In R. Glaser (Ed.), *Advances in Instructional Psychology*. Mahwah, NJ: Lawrence Erlbaum Associates (in press).

# Self-explaining: The dual processes of generating inferences and repairing mental models

**Michelene T.H. Chi**

**Learning Research and Development Center**
**University of Pittsburgh**
**Pittsburgh, PA 15260**
**Chi@vms.cis.pitt.edu**

Sept. 15, 1998

The author is grateful for the support from the Spencer Foundation for this research by grants #199400132 and #199500102. The paper was written in part while the author was a Fellow at the Center for the Advanced Study in the Behavioral Sciences. The author would like to thank Heisawn Jeong, Stephanie Siler, Roger Taylor, and Cindy E. Hmelo, for comments on an earlier draft; Kurt VanLehn, an anonymous reviewer, and Bob Glaser for comments on later drafts.

## Table of Contents

*Tell me and I forget.*

*Teach me and I remember.*

*Involve me and I learn.*

—Benjamin Franklin 1706–1790

The dream of psychologists and educators has always been to identify skills or strategies that can be used across domains. The pursuit of domain-general strategies largely characterized the literature in the seventies. It was then that general memory strategies (such as rehearsal or method of loci) and problem-solving strategies (such as means-ends analysis) dominated the experimental studies in psychology and simulation research in artificial intelligence. These studies reached conclusions such as the following: memory retrieval is facilitated when one uses a strategy; memory retrieval improves with age because children are increasingly better at using memory strategies; and experts are better problem solvers because they use sophisticated means-ends strategies (i.e., you find the next solution step by reducing the difference between the goal state and the initial state).

The enthusiasm for domain-general strategies was challenged in the eighties by work on expertise. This research showed that people with expert knowledge in a particular domain have not necessarily acquired more skillful use of strategies. Instead, it seems that their domain knowledge bypassed the need for general strategies, or else their strategic knowledge is not the source of their superior performance. In the developmental arena, Chi & Koeske (1983) and Chi (1978) have shown that young children can remember and recall just as many items as adults in a memory retrieval task if the items are drawn from a content domain in which the child knows something, such as dinosaurs or chess. Likewise, Chi, Feltovich and Glaser (1981) have shown that it is not the use of a means-ends strategy that allows experts to solve problems so readily; rather, experts have richly organized knowledge of the problems (analogous to problem schemas) that allow them to represent the problems in such a way that the solutions became transparent. Hence, experts are not solving problems successfully because they can apply and use strategies skillfully. Rather, the solutions become apparent as soon as they represent the problems correctly. Thus, how well one represents a problem depends on a person's domain-relevant knowledge rather than strategic skill.

Although the aforementioned types of studies challenged the utility of domain-general strategies, in the nineties, the attention has been turned away from strategies that are effective for *performing* a task (such as problem solving and remembering) and to the idea that there may be domain-general activities for *learning*. That is, surely there may be *ways of learning* that are more effective and can be beneficial across domains. Engaging in these learning activities might enable some individuals to become experts while others remain novices. In order to differentiate the conventional performance strategies from learning ones, the former will be referred to as *strategies* whereas the latter ones will be referred to as *activities.* These two terminologies also fit our intuitive notions of strategies and activities. One uses strategies, as if they are a kind of rule that can be applied in a particular circumstance, whereas one engages in activities, and the engagement itself promotes learning. These differences, although subtle, are important.

A learning activity that seems to be domain-general, effective for learning both procedural and conceptual type of domains, easily used, and beneficial to students of all abilities, was identified and described in Chi, Bassok, Lewis, Reimann, and Glaser (1989) and Chi, de Leeuw, Chiu, and LaVancher (1994). The discovery of this learning strategy hinges on an assumption about learning, which is that new knowledge (whether declarative or procedural in nature) cannot be readily and perfectly assimilated (or encoded) by the student from direct instruction, either in the form of *listening* to a teacher's explanations, or in the form of *reading* from a textbook. Instead, the acquisition of new knowledge requires the students to be actively involved in the *construction* of their own knowledge.

What does being constructive mean? Although the concept of constructivism has been around since Piaget, it has been discussed theoretically in an epistemological way (von Glasersfeld, 1984, 1989) and defined generally as a contrast to either passive learning (in which students merely encode and store what is presented) or to instructionism (in which the research agenda focuses on new ways for teachers to instruct, Papert, 1991). Construction is a very broad term denoting both the external behavioral or activity aspects of learning as well as the internal processes of cognitive reorganization (Cobb, 1994).

Evidence for constructivism is indirect. On the behavioral side, to be constructive merely means to be actively doing something while learning. By this definition, engaging in any externally observable activities such as drawing a diagram, solving problems, answering and asking questions, designing, summarizing, and reflecting, can be broadly construed to be constructive, as opposed to passively sitting there assimilating instruction. Being constructive in this activity sense does facilitate learning, as intervention studies repeatedly show that students who were required to construct knowledge (by engaging in any of the activities listed above) are shown to learn more and perform better than students who did not actively engage in any of the above-listed activities. Thus, methods such as reciprocal teaching (Palinscar & Brown, 1984), collaboration (Webb, 1989), question-asking (King, 1992), can all be construed as interventions that engage the students in constructive activities, and such engagement produces more effective learning.

The externally observable activity of constructivism obviously corresponds to some internal processes of reorganization. Without specifying what these internal processes of reorganization are, one can nevertheless show, again indirectly, that constructivism occurs. Such evidence comes from studies that show students have alternative conceptions that are qualitatively different from the ones held or taught by the teachers or the textbooks. The existence of misconceptions suggests that these alternative understandings must have been constructed by the learners themselves. Thus, there is abundant indirect evidence showing that engaging in constructive activities facilitates learning, and abundant indirect evidence showing that students do construct alternative conceptions. However, no evidence showed a direct link between engaging in a constructive activity and the processes of reorganization.

The constructive activity to be focused on in this paper is self-explaining, which is the activity of explaining to oneself in an attempt to make sense of new information, either presented in a text or in some other medium. The self-explanation effect, to be described in Section II, provides evidence of a direct link between a constructive activity and knowledge re-organization. The goal of this paper is to specify what the relationship is between the processes of self-explaining and knowledge reorganization. This specification is illustrated by a microgenetic analysis of four self-explanations of a single subject. This qualitative analysis revealed the underlying mechanism of self-explaining to be the process of *revising* one's knowledge structure or mental representation. This revision process is to be contrasted with the process of generating inferences, which was the process assumed to underlie self-explaining in Chi, et al. (1989, 1994). (The term "revision" will be used henceforth instead of the more general term "reorganization" because the latter term has implications for the more radical kind of knowledge change, commonly referred to as conceptual change. See the later discussion under Two Caveats in Section IV of this paper.)

