one point. The remaining two terms "heartbeat" and "systemic circulation" carried only one point each, because their meanings and functions can be described in a single proposition.

2.3.2. Learning materials

The learning materials consisted of a flawed single-loop diagram of the circulatory system (see Appendix B), a correct double-loop diagram of the circulatory system (see Appendix C), and an expository text that described the structure of the human heart, the function of various components, and the path of blood flow (see Appendix D). The diagrams were hand-drawn on an outline of the human body, with the heart and other organs drawn in black, and red and blue lines indicating the oxygenated and deoxygenated blood respectively. Arrows indicated the direction of blood flow. In the correct double-loop diagram, only the basic parts, that is, the atria and the ventricles, the septum, arteries, veins, and the lungs were labeled, whereas in the flawed single-loop diagram, the heart, the arteries, veins, and lungs were labeled. Only the basic parts were labeled in order to keep the level of detail similar across the two diagrams. The flow of blood to different parts of the body was indicated by arrows. We chose the single-loop model of the circulatory system as our misconception model because we screened for only those participants who held this model as determined by their pretest scores. Students in the compare condition were shown the single-loop diagram as well as a diagram of the correct double-loop model. Students in the explain-diagram condition were shown only the correct double-loop model. The text about the heart and circulatory system was common to both conditions and was taken from a popular college-level Biology textbook (Shier, Butler, & Lewis, 2006). The text was 72 sentences long, had no accompanying diagrams, and was presented on the computer one line at a time.

2.3.3. Posttest materials

The posttest consisted of the same mental model and declarative knowledge tests given at pretest. In addition, there was a diagram test and a knowledge inference test. In the *diagram test,* students were given a template of the human body on paper, and asked to draw the path of blood flow through the circulatory system on that template. This test provides a fine-grained assessment of students' mental models of the circulatory system because in order to draw the correct path one must have a correct mental model. Furthermore, prior work has developed a fine-grained coding rubric to assess the correctness of the learner's model (Chi, de Leeuw, et al., 1994; Chi et al., 2001; Chi, Slotta, et al., 1994; see Section 2.4 for our adaptation of the coding rubric). The diagram test was not given at pretest, because in the self-explain condition, drawing the diagram at pretest may have prompted comparisons to their prior models while self-explaining the correct diagram, thus making the distinction between the two conditions less clear.

The *knowledge inference test* contained questions that were not directly addressed in the test. Participants had to make connections either across sentences in the text, or between something described in the text and their prior knowledge to answer correctly. These questions required inference generation in order to reason using one's mental model. For example, the answer to the question, "why does the heart have four chambers?" is not directly described in the text. A student who continues to have a single-loop model after instruction, and therefore believes that the heart oxygenates blood, will not be able to explain why a heart has four chambers. In contrast, a student who has the correct double-loop model will be able to generate this inference by mentally running their own model, and concluding that the heart needs a divider to separate oxygenated and deoxygenated blood, and also separate chambers to take in and pump out blood.

The knowledge inference test had 18 items and the maximum possible score on this test was 25 points[3]. Each item on this test carried either one or two points based on the number of critical propositions it required to be answered correctly. For example, in the question, "Why is it not necessary for capillaries and arteries to have valves?", two distinct propositions were involved, one being that "capillaries are extremely small, and cannot contain valves" and the other being that "arteries carry blood away from the heart, which is already under high pressure, therefore does not require valves." Other questions such as "What will be the consequence of having a hole in the septum" carried only one point, because the correct answer required only a single correct proposition, that is, "it will cause the oxygenated and deoxygenated blood to mix." The knowledge inference test was administered only at posttest to avoid students' focusing specifically on those items during learning. The inter-item reliability for this test was high with Cronbach's alpha = 0.82.

2.4. Procedure

The procedure consisted of the pretest (mental model and declarative knowledge tests), instruction, and a posttest (diagram, mental model, declarative knowledge, and knowledge inference tests).

2.4.1. Pretest procedure

The pretest consisted of the mental model test and the declarative knowledge test. Each participant first took the mental model test. There were two purposes for conducting this test. The first was to screen participants for prior knowledge of the circulatory system. The second was to gauge the participant's current mental model. Only those participants who described a single-loop model and scored ten or less out of 20 points proceeded to the next part of the

[3] Full materials and scoring template available upon request.

experiment. The mental models were scored based on the rubric described under Materials and measures. Out of 84 participants, 58 were found to have a single-loop model and scored less than ten points on the mental model test. Participants who had models other than the single-loop model were debriefed and excluded from further participation. Next, the screened participants took the declarative knowledge test. The entire pretest took approximately 25 min. No feedback was given during either part of the pretest.

2.4.2. Learning procedure

Participants were randomly assigned to the two instructional conditions. In the *compare condition*, participants 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. Since all the participants were screened to have a single-loop model, they all agreed that the flawed diagram was correct. They were then told that this was not actually the correct model, and that they would be shown a new diagram of the correct model and asked to make comparisons between the two diagrams. The experimenter then informed the participant that they would be asked some questions to help them compare the diagrams, and that even though some questions may be repetitive, participants should try to answer them, and keep talking aloud as they are comparing. The prompts were presented in a pre-determined order. The first prompt for the compare condition simply asked them to compare across diagrams, *"What similarities and differences can you see in the two diagrams?"* Further prompts directed them to make specific comparisons, for example, "*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?"* (See Appendix E for a complete list of prompts). Participants' verbalizations were audio-recorded. If they fell silent for more than 10 s, the experimenter reminded them to keep talking. If they made vague statements, for example, "this seems to be going there…" the experimenter asked them "can you elaborate on what you just said?" No other feedback was given by the experimenter. The time spent comparing was limited to approximately 5–7 min. Participants then read the circulatory-system text presented one line at a time on the computer screen. They were told to read the text at their own pace, like they would study for a test.