The paper has four sections. The first section introduces and defines self-explaining (Chi & Bassok, 1989). The second section describes the phenomenon of self-explaining in learning a procedural skill (Chi et al, 1989; Chi and VanLehn, 1991) and a conceptual domain (Chi et al, 1994). The identification of this phenomenon will be complemented with largely quantitative analyses of verbal data. The third section describes an alternative possible process (revision) by which self-explaining fosters learning and constrasts it with the process entertained earlier (inferencing). This revision process is proposed as a result of qualitatively analyzing the verbal utterances generated by a single student. The fourth section addresses several additional pertinent issues. (For readers who know the definition and the phenomena, they may skip Sections I and II.)

## Self-explaining: Definitions and Examples

### Terminologies

Five terms have been used in previous papers to refer to different aspects of the self-explanation research. The term *self-explaining* refers to the *activity* of generating explanations to oneself, usually in the context of learning from an expository text. It is somewhat analogous to elaborating, except that the goal is to make sense of what one is reading or learning, and not merely to memorize the materials (as is often the case when subjects in laboratory experiments are asked to elaborate). In this sense, self-explaining is a knowledge-building activity that is generated by and directed to oneself. Additional comparisons to elaborations will be discussed later.

The generic term *self-explanation* (SE, singular) refers to a unit of utterances produced by self-explaining. That is, it is any content-relevant articulation uttered by the student after reading a line of text. A unit of SE may or may not contain a *self-explanation inference* (SEI), which is an identified piece of knowledge generated in the SE that states something beyond what the sentence said explicitly. Thus, an SEI is *a piece of new knowledge or inference*. This means that for any given SE, a content analysis may reveal several other types of statements besides inferences, such as paraphrases, monitoring statements, and nonsensical statements. (In our codings, we typically do not differentiate the form of the statement from the content. For example, an inference or a monitoring statement can be phrased either as a question or an assertion.) Because the focus of this research is often only on the content, monitoring statements and questions will be counted as an SEI if they contain knowledge inferences. Thus, in general, an SEI refers to a specific piece of knowledge within a unit of SE that has been coded as an inference.

The term *self-explanations* (SEs, plural) refers more generally to the entire corpus of collective utterances or verbal protocol data gathered from self-explaining in a particular study. Finally, the term *protocols* will be used more generally to refer to any kind of verbal data, including giving definitions or answering questions on the pre- and post-tests, as well as SEs.

## Examples of Self-explanations

Examples provide the best way to understand what an SE is. After some examples are presented below, additional clarifications will be made of what self-explanations are. Below are self-explanations taken from two different domains. One set of examples is taken from self-explanations generated while eighth grade students were studying the topic of the human circulatory system from explaining a biology text (Chi et al., 1994). The second set of examples is taken from college students self-explaining a worked-out example solution in a college-level physics text (Chi et al, 1989). The sentences in the text are always underlined with an identification of the sentence number such as S17, and the SEs are always in quotes. Student initials are in the parenthesis. The codings on the right tally how many SEIs are coded from the quote, and each SEI is italicized. Below are examples of self-explanations generated by three students, after having read the same sentence S17 in the biology text.

S17: The septum divides the heart lengthwise into two sides.

(#1) (AW) "...the septum, it sort of...um...would divide the heart so that you can like...*distinguish between the two parts.*" —1 SEI

(#2) (BB) "Well, what's a septum? I mean is it, um *a muscle? A bone?* You know,um, *an organ? I don't think it's an organ , though* —3 SEI

(#3) (MW) "...it's probably not like a wall, maybe *like a barrier*... probably *things can go through it*... I think it's probably *not like a solid wall."* —3 SEI

There are several features to notice about these SEs. First of all, the italicized parts of each SE gives the reader a sense of what constitutes an SEI. Take SE#1. The part that is italicized would constitute an explanation inference because it is stating something that the sentence did not say. The sentence merely discusses the fact that the septum is a divider of the heart in a lengthwise direction. The inference here is that such division allows one to distinguish the two parts. Although this is a commonsense inference (any divider allows one to distinguish the two parts), it is important to make it here, since it happens to be an important feature of the heart. (In fact, about a third of SEIs are generated by bringing in commonsense knowledge. This will be discussed again later.) For this SE#1 then, we would code it as having one SEI. In SE#2, the queries of whether the septum is a muscle, a bone, or an organ, are coded as three SEIs since three separate analogies were made. We could also code the last comment, the rejection of it being an organ, as another SEI, but we did not in this case, to be on the conservative side. Instead, we merely considered it to be redundant with the preceding query, analogous to an utterance such as “Is it an organ or not an organ?” In general, we tended to be conservative in our coding, and not over-attribute any utterance as an inference unless new knowledge is produced.

In SE#3 above, one could consider *barrier* to be distinct from *wall* since walls seemed to imply solidity whereas barriers can be penetrable. However, we only counted the *wall* reference as one SEI even though it was mentioned twice. Notice in SE#3 that when a phrase is repeated, as in the case of *not like a wall* and later *not like a solid wall*, we generally coded that repetition as a single SEI. In this case, we could have coded the second comment as a separate SEI since the adjective “solid” has been added. But again, in order not to overly attribute inferences to the student (assuming in this case that wall meant solid in the first place), this SE would be counted as containing three SEIs.

The next example shows two SEs generated by the same student in sequential order, each after reading a sentence:

S17: <u>The septum divides the heart lengthwise into two sides</u>.

(#4) (NH) “...so...there’s two sides of the heart and the thing that divides it is the septum, lengthwise...” —0 SEI

S18: <u>The right side pumps blood to the lungs, and the left side pumps blood to the other parts of the body.</u>

(#5) (NH) “So, the septum is a divider so that the *blood doesn’t get mixed up,* so the septum is *like a wall* that divides the heart into two parts...” —2 SEI

The important feature to notice about SEIs from this example is that they are not necessarily generated on the basis of information confined to a single sentence. SEIs can be an integration across information provided in multiple sentences. SE#4, generated after reading S17, is basically a paraphrase and did not contain any inferences, so it would not be coded as a SEI. However, after reading S18, the two inferences generated in SE#5 integrated information provided in both S17 and S18. That is, S17 talks about the septum being a divider, and S18 talks about the destinations to which blood is pumped from each side of the heart. One of the SEIs embedded in SE#5 is that the septum's purpose is to prevent the blood from the two sides from mixing, a very important inference.

The next example shows SEs taken from attempts to understand a physics worked-out example taken directly from a college physics text (Chi et al., 1989). The example starts with a description of the problem (which includes a diagram of three strings connected in a knot, with a body of mass W hanging from one of the string, as shown in Figure 1). Here we show three consecutive SEs by the same student:

Insert Fig. 1 about here

The diagram shows an object of weight W hung by strings.

Consider the knot at the junction of the three strings to be the body.

(#6) "*Why should this [the knot] be the body? I thought W was the body*."

The body remains at rest under the action of the three forces...

(#7) "I see. *So, the W will be the force and not the body*. OK."

. . . of the three forces shown in the diagram.

(#8) "*Un huh, so...so they refer to the point as the body*..."

Each of the quotes was considered a single SE in this earlier study. Note that the student had not realized that the knot (i.e., the intersection of the three strings) was the body,because people's naive conception normally considers the mass to be the body. Hence, the student had to work through and revise that conception.

## Grain Size and Format

The examples of SEs presented above also illustrate the approximate grain sizes of our codings. Earlier, in the physics work, SEs were coded at the level of a sentence or multiple sentences if they referred to the same idea. In the biology study, we coded SEIs at the phrase level. And finally, we have also coded the biology protocols at a more fine-grained proposition level, and the results provided basically the same pattern as the phrase level (see Chi et al., 1994). However, becausewe are mainly interested in knowledge inferences, coding at the grain size of the phrase seems to be more at the knowledge level, and the inference is more sensible, whereas coding SEs at a more fine-grained proposition level sometimes gets redundant. More detailed discussion of coding and grain size issues can be found in Chi (1997b).

Note finally that coding at the knowledge level is independent of coding for the format of the SEs. Coding for the format or structure of the explanations would consist of characterizing whether the student is making an analogy, posing a conjecture, or making a metacognitive statement. The latter two examples can be seen in SE#2, where the student is posing the conjecture that a septum might be like a muscle, a bone, or an organ. The student also rejected his last conjecture. This is a metastatement. An analogical SEI can also be seen in SE#5, where the student analogized the septum to a wall. The focus of the discussion in this paper is not on the format of the SEs, but only on the content.

## Context of Self-explaining

The research on self-explaining has been done in the context of students learning from texts, usually expository kinds of texts on a complex science subject matter. But in principle, self-explaining can be undertaken in any learning context, not necessarily learning from text. For instance, one could self-explain while examining a bus on the street, such as wondering why some buses (the connected ones) have an accordion midsection, as my young son once thought aloud to himself. Basically, self-explaining is a process that the learner uses to help him/herself understand external inputs that can be instantiated in any medium (e.g., text or video). The focus of the discussion throughout this paper is on learning some new domain of knowledge, with the assumption that there are some external materials that we want the learner to think about and acquire. For the sake of parsimony, a text will be referred to as the generic external input.

One could also self-explain or reflect without any external inputs. In this case, it would be like thinking to oneself. Finally, although in our studies, students self-explain overtly, in principle, one could self-explain and think covertly. We needed the students to self-explain overtly for the pragmatic reason that protocol data can be collected.

## What Self-explaining is Not

### Self-explaining versus Talking or Explaining to Others

In principle, self-explaining was differentiated from the act of merely talking, because talking may not produce any content-relevant inferences, even though in practice, talking often does produce SEIs. However, self-explaining should be a more focused activity than talking: The focus is on trying to understand the learning material and make sense of it, whereas talking is conveying information to a listener. Talking and explaining to others has the added demand of monitoring the listener's comprehension. However, self-explaining and talking and explaining to others do share similarity in that they are both constructive activities, and some researchers simply code the amount of talking as a substitute for SEs or SEIs (see Teasley, 1995, for example). Another difference between self-explaining and explaining to others is that SEs or SEIs need not be complete and coherent explanations in the Hempel (1965) sense, whereas there is an implicit demand for coherence when explaining to others. For example, suppose one was asked to explain how natural selection works or how a cut heals. Answering either of these two questions would require a coherent explanation, perhaps explicating the causal mechanism, otherwise the explanation would not constitute an answer. A self-explanation is not a complete and coherent answer in this sense. A self-explanation can be partial, fragmented, and at times, incorrect. These differences between *Self*-explanations and *Other*-explanations are important because they serve different goals: Self-explaing serves the goals of revising one's understanding, whereas explaining to others serve the goal of conveying information. They both facilitate learning, however, probably because they are both constructive activities (see findings in tutoring for example, where the tutors benefit as much as the tutees, probably from having to explain to the tutees; or findings from collaboration research).

### Self-explaining versus Thinking Aloud

Although one could think of self-explaining as thinking aloud, it is also important to differentiate it from what has traditionally been called "think-aloud protocols" in the problem-solving literature (see Ericsson & Simon 1993; Chi, 1997c). Think-aloud protocols, often collected in the context of problem-solving research, is a method of collecting verbal data that explicitly forbids reflection. Think-aloud protocols presumably ask the subject to merely state the objects and operators that s/he is thinking of at that moment of the solution process. It is supposed to be a dump of the content of working memory. The analysis of such think-aloud protocols thereby reveals the search space and search operators that the subject is using. Contrasts between think-aloud protocols and self-explaining utterances are described in greater detail in Chi (1997b). Self-explaining is much more analogous to reflecting and elaborating than to "thinking aloud." Talking aloud is simply making overt whatever is going through one's memory (see Ericsson & Simon, 1993), without necessarily exerting the effort of trying to understand. A recent Ph.D. thesis compared students who were prompted to either talk aloud or to self-explain, and found self-explaining to produce greater learning gains (Wathen, 1997).

### Self-explaining versus Elaborating

Elaboration is a broad concept that can be conceived of either as a strategy or as an inference. Research that conceives of elaboration as a strategy is the kind that typically asks students to create a relationship among concepts, whether or not they are meaningful, usually for the sole purpose of remembering the concepts. In fact, the more bizarre the elaborated relationships among concepts are, the more memorable the materials become. Clearly, this form of elaboration is not appropriate for learning with understanding.

Research that conceives of elaborations as inferences sometimes supplies inferences for the students, and different types of elaborations are examined to see which ones are more effective for remembering. Elaboration has also been studied in the context of learning. In these contexts, students are either asked to elaborate or else a variety of text modifications is undertaken to incorporate elaborations. Whereas elaboration has been perceived as an intervention that either a researcher can undertake (by inserting inferences into the text materials) or as a strategy that a subject can undertake (to create relationships and embellishments either between two words or between sentences), self-explaining, on the other hand, only makes sense if it is undertaken by the subject. Thus, SEI is a specific kind of elaboration, the kind that is conceived of as an inference rather than a strategy.