Participants in the *explain-diagram condition* were only shown the correct diagram and asked to explain it. Similar to the compare condition, the experimenter told the participant that they would be asked some questions to help them explain the diagram, and that even though some questions may be repetitive, participants should try to answer them, and keep talking aloud as they are explaining. They were given prompts equivalent to those in the compare condition, presented in a pre-determined order. The first prompt simply asked participants to explain what they saw in the diagram; "*What do you see happening in this diagram?*" Further prompts directed them to explain specific aspects of the diagram, for example, *"Can you trace the path of blood flow in the diagram?" and "What are the important components of the circulatory system based on this diagram?"* (See Appendix E for a complete list of prompts). Participants' verbalizations were audio-recorded. The time spent explaining was also limited to approximately 5–7 min. If they fell silent for more than 10 s, the experimenter reminded them to keep talking. If they made vague statements, the experimenter asked them to elaborate on what they said. No other feedback was given. Following the explanation of the diagram, participants read the circulatory-system text. Similar to the compare condition, the text was presented one line at a time on the computer screen, and participants were told to read it at their own pace, like they would study for a test. The total time spent learning for both conditions was approximately 25 min.

2.4.3. Posttest procedure

Immediately after reading the circulatory-system text, participants took the posttest. They first completed the diagram test in which they were asked to draw a diagram showing their current understanding of the heart and circulatory system. They were given a blank outline of the human body, and asked to draw and label as many parts of the circulatory system as they could remember, based on the diagram/s they had seen earlier, and the text that they had read. They were told not to worry about drawing the most realistic representation, but to make a schematic drawing of the circulatory system. They could use colors to indicate different parts. They were asked to explain aloud as they drew, and audio protocols were collected. If they fell silent for more than 10 s, the experimenter reminded them to keep talking. Next, participants completed the remainder of the posttest including the mental model, declarative knowledge, and knowledge inference tests. The entire posttest took approximately 55 min.

2.4.4. Coding of constructive statements in verbal protocols: definition and coding categories

We adapted Renkl's (1997) coding scheme for categorizing types of statements that students generated. Consistent with the scheme proposed by Renkl (1997), we defined constructive statements very broadly to include any statement that went beyond simply paraphrasing information in the diagram/s. These include several fine-grained categories such as function-related inferences, elaboration, etc. described later. Our adaptation did not include the anticipative reasoning and principle-based explanation categories proposed by Renkl (1997), because they are more relevant to a problem-solving situation. We also added a category of comparison statements to capture statements aligning or mapping features and relations between the flawed and correct models (see below). We first coded for the total number of constructive statements and then coded for the *types* of constructive statements according to the following categories:

i. *Function-related*: Students explanations were coded as function-related 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*: 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*: 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.
   For example, *"Oh, so* the deoxygenated blood comes from the body, goes through the heart into the lungs, and then it gets oxygenated IN the lungs, okay, and it comes back through the right side."
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 were coded as comparisons if students were explicitly comparing the flawed model with the correct one.

For example, *"Well this one has the different chambers in it instead of just one unified thing."*

We analyzed a randomly selected subset of students' verbal protocols from each instructional condition. We analyzed 76% of all the protocols ($n = 22$ of compare condition and $n = 22$ of the explain-diagram condition). Each protocol was segmented into a set of statements. A statement was operationalized as a single complete sentence. The length of the protocols ranged between 35 and 50 statements. A primary and a secondary coder first coded 25% of the protocols, based on the coding scheme described earlier. Inter-rater reliability was high; Cohen's Kappa = 0.87. Disagreements were resolved via discussion, and the primary coder proceeded to code the remaining protocols.

2.4.5. Diagram scoring procedure

The same classification described for classifying mental models at pretest described in the pretest materials was used to classify students' drawings. Only those participants who clearly showed the path of the blood as follows: deoxygenated blood coming in to the right atrium, going to the right ventricle, then to the lungs, becoming oxygenated and returning to the left atrium, going to the left ventricle, and from the left ventricle to the rest of the body via arteries, to return to the right atrium via veins were scored as the double-loop model. Any participant who mentioned that the lungs oxygenated blood, but drew an incorrect path was classified as having a partially correct double-loop model. If a participant showed that the blood goes to the lungs just like it goes to the other parts of the body, that is, oxygenated, then it was classified as a single-loop with lungs model. If lungs were not mentioned at all, and blood was just shown to go from heart to the body and back through arteries and veins respectively, it was classified as a single-loop model. The few models that could not be classified by this scheme were coded as "Unclassifiable."

3. Results

The results are presented in two sections: learning outcomes and protocol analyses. In the learning outcomes section we report performance on the posttest including the diagram, mental model, declarative, and inference tests. In the protocol section we report data on learners' verbalizations as they explained or compared diagrams. These data provide insight into the kinds of cognitive processes participants engaged in during the learning phase. We set the alpha level at .05 for all main effects, interactions, and planned comparisons (Keppel, 1991). We calculated effect sizes (eta squared, η^2) for all significant main effects, interactions, and planned comparisons. We followed the guidelines by Cohen (1988; see also Olejnik & Algina, 2000) according to which effects are regarded as small when $\eta_p^2 < .06$, medium when $\eta_p^2 < .14$, and large when $\eta_p^2 > .14$.