Notice that students' interpretation of what they are supposed to do under a self-explanation instruction may not be fundamentally different from what they do under a general elaboration instruction. The instruction we have used to prompt students for self-explanation is shown below:

*We would like you to read each sentence out loud and then explain what it means to you. That is, what new information does each line provide for you, how does it relate to what you've already read, does it give you a new insight into your understanding of how the circulatory system works, or does it raise a question in your mind? Tell us whatever is going through your mind—even if it seems unimportant.*

Thus, although there may be some differences in the instructions given to the subject about elaborating versus self-explaining (especially in the extreme case when students are asked to generate meaningless and bizarre elaborations), from the learner's point of view, conventional elaboration and self-explanation instructions may be more-or-less the same. The difference lies mainly in the researcher's conceptualization of them. To be succinct and contrastive, one could say that elaboration researchers conceive of elaborations (either supplied by the experimenter or the learner) as serving the purpose of improving an *imperfect text*, whereas self-explanations should be thought of as serving the purpose of improving one's *imperfect mental model (or representation).* Thus, in the imperfect mental model view, it would not make sense to have the researcher or experimenter supply the elaborations. Alternatively, one could say that self-explanation is a form of elaboration (only the ones that are supplied by the learner), but not all elaborations are self-explanations (since some elaborations include the ones supplied by the experimenter). These differences will be further elaborated later.

### The Self-explanation versus the Generation Effect

The self-explanation phenomenon is reminiscent of the generation effect discussed two decades ago (Slamecka & Graf, 1978). The generation effect is a phenomenon uncovered in a traditional paired-associate memory test. The design is for subjects to either read word pairs such as *rapid-frequent*, or for the subjects to generate the associated word given the stem word, an initial letter, and a rule. So, suppose the rule is synonym, and the stem and initial letter given are: *rapid-f*. In this generation condition, the subject was asked to generate the associative word on the basis of the rule and the initial letter. In this case, the majority of the subjects would generate the word *fast*. It turns out that memory for the associated words are significantly better if they were generated rather than simply read.

Actually, a more related study to the self-explanation effect is one by Stein and Bransford (1979), who studied subjects' memory for a target word embedded in a sentence, such as "The *fat* man read the sign." Recall is tested by the sentence frames with a blank for the target word *fat*. Consistent with traditional elaboration research, they elaborated each sentence frame with additional information, such as "The fat man read the sign that was two feet high" or "The fat man read the sign warning about thin ice." As expected, recall was better for elaborated sentences than non-elaborated, fostering the traditional view that elaborating a text can help memory. However, the more important finding was that if the subjects were asked to generate elaborations themselves, then their recall was even better. This suggests that providing one's own elaboration somehow facilitated access. The generation effect in the context of memory retrieval and the self-explanation effect in the context of learning share one thing in common: They are both constructive activities. To that extent, they facilitate memory retrieval and learning, possibly because being generative means one is being more attentive and actively laying down memory traces.

### Self-explanation Inferences versus Inferences

How are SEIs different from other types of inferences? SE and a SEI can be thought of as knowledge inferences and can be contrasted with four types of inferences that are commonly studied in the psychological literature. First of all, SEI is not a bridging type of inference, the kind commonly studied in the comprehension literature. A bridging inference is one whereby a referent is explicitly provided. For example, if a sentence had referred to the septum as “it” and the student explicitly refers to it as “the septum,” then that is providing a bridging inference. By our analysis, such bridging inference provides no additional knowledge about septum, so it is not counted as a piece of knowledge inference. Second, an SEI is not a paraphrase because paraphrasing often does not add new knowledge, as shown in SE#4. Third, nor is a SEI a logical-type of inference, as the kind commonly studied in the psychological literature on inferencing from quantifiers, such as All men are mortals, if John is a man, is John a mortal?

A fourth kind of commonly studied inferences is schema-based inferences. Schema-based inferences refer to the supply of inferences based on prestored knowledge. In many ways, one cannot call these inferences; they are more analogous to retrieval of prestored knowledge. For instance, if a reader was asked to read the following sentences:

(1) John sat down and talked to the waiter.

(2) Five minutes later, the food arrived.

an inference that is needed to comprehend the second sentence is that John had ordered some food. Such an inference can easily be supplied by the reader. To do so, the reader merely activates and retrieves a prior script. One can ascertain that this piece of missing inference in the text is already a prestored piece of knowledge because we can ask the reader, prior to reading such sentences, what one does in a restaurant, and the reader will typically say that a diner sits down, orders food, eats, and pays the bill. Thus, prior to reading, the reader already knows the script or schema for what happens at restaurants; so that from the perspective of the text, this is an inference; but from the perspective of the reader, this is not an inferred piece of new knowledge. A researcher can disambiguate whether an inference is retrieved or constructed by first assessing the subject’s prior knowledge.

To recap, SEIs have been contrasted with four kinds of inferences commonly studied in the literature.  SEIs are unlike all of them in that SEIs are pieces of new knowledge constructed/generated by the students.  How these knowledge inferences can be generated when the students are learning a new domain will be considered later.

## The Self-explanation Effect: The Phenomenon

In this section, two of our studies showing the self-explanation effect will be briefly described. The two studies were carried out in two different domains, in biology (on a passage about the human circulatory system consisting of 101 sentences) and in physics (Chapter 5 of Halliday & Resnick, 1981, on mechanics).  The details of these studies have been published elsewhere, so the goal here is to highlight their similarities and differences for comparison purposes, and more importantly, to address the crucial individual differences finding that exists in both sets of data.

### The Physics and the Biology Studies

The design of the two studies are basically the same.  Each study has three phases.  The initial phase consists of some kind of assessment of both the students' prior knowledge about the domain topic and some indication of their general ability.  The learning phase consists of studying the text (or examples in the text) and explaining while they study.  The final phase consists of a post-learning assessment, always taken one week later to make sure that the learning was long term.  The design of both studies are laid out in Figure 2.  We will refer to the Chi et al. (1989) study as the physics study, and the Chi et al. (1994) study as the biology study.

Insert Figure 2 about here

The differences between the two studies are laid out in Figure 3:  acquiring declarative (system-related) knowledge versus procedural knowledge; using eighth graders versus college students as subjects; explaining expository text versus worked-out example solutions; answering questions in the post-tests versus solving problems.  The most important difference between the two studies is that in the physics study, the self-explanation effect was obtained by encouraging students to spontaneously generate SEs, whereas in the biology study, students were prompted (thus enforced) to self-explain.

A second important difference was that the biology study had a control group of students who were not asked to self-explain but were asked to read the same text twice. Reading the text twice took about as much time as reading it once along with self-explaining.

Insert Figure 3 about here

## **The Findings**

In reporting the results, the unit of analysis for the physics study was at a coarser grain than the biology study, primarily because the students in the physics study were not enforced to self-explain. This means that whenever they did explain, we only had to judge whether what they said contained ideas relevant to and extending beyond the materials they were studying. Every relevant and extended idea was counted as an SE. In the biology study, on the other hand, because the students in the prompted group were enforced to self-explain after every line, this means that they could have just stated nonsense, so we had to decipher whether there was an SE *inference* within each articulation. This required a meticulous coding to determine whether a SEI was embedded in each articulation. The way an SEI was coded has already been discussed in the Definition section above. To put it another way: in the physics protocols, whenever a student voluntarily said something, what they said was either an SE or not an SE, in which case the utterances could either be some kind of metastatement ("I don't know what's going on here") or some kind of mathematical statement. In the biology protocols, because the students were enforced to self-explain, there were lots of utterances, but only some of them were SEIs.

### Individual Differences or Range of Self-explanations Generated

In both studies, there was a continuous range in the number of SEs generated by the students, even though in the one case it was generated spontaneously and in the other case it was enforced. In the physics case, the average number of SEs generated by each student while studying a worked-out example problem ranged from a low of less than 2 to a high of more than 25. The number of SEIs generated in the biology case ranged from a low of 7 to a high of 111 for the entire passage of 101 sentences on the circulatory system. In both cases, it was a continuous range rather than a bimodal split, even though our analyses often refer to the students as either high or low explainers.

### Learning Correlated with the Number of Self-explanations

In both studies, learning correlated with the number of self-explanations generated: That is, the greater the number of self-explanations generated, the better the students learned. Instead of correlating the number of self-explanations generated and the number of problems solved correctly (because of low *N*), an analysis based on a median split of the number of SEs generated shows that in the physics study, students who generated, on average, 16 SEs per example solved 86% of the problems correctly, whereas students who generated on average 3 SEs per example solved only 42% of the problems correctly. All differences to be reported here are statistically significant. Henceforth, students who generated a large number of explanations, on average, will be considered the high explainers, whereas students who generated fewer number of SEs will be referred to as the low explainers.[1]

The correlation between learning and self-explaining can be seen in two ways in the biology study. First, students who were prompted to explain learned more than students who were in the control group (those who were not prompted to explain, but did read the text passage twice). The prompted group gained 26% from the pre-test to the post-test on answers to the assessment questions, whereas the control group gained only 16%. However, such difference in the amount of gain between the two groups is impressive considering that (a) this passage is taken from a very well-written text, (b) the control group did read the passage twice, and moreover, (c) there undoubtedly

[1] Note that if the numbers reported here vary slightly from the published report, that is because here, we report the results in terms of high and low explainers, whereas the published paper (Chi, et al., 1989) reported the results in terms of the successful and less successful solvers. In the original study, we did a median split on the basis of the number of problems solved correctly, then considered the number of explanations generated as the dependent variable. Here, we divided the students into two groups on the basis of the number of SEs they generated, then compared the number of problems they solved correctly. Because the results are correlational, the pattern of the data is the same.

were some spontaneous (albeit covert) explainers in the control group as well, and most importantly (d) the difference between the two groups gets more pronounced on the harder questions. This last result suggests that the prompted explainers understood the materials more deeply than the non-prompted group of learners. This means that generating self-explanations per se was beneficial, assuming that the self-explanations generated contained inferences.

A second way to show the benefit of self-explaining is to contrast the high and the low explainers within the prompted group. As in the physics result, the high explainers (those who generated, on average, 87 SEI) learned considerably more than the low explainers (those who generated, on average, 29 SEIs). The high explainers answered significantly more questions correctly (78%) than the low explainers (61%). This difference again became even more pronounced for the harder questions. Notice that Wathen's findings (1997) replicated this result in that her prompted to self-explain group learned more than the group who was simply prompted to talk-aloud. One could interpret her result to mean that talk-aloud did not generate as many knowledge inferences as self-explaining, so that the talk-aloud group is comparable to the low explainers.

The individual difference results in the number of self-explanations generated, either spontaneously (in the physics case) or enforced (in the biology case) raise two important questions. On the surface, the intervention question one immediately thinks of is *how* one can encourage students (the low self-explainers) to generate more SEs or SEIs. However, a deeper question might be *why* some students generated more SEIs than others. Understanding the processes of self-explaining (to be discussed below) may shed light on understanding this *why* question.

Taken together, the results of the two studies show that generating self-explanations per se is useful for enhancing learning, since the prompted group did learn more than the unprompted group in the biology study, and the high explainers (whether spontaneous or enforced) learned more than the low explainers in both studies. Moreover, the more inferences the students' self-explanations contain, the better they learn. This suggests that generating self-explanations is useful in general (because it's a constructive activity), but it's even better if one could encourage students to generate SEs that contain lots of inferences. In order to understand whether or not this can be done, we need to understand the processes of self-explaining that can account for such individual differences.

### Robustness and Generality of the Phenomenon

Although our findings of the self-explanation effect are primarily correlational, the phenomenon is robust because many other researchers have found similar results, using a variety of manipulations and research designs. Moreover, the domains that other researchers have explored expanded beyond the ones we have studied (mechanics and the circulatory system) to include different domains such as learning LISP coding (Pirolli & Recker, 1994), electricity and magnetism (Ferguson-Hessler & de Jong, 1990), and probability (Renkl, 1997). We have shown that self-explaining is effective for both college students and eighth graders, and Siegler (1995) has extended this to five-year-olds, in the context of asking them to explain the experimenter's reasoning for a number conservation task.

## The Influence of Prior Knowledge and Ability in Understanding the Self-explanation Effect

Before advancing reasons for why and how self-explaining works in enhancing learning (to be discussed in the third section of this paper), we should first consider what can cause individual differences in the amount of self-explanations generated. Assuming that all students have the goal of trying to learn (at least in the context of our laboratory study), two possible factors, prior knowledge and ability, are tentatively rejected.

Prior knowledge can be divided into four types: domain-specific knowledge, domain-relevant knowledge, misconceptions, and domain-general world knowledge. Domain-specific knowledge refers to knowledge that is directly related to the content of the materials. Prior domain-specific knowledge was controlled first in a general way (by selecting students who had not taken college courses in the relevant topic, for example) and then assessed more specifically in a number of pre-tests. In the physics study, domain-specific knowledge was controlled and assessed in the following ways. First, students were selected to participate in the physics study only if they had the kind of profile we were looking for, such as not having taken any college physics courses, but having taken one high-school physics course. Second, to equate for the amount of knowledge students gained from the first three chapters of the physics text, all students had to master the materials by being able to solve the problems at the end of each chapter to a preset criterion.  Otherwise they had to re-read the chapter and re-solve the end-of-the-chapter problems.  Third, to assess how much knowledge they did acquire from the expository part of the target Chapter 5, they were asked to define Newton's three laws *after* they read that part of Chapter 5 but *before* they self-explained the worked-out solution examples.  Similarly, prior domain-specific knowledge about the circulatory system was assessed by having the students define terms in the pre-test, prior to reading the passage, to see if they knew, for example, that the atrium is a chamber of the heart.  Sometimes we also asked students to draw the path of blood flow, as well as answer some pre-test questions.  Hence, prior domain-specific knowledge was usually thoroughly assessed in multiple ways.