3.1. Learning outcomes

3.1.1. Diagram test

Recall that only those participants who had a single-loop model to begin with were included in the experiment. The drawings that participants generated after the learning session were coded for the type of mental model that they had after instruction. Two raters independently coded the diagrams. Each rater listened to the participants' explanations while drawing the diagram, because not all participants drew equally detailed diagrams, and some details were captured only in the explanations. Inter-rater reliability was high; Kappa = 0.94. Fig. 2 shows the proportion of participants who generate each type of model. Since none of participants generated the single-loop model, this category is not represented in the figure.

Participants in both conditions showed evidence of learning from instruction because no one generated the single-loop model at posttest. However, the two conditions differed in the proportion of drawings for each model type. As predicted, the participants in the compare condition generated a higher proportion of the correct double-loop drawings than those in the explain-diagram condition (90% vs. 64% respectively). We conducted a chi-square test comparing the number of correct (double-loop) diagrams to the number of incorrect diagrams (partially correct double-loop, single-loop with lungs, and unclassified) generated by each condition. Analysis revealed a large effect of instruction, χ^2 (2, $N = 58$) = 7.12, $p < .05$; phi = .29. Consistent with the prediction in Hypothesis 1, this result shows that participants in the compare group underwent more systemic conceptual change than participants in the explain-diagram group. Next, we examine whether the comparison group showed similar advantages on the mental model questions at posttest.

3.1.2. Mental model test

The *mental model score* was based on the proportion of correct solution components out of 20. Fig. 3 shows the adjusted mean posttest scores for the two conditions after controlling for their pretest scores. A one-way analysis of covariance (ANCOVA) was conducted to examine the effect of instruction on the mental model posttest performance using the pretest scores as a covariate. The analysis revealed a large effect of the covariate, $F(1, 55) = 20.14$, $p < .05$, $\eta_p^2 = .27$, showing that participants pretest scores significantly predicted their posttest scores. In addition, there was a medium effect of instruction, $F(1, 55) = 4.00$, $p = .05$, $\eta_p^2 = .07$, with participants in the comparison condition scoring higher than participants in the explain-diagram condition. Thus, consistent with our prediction in Hypothesis 2a, this result shows that comparing diagrams led to more systemic mental model revision than explaining the expert model. Next, we examine how the two conditions performed on the declarative knowledge test.

3.1.3. Declarative questions

The *declarative question score* was based on the proportion of correct answers out of 12. Fig. 3 shows participants' adjusted mean posttest scores after controlling for their pretest scores. An analysis of covariance (ANCOVA) was conducted to determine the effect of instruction on the posttest performance using the pretest scores as

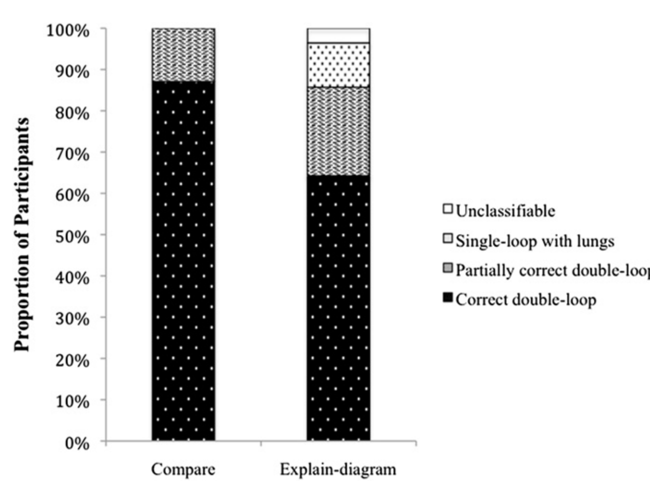


Fig. 2. Proportion of participants' diagrams for each category by condition.

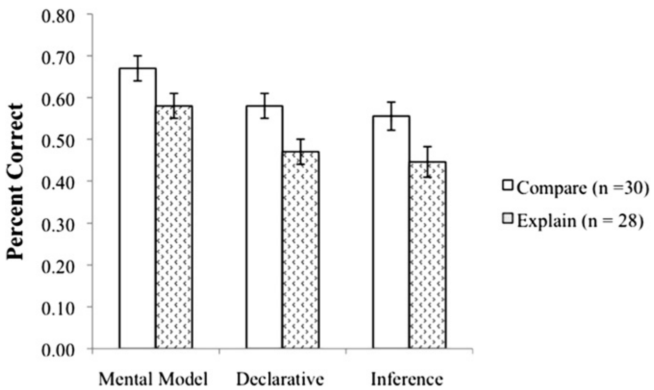


Fig. 3. Adjusted mean posttest performance and standard errors on mental model and declarative questions, and mean posttest performance on knowledge inference questions.

a covariate. The analysis revealed a large effect of the covariate, $F(1, 55) = 32.78$, $p < .05$, $\eta_p^2 = .37$, showing that participants pretest scores significantly predicted their posttest scores. In addition, there was a medium effect of instruction, $F(1, 55) = 5.80$, $p < .05$, $\eta_p^2 = .09$, with participants in the comparison condition performing better than participants in the explain-diagram condition. Thus, in contrast to our predictions in Hypothesis 3, this result shows that participants comparing the diagrams showed greater declarative learning than participants explaining the expert diagram even though both groups read the same instructional text.