In both the biology and the physics studies, there were no differences between the high and the low explainers in terms of their prior domain-specific knowledge.  For example, in the biology study, there were no significant differences in the pre-test scores between the prompted group (39%) and the unprompted group (42%).  Likewise, in the physics study, the most sensitive prior specific knowledge assessment was the definition of Newton's three laws.  Newton's three laws were scored by decomposing each law into 3-5 subcomponents.  Thus, prior to the study of the examples embedded in Chapter 5, both the high and the low explainers scored 5.5 of a possible 12 components for their definitions of Newton's Laws.  Overall, in both studies, there were no noticeable differences among the students in their prior domain-specific knowledge.

A second type of prior knowledge is domain-relevant knowledge.  This refers to general knowledge that is relevant to understanding the domain under study (as in the circulatory system), such as knowing that distance is inversely related to pressure or that lungs are proximal to the heart, as opposed to domain-specific knowledge such as knowing what a septum is.  These latter two pieces of knowledge are relevant for making inferences and understanding why only the pulmonary vein has no valves, because lungs are so proximally close that the pressure coming from the heart will remain high for such a short distance.  In a current yet unpublished study (see Chi, 1997c, for a preliminary discussion), we administered to students a post-test of domain-relevant knowledge after they learned the unit on the human circulatory system, and again, we found no differences between the high and low explainers in their domain-relevant knowledge in the context of tutoring.

A third kind of prior knowledge is the amount of misconceptions or false beliefs that students initially possess. This knowledge is the complement of students' domain-specific knowledge. That is, domain-specific knowledge assesses what correct knowledge about the human circulation a student knows. False beliefs and misconceptions assess how much incorrect knowledge the student has about a domain. False beliefs are of the type such as believing that the heart oxygenates blood, and misconceptions are of the type such as conceiving of heat as a kind of substance rather than as a kind of process. Although our data are scant, our physics data show that high explainers do have fewer initial misconceptions (10.0) than low explainers (13.3), as assessed by a test presenting problems and questions of the kind "Which car will suffer more damage in a head-on collision, the heavy Cadillac or the small Volkswagen?". In the biology data, the high explainers also had slightly fewer initial false beliefs (2.25) than the low explainers (3.00). However, neither differences reached statistical significance (perhaps because of our low N) even though they were both in the right direction, so we would have to conclude that the number of misconceptions and false prior beliefs do not correlate with the quantity or quality of self-explanations generated.

What about domain-general world knowledge? One way to assess domain-general world knowledge is to consider tests such as the California Achievement Test (CAT) as a measure of domain-general knowledge. Again, in the biology study, there were no significant differences in the CAT scores of the students of the prompted (87%) versus the unprompted (83%) groups. If one assumes that the CAT measures one's general knowledge, then we can again conclude from these results that general world knowledge did not correlate with learning from self-explaining. Thus, none of the measures of the four types of prior knowledge could have predicted differences in learning from self-explaining.

What about ability, can it be a cause of differences in the amount of self-explanations generated? Ability was assessed in a number of different ways, such as more generally by students' grade point averages. In general, there were no differences between groups in their general profile, such as the number of science courses taken, or the grade-point averages.  Within each group, ability was further assessed by specific tests.  In the biology study, if we want to consider CAT as a measure of ability, then we can differentiate students on the basis of their CAT scores, and show that those who scored higher on the CAT (98%, with 0.8 SD) did not benefit more from self-explaining than those who scored lower on the CAT (72%, 17.4 SD).  Both groups gained about 35% from the pre-test to the post-test.

Ability in the physics study was assessed by the Bennett Mechanical Ability test (because it correlates with success in domains such as physics).  The assumption was that scores on this kind of test would predict ease of learning physics-related materials. There was absolutely no difference between the successful and less successful solvers in their performance on the Bennett Mechanical Ability test.  Both groups scored around 22-23 points out of a possible 29 for the mechanical test.

In general, these results suggest that individual differences in the number of  self-explanations generated is not due to prior knowledge or ability.  Prior knowledge has been examined in terms of prior domain-specific knowledge, prior domain-relevant knowledge, prior false beliefs and misconceptions, and prior general world knowledge (if one considers the CAT to be basically a measure of general world knowledge).  In all these cases, there were no significant differences in any of these measures between the high and the low explainers.  Thus, it appears not to be the case that students who self-explained more often have more prior knowledge or have higher ability.  Rather, it appears that the mere act of self-explaining fosters greater learning.

## Two Contrasting Approaches to Understanding the Self-explaining Effect

Why does self-explaining facilitate learning?  To say that self-explaining has a positive effect on learning because it’s a constructive activity not only fails to explain the internal processes or mechanism of construction, nor does it explain the fundamental individual differences result, that some students generate more SE inferences than others (whether spontaneously or enforced), and thereby learn more.  Thus, the ideal explanation needs to identify the processes of self-explaining such that the processes also account for the finding of variance in the amount of SEIs generated.

The internal processes corresponding to self-explaining were tacitly implicated in the way we coded and interpreted the SE protocols in the two earlier studies. The perspective we took in our earlier analyses conceived of self-explanations as the product of self-explaining. That is, self-explaining was conceived of as a constructive activity that generated inferences, and these inferences were determined on the basis of information that was missing from the text sentences (thereby missing in one's mental model). Thus, the perspective taken in the earlier work (Chi et al., 1994), was to think of self-explaining as the process of generating inferences. This perspective thus made three implicit assumptions: that 1) the text is incomplete, 2) the goal of self-explaining is to generate inferences to fill the omissions in the text, and 3) the "omissions" in the text correspond to the "gaps" in one's mental model, thus implying a direct correspondance between the model conveyed by the text (henceforth referred to as the text model) and the learner's mental model. That is, the inference-generating perspective implicitly assumed that learning is directly encoding materials presented in a text, and the result of the encoding is a mental model that is isomorphic to the text model. When the text is incomplete, generating inferences will fill gaps in one's encoded mental model.