3.1.4. Knowledge inference questions

The *knowledge inference score* was based on the proportion of correct components out of 25. Inspection of the means shows that participants in the compare condition scored higher than those in the explain-diagram condition (see Fig. 3). Since the knowledge inference test was given only at posttest, we conducted a one-way ANOVA to examine the effect of instruction on inference questions. Similar to the findings for the mental model and declarative questions test, the analysis revealed a medium effect of instruction, $F(1,56) = 4.86$, $p < .05$, $\eta_p^2 = .08$, with participants in the compare condition performing better than participants in the explain-diagram condition. Thus, consistent with our prediction in Hypothesis 2b, this result shows that the compare condition generated more correct inferences than the explain-diagram group. In the next section, we examine process data in the form of participants' verbal protocols to determine whether the effects of comparison can be attributed to the kinds of processes that students engaged in.

3.2. Verbal protocol analysis

In the verbal protocol analysis, we first examined the total number of constructive statements generated by each condition across all types of constructive statements. Fig. 4 presents the average number of constructive statements and standard errors for each condition, which includes all the fine-grained categories.

We first compared the overall number of constructive statements generated by each condition, followed by separate ANOVAs for each of the sub-categories. Students in the compare condition generated more cumulative constructive statements than the participants in the explain-diagram condition, $F(1, 42) = 17.59$, $p < .05$, $\eta_p^2 = .35$. Supporting our prediction in Hypothesis 4, this result shows that being prompted to compare the flawed and expert models facilitates more constructive activity than being prompted to self-explain the expert model alone. Also consistent with our prediction in Hypothesis 4, we found that participants in the compare condition made many more comparison statements than those in the explain-diagram condition. This provides both a manipulation check of the function of including an external representation of the flawed model to facilitate comparison as well as evidence that being prompted to self-explain an expert model does not result in the spontaneous generation of comparison statements. It is possible that students in the self-explain condition could have compared the correct diagram to their own mental model and made statements such as "I had thought the blood goes to the lungs oxygenated, but I can see that it is not oxygenated." We observed only one instance across all participants of such a comparison. This suggests that without an external representation, it is not likely that students will generate comparisons, and notice the different relational structures between the models, focusing instead on an individual propositional level.

In addition, the participants in the compare condition also generated marginally more function-related explanations than those in the explain-diagram condition, $F(1, 42) = 3.37$, $p = .07$, $\eta_p^2 = .07$. This result provides some support for our prediction in Hypothesis 5 showing that the comparison is more likely to promote the generation of inference statements about the specific function of the different parts of the circulatory system. This is consistent with the notion that the function-related statements describe the interrelations between features and may be critical (or indicative) of a correct mental model. Finally, consistent with our prediction in Hypothesis 6 there was also a significant difference in the number of revision statements, with the explain-diagram condition generating more than the compare condition $F(1, 42) = 5.64$, $p < .05$, $\eta_p^2 = .19$. This suggests that although the explain group did not explicitly

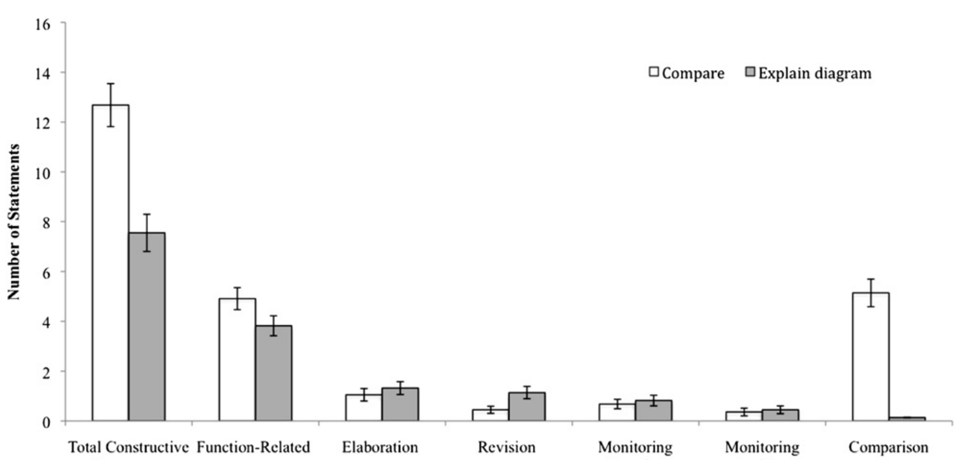


Fig. 4. Mean and standard errors for number of statements of each category.

mention their flawed model they did engage in correction of propositions they made about the correct model. This result suggests that the explain-diagram group engaged in knowledge revision at the false belief (or proposition level) and were able to make the change at the mental model level only insofar as they could learn from the accumulation of these revised beliefs.

4. Discussion

We hypothesized that analogical comparison would be more effective than self-explanation in facilitating conceptual change of a mental model. We tested this hypothesis in a laboratory experiment in which students who had a flawed mental model of the circulatory system were prompted to either compare their model to an expert model, or simply explain the expert model. After completing the prompting task, both groups read a text about the circulatory system and were assessed on a number of dependent measures. We expected analogical comparison prompts to make the interrelations between the features of the model salient, thereby facilitating systemic change rather than revision of single flawed propositions. Overall, our hypotheses were supported, and we observed that analogical comparison was more effective in driving mental model change as shown by students' performance on the diagram and inference tests.

The most stringent test of conceptual change was participants' performance on the diagram test. Consistent with the prediction in Hypothesis 1, approximately 90% of the learners in the compare condition successfully generated the correct model after instruction as compared to about 64% in the explain-diagram condition. The performance of the learners in the explain-diagram condition was consistent with past research that has shown that prompted self-explanation in the domain of the circulatory system led to conceptual change in approximately two-third of the learners (Chi, de Leeuw, et al., 1994; Chi, Slotta, et al., 1994). Our results show that an instruction that embodied holistic confrontation (i.e., compare and contrast) was even more effective in facilitating conceptual change at the mental model level than this already moderately successful approach. Further, results for Hypothesis 2a show that participants in the compare condition performed significantly better than the explain-diagram group on the mental model test. These results are consistent with the interpretation that the process of holistic confrontation facilitates noticing the relations between the different propositions and features of the model leading to system-wide mental model revision.

The declarative questions assessed propositional knowledge. As we described in the introduction, prior work has shown that a student can have a flawed mental model and yet acquire new correct propositions and incorporate them into their flawed model. Therefore, we predicted in Hypothesis 3 that the participants in the two conditions would perform similarly on this measure, because continuing to hold a flawed mental model did not necessarily mean that the student would acquire fewer correct propositions. However, in contrast to this prediction, we found that the participants in the compare condition outperformed the participants in the explain-diagram condition. This suggests that the learning activity (i.e., comparing diagrams versus explaining the expert diagram) differentially prepared the participants to learn from the subsequent instructional text. Past work on *preparation for future learning* has shown similar results for students showing benefits from an initial learning activity when learning new material, whether it was a lecture or a worked-out example (Schwartz & Bransford, 1998; Schwartz, Bransford, & Sears, 2005; Schwartz & Martin, 2004). For example, Schwartz and Bransford found that students given a comparison activity before a lecture better prepared them to learn from that lecture as compared to others who did not compare but read a summary of the data features. The authors hypothesized that the advantage of the compare group resulted from their acquisition of a more articulated or differentiated knowledge structure that they later used to understand and organize the information presented in the lecture. Our result is consistent with this interpretation in that the participants in the compare condition presumably revised their mental model creating a more differentiated and complex knowledge structure that they later used to help understand and organize the declarative knowledge presented in the text.

On the knowledge inference questions, as predicted in Hypothesis 2b, the compare condition performed significantly better than the explain-diagram condition. We predicted that comparison would lead to noticing multiple features and would provide an opportunity to generate inferences about those features. This interpretation is also supported by the results for Hypothesis 5, which show that the compare group made more function-related inferences during learning than the explain-diagram group. In sum, the participants that engaged in comparison and contrast activities were more likely to acquire the correct mental model, generate inferences supported by such a model, and learn declarative knowledge about that model compared to participants self-explaining an expert model.

It was not simply the amount of constructive processing that differentiated the two groups, but also the type of knowledge that was revised. Targeting constructive processes at individual beliefs versus a mental model has implications for the amount and type of subsequent knowledge change. This was supported by the findings for Hypothesis 4 and 5 in the protocol analyses, which showed that the compare condition generated more comparison and function-related statements, whereas consistent with Hypothesis 6, the explain-diagram group led to more statements revising false beliefs. By engaging in comparison, learners were able to focus on the key features and relations of the expert model, and therefore were more successful in acquiring a more expert-like representation.

A key design feature of our instruction was that the compare condition saw two diagrams, whereas the explain-diagram condition only saw the expert diagram. This raises the question of whether the type of instructional technique (compare versus explain) and model presentation (only expert versus flawed + expert) were potentially two different factors driving the effects. Since we were interested in testing differences in the two types of cognitive processes on conceptual change our instructional manipulations were designed to maximize the probability of students either engaging in analogical comparison *or* self-explanation. Had we given students in the explain-diagram condition both models simultaneously they may have spontaneously generated comparisons, making this condition indistinguishable from the compare condition. Furthermore, the pretest results provide evidence that participants in the explain-diagram condition also had the same flawed model as those in the compare condition, just not explicitly in a diagram form in front of them. Our manipulation enables us to test claims of the effect of each type of cognitive process on conceptual change because those processes were clearly distinct as the data from the verbal protocols show. That said, future work should further examine these two aspects to determine the contribution of each to conceptual change and analogical comparison.

We also chose not to tell the participants in the explain-diagram condition that their prior model was incorrect because this also might have invited spontaneous analogical comparisons. If the explain-diagram group made comparisons to their implicit model (prompted by the incorrect prompt) they would have effectively engaged in the same cognitive process as the analogical comparison group thus making it difficult to compare the effect of the two types of processes on conceptual change of a mental model. Future work should address whether there is an effect of simply telling students that their model is incorrect and furthermore how this instruction impacts the likelihood of engaging in analogical comparison. We hypothesize that students who are told their initial model is incorrect would perform better than those who are not while self-explaining because they would be more likely to engage in analogical comparison. Another related issue for future research to address is the role of motivational variables such as students' interests, goals, and epistemological understanding. In the present work we have considered conceptual change only from a cognitive viewpoint, but we acknowledge that motivational considerations also play an important role in determining whether or not conceptual change can take place (see Pintrich, Marx, & Boyle, 1993 for a detailed review of the role of motivational factors in conceptual change). Some past work has explored the effect of motivational factors such as emotional variables and goals on analogy generation (Blanchette & Dunbar, 2001). Future work should address how this interacts with the different levels of representations, and whether motivational variables are more likely to impact one kinds of representation change over others.