The inference-generating view was supported in three broad ways: 1) by the success in our coding method, 2) by the empirical results, and 3) by the postulation of inference mechanisms that showed how it was plausible to generate new knowledge without being told (either by the text or by an agent, such as a teacher). With respect to our coding method, we have already shown how coding was undertaken, how an utterance was counted as a self-explanation if it contained knowledge beyond what the text sentences stated. Thus, SEs were treated as a product (the inferences) of self-explaining. Even though the inferences were capturable externally in the SE protocols, it was assumed that they were stored and internalized in the mental representation as well. In this view, a direct link between a constructive activity (self-explaining) and the external results of some internal processes (the inferences) was established. The implicit assumption was that these generated inferences filled gaps in the students' mental representations, much as they filled omissions in the text. Obviously then, a mental representation that contained more inferences (with fewer gaps) had to be better than a mental representation with fewer inferences (and more gaps). Despite the fact that this inference-generating perspective was able to establish a direct link between a constructive activity and the resulting product (a more enriched, correct, and coherent mental model), this perspective cannot explain why some students generated more SEIs than others. Below, we first detail this perspective, and then present an alternative perspective.

## Generating Inferences: The Incomplete Text View

The hypothesis that self-explaining is the process of generating inferences beyond information contained in the text sentences tacitly assumed that the text is incomplete in some ways. There are two kinds of incomplete text: a poorly written text and a well-written text. A poorly written text can be incomplete in omitting anaphoric references and/or connective ties between sentences, thus destroying structural coherence. Making a text more structurally coherent facilitates comprehension in general for all learners (Kintsch & Vipond, 1979). A poorly written text can also be explanatorily incoherent (Kintsch & van Dijk, 1978), in the sense that some crucial piece(s) of background information is left unstated. When crucial pieces of information are left unstated, not surprisingly, subjects with high-domain knowledge can supply their own, so that they learn just as well from an explanatorily incoherent text, but low-domain knowledge subjects obviously will profit more from having this explanatory background information supplied (McNamara, Kintsch, Songer, & Kintsch, 1996).

However, the kind of texts we use in our studies are typically very well-written and coherent, ones that have been popular and frequently used in the classrooms. Thus, an incoherent text is not the source of learning failures in the kind of learning situations that we have been considering. More over, an incomplete text is not necessarily an incoherent text,. Nevertheless, even if a text has both structural and explanatory coherence, it can contain omissions. Thus, even well-written texts can be incomplete because they cannot possibly contain all the information that can be conveyed about a topic. For instance, suppose we assume that for each component of the circulatory system, such as a valve or the atrium, there are three dimensions that a text can discuss: its structure, its function, and its behavior.  For example, for the valve, a text can describe how it is made (like a flap of tissue), what its function is (to prevent blood from going backward), and what its behavior is (closing and opening depending on the pressure created by the blood flow).  Similarly, a text passage can also describe the structure, function, and behavior of the atrium.  But typically, for each component, not all three dimensions are mentioned in the text.  In the text passage used in our biology study, the functional dimension is omitted about half of the time.  But, aside from these "first order" features about the components, a text can also describe the relations between a feature of one component and a feature of another component.  For example, the text could describe the relationship between the behavior of the atrium (contraction and relaxation) and the function of the valve.  Again, the majority of these kinds of "second order" relations are omitted.  Besides the second-order relations, one can also imagine the text explicating "third-order" relations, such as the relationship between a feature of a component (such as the function of the atrium) and the overall goal of the circulatory system, such as the relationship between the need for the atrium to act as a temporary reservoir in order for blood to circulate with a sufficiently high pressure.  Knowing this kind of relationship allows a student to answer complex "why" questions, such as "Why should the valve of the atrium close for a short time?" Again, these kind of third-order relationships are hardly ever mentioned.  Thus, from this simple exercise of analyzing the omissions in our circulatory text passage, one can only conclude that all texts are incomplete, without necessarily being incoherent structurally or explanatorily.  Thus, it seemed natural to assume that the source of learning failures is the omissions existing in a text, even if it is well written.

## Is an Incomplete Text Detrimental to Learning?

The preceding example of an incomplete text, one that omits some information about the topic, was illustrated in the context of a science text.  In principle, it is no different from narrative texts, as we saw in the sentences "John sat down and talked to the waiter.  Five minutes later, the food arrived."  In a narrative text, much information is omitted as well.  However, in a narrative text, presumably the reader has the prior knowledge to make the correct inferences, whereas, it was not clear how students could generate inferences when they lack the appropriate prior schemas and scripts (to be entertained in the next section). In this section, we recap the empirical evidence that supported the assumption that self-explaining is generating inferences of omitted information. The empirical evidence gives indirect, direct, and causal support.

Indirectly, the results we cited earlier showed not only that explainers learned more than non-explainers, but also that high explainers learned more and deeper than the low explainers. That is, the high explainers not only could answer more questions correctly in general, but they excelled particularly in the hardest questions. This confirmed an inference view because this view assumes that inferences fill gaps in one’s representation, so that the more enriched one’s representation is, the better one has learned.

We were able to support the inference view directly by showing that students were able to induce information omitted from the text (such as the function of an atrium). For example, in SE#5, the student correctly induced the function of the septum, as a device to separate the right and the left chambers. Similarly, in SE#2 and SE#3, the students were trying to induce the structure of the septum, which was again not stated explicitly in the text. To confirm that students were in fact trying to induce the omitted information, we explicitly assessed the students’ knowledge of the functions of 11 components during the reading phase (Chi et al., 1994). The assessment consisted of asking the students to explain what the function of each component was shortly after they had read and self-explained a sentence describing a particular component. The high explainers were significantly more successful at inducing correctly the functions of components (10.5 functions out of 11 components), whereas the low explainers were less successful (7.8 out of 11).

Finally, we were able to support the inference view causally by the way we constructed our questions in the biology study. That is, many of the questions we posed to the students (such as the “why” questions) required knowledge of the function of the components in order to answer them correctly. This probably allowed the high explainers to be better at answering the hardest “why” questions than the low explainers.

Thus, it seemed clear that the information omitted from the text was necessary for deep understanding, as assessed by the students’ ability to answer the missing functions of components, as well as the “why” and other types of complex questions in the biology study. However, because students seemed to be able to “recover” this information by self-explaining, their success reinforced the view that self-explaining serves the purpose of inducing the omissions in the text. More generally, the incomplete text, corresponding to an incomplete mental model view, persisted, with the availability of indirect, direct, and causal empirical evidence.

### Inference-generating Mechanisms for New Knowledge

An incomplete text view assumes that self-explaining plays the role of filling omissions in the text, and that the text model and the mental model are isomorphic in that they both contain the same gaps. If self-explaining is the process of generating inferences to fill gaps while *learning* a new domain, then one needs to postulate what kind of mechanisms can generate inferences when there is no prior knowledge or schemas. That is, what kind of inferencing mechanism can generate new knowledge that can facilitate learning a new domain; we are not talking about comprehension kinds of inferences that make a poorly written text more coherent, such as supplying the anaphoric referents, nor inferences from prior scripts and schema because these are assumed not to exist for learning new domains. Several inferencing mechanisms can be postulated that can account for learning a new domain without prior knowledge.