We identified students with a specific misconceived model, and encouraged holistic confrontation by the means of analogical comparison with an expert model. This constructive activity helped students align their prior mental model with the correct model and helped students revise them at a system-wide level. Although we demonstrated this with a particular type of misconceived mental model (i.e., single-loop), we hypothesize that this finding would generalize to other misconceived models such as the no-loop or the ebb-and-flow models (as described in Chi, de Leeuw, et al., 1994; Chi, Slotta, et al., 1994), because students could conceivably engage in the same cognitive processes that while making such comparisons, even though their initial mental model may be different.

This research extends past work on the effect of analogical comparison on learning and applies it to another important area of education, conceptual change. This adds to a body of work in science education literature addressing the design of interventions, which draw upon findings in psychology about learning mechanisms and processes (see Ruthven, Laborde, Leach, & Tiberghien, 2009 for a recent review). Based on our results, we can conclude that analogical comparison may be particularly helpful in bringing about conceptual change, when this change requires considerable restructuring of the relations between features of a students' prior model. This can have important implications for instruction, particularly in science education. Current science curricula do not stress the processes of comparing and contrasting one's own prior knowledge with scientific models (e.g., Richland, Zur, & Holyoak, 2007). Incorporating analogical comparison in instruction will be helpful and effective way of achieving radical restructuring of students' prior science concepts.

Analogical comparison can also augment learning from cognitive conflict, which by itself has found to be of limited efficacy in bringing about conceptual change (Limón, 2001). Analogical comparison can provide students the necessary resources to restructure their prior knowledge rather than simply confronting them with conflicting information. When comparing one's own prior model to the correct model, a student has to deliberately engage in constructive processes of inference generation and knowledge revision aimed at the interrelationships of the features and functions of the model. This helps students discover gaps in their own knowledge and reconcile them with the new, scientific information, rather than rely on accretion of revised individual beliefs to accumulate into a correct mental model.

We found that analogical comparison can be successfully employed to facilitate conceptual change of a mental model in the domain of biology. Future research should also examine whether such an approach would facilitate overcoming other types of misconceptions in other domains. In addition, further research should explore the relationship between comparing and contrasting and explaining. Past research has shown that not all students are good self-explainers and given that the process of comparing and contrasting generated more constructive statements in the current experiment, it would be interesting to see whether engaging in such an activity is more beneficial to poor self-explainers. Moreover, it will be important to provide more extensive training in self-explanation and analogical comparison before students engage in those tasks, and then look at how it impacts their performance. Finally, for greater external validity, it would be important to replicate this study in actual classrooms as opposed to a controlled laboratory setting and with different populations such as middle school and high-school age students.

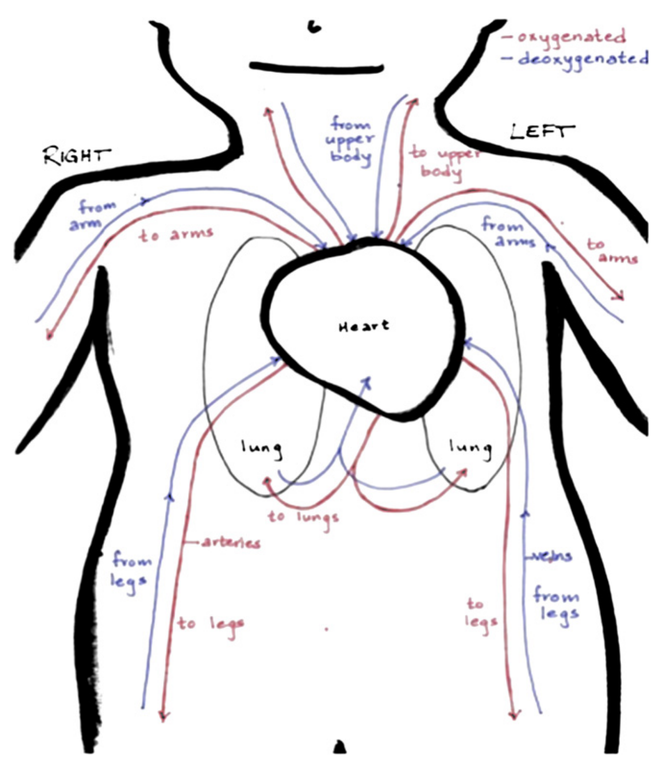
Acknowledgments

We would like to acknowledge the contributions of Marguerite Roy and Kirsten Butcher in the conceptualization of this project and their help with data collection.

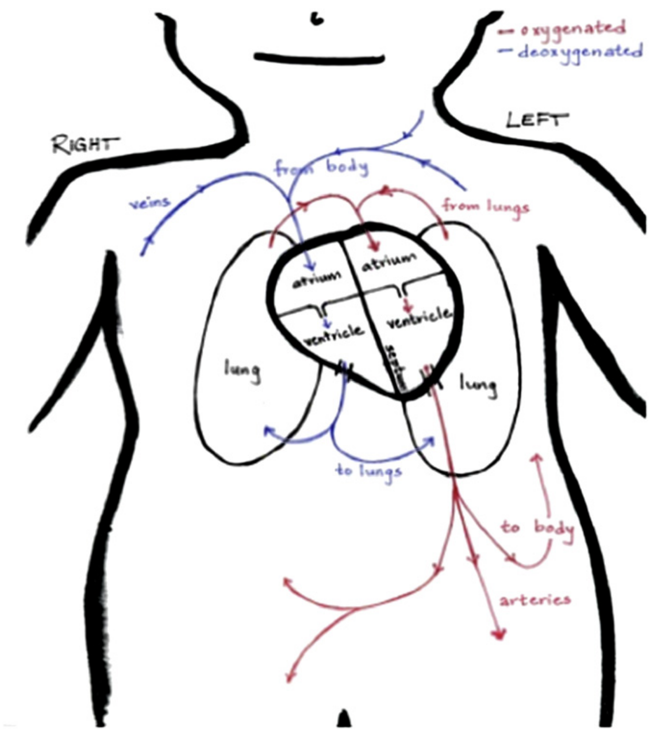
Appendix A. Mental model questions

1. Describe in a few lines the path of the blood in the circulatory system.
2. What types of blood vessels are present in the circulatory system and what are their functions?
3. Describe the structure of the heart in a few lines and explain the functions of each part.
4. What are important components of the circulatory system and what role do they play in circulation?
5. What are the primary and secondary functions of blood? 6. What is the main function of the heart?

Appendix B. Single-loop model (Flawed)



Appendix C. Double-loop model (Correct)



Appendix D. Experimental text (Shier et al., 2006)

1. The heart is a hollow, cone-shaped, muscular pump.
2. The heart pumps 7000 L 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 9 cm 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. These 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 loop.
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 E

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?"

Appendix. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.learninstruc.2011.06.002.

References

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*, 221–252. doi:10.1207/s1532690xci1302_3.

Blanchette, I., & Dunbar, K. N. (2001). Analogy use in naturalistic settings: the influence of audience, emotion, and goals. *Memory & Cognition, 29*, 730–735. doi:10.3758/BF03200475.

Blanchette, I., & Dunbar, K. (2002). Representational change and analogy: how analogical inferences alter target representations. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 672–685. doi:10.1037/0278-7393.28.4.672.

Brown, D., & Hammer, D. (2008). Conceptual change in physics. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 127–154). New York, NY: Routledge.

Carey, S. (1991). Knowledge acquisition: Enrichment or conceptual change? In S. Carey, & R. Gelman (Eds.), *The epigenesis of mind: Essays on biology and cognition* (pp. 257–292) Hillsdale, NJ: Erlbaum.

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, DC: American Psychological Association.

Chi, M. T. H. (2000). Self-explaining expository texts: The dual process of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161–238). Mahwah, NJ: Lawrence Erlbaum Associates.

Chi, M. T. H. (2005). Commonsense conceptions of emergent processes: why some misconceptions are robust. *Journal of the Learning Sciences, 14*, 161–199. doi:10.1207/s15327809jls1402_1.

Chi, M. T. H. (2008). Three kinds of conceptual change: Belief revision, mental model transformation, and ontological shift. In S. Vosniadou (Ed.), *Handbook of research on conceptual change* (pp. 61–82). New York, NY: Routledge.

Chi, M. T. H., de Leeuw, N., Chiu, M. H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. *Cognitive Science, 18*, 439–477. doi:10.1016/0364-0213(94)90016-7.

Chi, M. T. H., Siler, S. A., Jeong, H., Yamauchi, T., & Hausmann, R. G. (2001). Learning from human tutoring. *Cognitive Science, 25*, 471–533. doi:10.1016/S0364-0213(01)00044-1.

Chi, M. T. H., & Slotta, J. D. (1993). The ontological coherence of intuitive physics: Commentary on diSessa's "Towards an epistemology of physics". *Cognition & Instruction, 10*, 249–260. doi:10.1207/s1532690xci1002&3_5.

Chi, M. T. H., Slotta, J. D., & de Leeuw, N. (1994). From things to processes: a theory of conceptual change for learning science concepts. *Learning and Instruction, 4*, 27–43. doi:10.1016/0959-4752(94)90017-5.

Clement, J. (1993). Using bridging analogies and anchoring intuitions to deal with students' preconceptions in physics. *Journal of Research in Science Teaching, 30*, 1241–1257. doi:10.1002/tea.366030100.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Erlbaum.

Dagher, Z. R. (1994). Does the use of analogies contribute to conceptual change? *Science Education, 78*, 601–614. doi:10.1002/sce.3730780605.

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*, 335–356. doi:10.1016/S0361-476X(02)00039-5.

diSessa, A. A., Gillespie, N. M., & Esterly, J. B. (2004). Coherence versus fragmentation in the development of the concept of force. *Cognitive Science, 28*, 843–900. doi:10.1016/j.cogsci.2004.05.003.

Duit, R., & Treagust, D. F. (2003). Conceptual change: a powerful framework for improving science teaching and learning. *International Journal of Science Education, 25*, 671–688. doi:10.1080/09500690305016.

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

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: a general role for analogical encoding. *Journal of Educational Psychology, 95*, 393–408. doi:10.1037/0022-0663.95.2.393.

Gentner, D., & Stevens, A. L. (Eds.). (1983). *Mental models*. Hillsdale, NJ: Lawrence Erlbaum.

Gick, M., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15*(1), 1–38. doi:10.1016/0010-0285(83)90002-6.

Hummel, J. E., & Holyoak, K. J. (1997). Distributed representations of structure: a theory of analogical access and mapping. *Psychological Review, 104*, 427–466. doi:10.1037/0033-295X.104.3.427.

Hynd, C., & Alvermann, D. E. (1986). The role of refutation text in overcoming difficulty with science concepts. *Journal of Reading, 29*, 440–446. http://www.jstor.org/stable/40025804 Retrieved from:.

Hynd, C. R., McWhorter, J. Y. V., Phares, V. L., & Suttles, C. W. (1994). The role of instructional variables in conceptual change in high school physics topics. *Journal of Research in Science Teaching, 31*, 933–946. doi:10.1002/tea.3660310908.

Johnson-Laird, P. N. (1980). Mental models in cognitive science. *Cognitive Science, 4*, 71–115. doi:10.1207/s15516709cog0401_4.

Kendeou, P., & van den Broek, P. (2007). The effects of prior knowledge and text structure on comprehension processes during reading of scientific texts. *Memory and Cognition, 35*, 1567. doi:10.3758/BF03193491.

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

Kurtz, K. J., Miao, C. H., & Gentner, D. (2001). Learning by analogical bootstrapping. *Journal of the Learning Sciences, 10*, 417–446. doi:10.1207/S15327809JLS1004new_2.

Limón, M. (2001). On the cognitive conflict as an instructional strategy for conceptual change: a critical appraisal. *Learning and Instruction, 11*, 357–380. doi:10.1016/S0959-4752(00)000.

Medin, D. L., Goldstone, R. L., & Gentner, D. (1993). Respects for similarity. *Psychological Review, 100*, 254–278. doi:10.1037/0033-295X.100.2.254.

Nokes, T. J., & Ross, B. H. (2007). Facilitating conceptual learning through analogy and explanation. In L. Hsu, C. Henderson, & L. McCullough (Eds.), *Physics Education Research Conference* (pp. 7–10).

Nokes, T. J., Hausmann, G. M., VanLehn, K., & Gershman, S. Testing the instructional fit hypothesis: the case of self-explanation prompts, Instructional Science, in press doi:10.1007/s11251-010-9151-4.

Nokes, T. J., & VanLehn, K. (2008). Bridging principles and examples through analogy and explanation. In *The Proceedings of the 8th international conference of the learning sciences*. Mahwah: Erlbaum.

Ohlsson, S. (2009). Resubsumption: a possible mechanism for conceptual change and belief revision. *Educational Psychologist, 44*, 20–40. doi:10.1080/00461520802616267.

Olejnik, S., & Algina, J. (2000). Measures of effect size for comparative studies: applications, interpretations, and limitations. *Contemporary Educational Psychology, 25*, 241–286. doi:10.1006/ceps.2000.1040.

Pintrich, P. R., Marx, R. W., & Boyle, R. A. (1993). Beyond cold conceptual change: the role of motivational beliefs and classroom contextual factors in the process of conceptual change. *Review of Educational Research, 63*, 167–199. doi:10.3102/00346543063002167.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: toward a theory of conceptual change. *Science Education, 66*, 211–227. doi:10.1002/sce.3730660207.

Renkl, A. (1997). Learning from worked-out examples: a study on individual differences. *Cognitive Science, 21*, 1–29. doi:10.1207/s15516709cog2101_1.

Richland, L. E., Zur, O., & Holyoak, K. J. (2007). Cognitive supports for analogies in the mathematics classroom. *Science, 316*, 1128–1129. doi:10.1126/science.1142103.

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*, 561–574. doi:10.1037/0022-0663.99.3.561.

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* (pp. 37–53). Hillsdale, NJ: Erlbaum.

Ruthven, K., Laborde, C., Leach, J., & Tiberghien, A. (2009). Design tools in didactical research: Instrumenting the epistemological and cognitive aspects of the design of teaching sequences. *Educational Researcher, 38*, 329–342. doi:10.3102/0013189X09338513.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16*, 475–522. doi:10.1207/s1532690xci1604_4.

Schwartz, D. L., Bransford, J. D., & Sears, D. (2005). Efficiency and innovation in transfer. In J. Mestre (Ed.), *Transfer of learning: Research and perspectives* (pp. 1–51). Greenwich, CT: Information Age.

Schwartz, D., & Martin, T. (2004). Inventing to prepare for future learning: the hidden efficiency of encouraging original student production in statistics instruction. *Cognition and Instruction, 22*, 129–184. doi:10.1207/s1532690xci2202_1.

Shier, D., Butler, J., & Lewis, R. (2006). *Hole's essentials of human anatomy and physiology*. Boston: McGraw Hill.

Smith, J. P., diSessa, A., & Roschelle, J. (1993). Misconceptions reconceived: a constructivist analysis of knowledge in transition. *The Journal of the Learning Sciences, 3*, 115–163. doi:10.1207/s15327809jls0302_1.

Tippett, C. D. (2010). Refutation text in science education: a review of two decades of research. *International Journal of Science and Mathematics Education, 8*, 1–20. doi:10.1007/s10763-010-9203-x.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4*, 45–69. doi:10.1016/0959-4752(94)90018-3.

Vosniadou, S. (2008). Conceptual change research: an introduction. In S. Vosniadou (Ed.), *International handbook of research on conceptual change* (pp. 127–154). New York, NY: Routledge.

Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. J. Hacker, J. Dunlovsky, & A. C. Graesser (Eds.), *Metacognition in educational theory and practice* (pp. 277–304). Mahwah, NJ: Lawrence Erlbaum Associates.

Zimmerman, B. J., & Campillo, M. (2003). Motivating self-regulated problem solvers. In J. E. Davidson, & R. J. Sternberg (Eds.), *The psychology of problem solving* (pp. 233–262). Cambridge University Press.