First, students can produce inferences by integrating information presented across different sentences (as shown in the SE#5, where SE#5 integrates information contained in S17, relating the septum as a divider to information contained in S18 that the different sides of the heart pump to different locations in the body, therefore the septum must serve the purpose of preventing blood from mixing).

Second, inferences can be generated by integrating information presented in sentences with prior (commonsense world or domain-relevant) knowledge, using processes of analogy or any kind of comparison to integrate them. Once a comparison is made, then attributions can be made about the new information on the basis of the properties of the analogized entity. SE#3, shown earlier, exemplifies both of these types of inferencing processes. In SE#3, the inference that the septum is not like a wall was generated by analogizing the septum to some kind of non-wall-like penetrated barrier (commonsense world knowledge). Moreover, once the idea of a penetrated wall is retrieved, then the septum is attributed with the possibility that things can go through it.

A third way of generating inferences is to use the meaning of words to imply what may also be true. Again, SE#1 illustrates this. In that SE, the septum is inferred to...*distinguish between the two parts* . Distinguishing between the two parts might be an inference derived from the semantics of the word divide. That is, one often divides things so that there are two discrete subparts. So this might be an inference generated from the meaning of the word divide.

Fourth, an inference can also be generated by combining any of these inferencing mechanisms. Various permutations of combination of the above three inferencing mechanisms can be proposed. The next self-explanation illustrates the combination of using inferencing from the meaning of a word with inferencing from using commonsense knowledge:

> S22: In each side of the heart blood flows from the atrium to the ventricle.
>
> SE#9 (DA) "Well, so that,...um..,the heart, the, *the atrium is up on the top and the ventricle's on the bottom,...*"

Here, the idea in the text stating that blood flows from the atrium to the ventricle may have activated the commonsense knowledge that liquid-like entities can only flow from a higher to a lower position (due to gravity). Hence, if blood flows from the atrium to the ventricle, then the atrium must be above the ventricle. Thus, in this case, the word flows activated the context in the real world under which such situations can occur, leading to the inference that the ventricle must be on top.

Two important conclusions can be drawn from the above analyses. First, one can conclude that several knowledge inferencing mechanisms can be postulated, and that they can be instantiated in the protocol data. The inference mechanisms illustrated above make it plausible to see how learning of a new complex domain can occur when the learner simply attempts to connect new knowledge with prior knowledge, even if the prior knowledge is non-domain-specific commonsense world knowledge. Thus, postulating these mechanisms takes the mystery out of understanding how learning can occur by self-explaining. The possibility that these three types of inferencing mechanisms (along with their combinations and variations) can construct knowledge that is needed and missing from the text, reinforces the view that self-explaining is an inference-generating process. Second, one can conclude that via these mechanisms, using largely commonsense background knowledge, one can learn a *new* domain of knowledge without being taught. That is, one can learn by explaining to oneself, using one's existing knowledge, without having someone else (a more knowledgeable person) explain to us. These two conclusions seem sound, well-grounded, and counterintuitive, thus lending credibility and popularity to the self-explanation effect.

### Skepticisms About the Inference-Generating View

The preceding section assumes that self-explaining serves the purpose of inferring missing information not explicitly stated in the text sentences, so that information in the text served as the context by which self-explanations were interpreted. This particular view was derived by an analysis of the missing information in the text. This view was then further reinforced by the success in coding for inferences in the self-explanations that were not stated in the text, and further supported by indirect, direct, and causal empirical findings. Finally, this view was further bolstered by the postulation of a variety of plausible inferencing mechanisms that can generate new knowledge in a learning context. However, if generating inferences of missing information is the only mechanism underlying self-explaining, then it should predict several additional secondary results.

First, a content analysis of any well-written text from a normative point of view, as discussed earlier, will show omissions of information distributed throughout the text passage. Thus, in principle, if self-explaining was generating inferences, one would expect a *uniform* distribution of self-explanations throughout the text sentences. However, this prediction was not supported. Figures 4A and 4B show the distribution of self-explanations for the four high explainers, taken from the data collected in the Chi et al. (1989) physics study, for Example 5 and 8 (without being redundant, only two of the three examples will be shown, when data of individual subjects are presented). Instead of a uniform distribution of a small number of self-explanations across all the sentences in the example, the self-explanations seem to be clustered at several key locations. Yet, these key locations cannot be the sites at which crucial information was missing due to explanatory incoherence, since there is little consensus in the loci at which the majority of self-explanations were generated by each student. For example, in studying example 5 in the text (Figure 4A), subject P1 spontaneously self-explained four idea units at sentence 10, whereas the other three high explainers did not self-explain much in that location. Figure 4B shows a similar lack of consensus in the loci for Example 8 from the physics text. Figures 4C and 4D show the distribution of the four poor learners. Again, although the poor learners generated far fewer spontaneous self-explanations, they nevertheless show the same pattern of non-uniform distribution across the sentences of an example. The same pattern can be seen in the biology study in which students were enforced to self-explain. Here, we simply counted the number of lines students uttered, just to get a quick count. For the best learner (Figure 4E), although there is a baseline of explanations distributed across all the sentences (because they were required to self-explain in this study), there is nevertheless a pattern of spikes occurring unpredictably across the sentences. The same pattern can be seen for the poorest learner (Figure 4F). Thus, it does not appear to be the case that self-explanations were generated to correspond to information omitted from the text.

Insert Figures 4A, 4B, 4C, 4D, 4E and 4F about here

Second, the inference-generating view should also predict that when students do explicate an inference at the same location, the inferences should be semantically equivalent presumably because a relevant piece of information is missing. However, this is simply not the case: see again, SEs#1, #2, and #3, and #4, which were generated by four different students at the same location in the text (after reading S17).

A third finding that questions the inference-generating view is that unlike elaborations supplied by the researchers, self-explanations do not always make sense to the analyzers. They are often fragmented, and sometimes incorrect, whereas elaborations supplied by the researchers are always scientifically correct and meaningful. In fact, when a researcher supplies less meaningful elaborations (such as imprecise ones), then typically they are less helpful for recall (Stein & Bransford, 1979). In contrast, the fact that self-explanations are often either incorrect, fragmented, or meaningless to the coders, and yet still helpful to learning, raises the question of what purpose they serve. If these incorrect and fragmented SEs are viewed as incorrect inferences, then in principle they should pose a problem for the learners, in that they should hamper learning somehow. But the fact that about 25% of SEs are erroneous, and yet students nevertheless learn from generating them, suggests that these may not be considered "incorrect inferences" analogous to incorrect elaborations. Again, the fact that incorrect SEs are not damaging to learning suggests that they may serve another purpose.

Thus, not only did we not find explanations generated uniformly across all sentences, because the text passage as a whole omitted numerous pieces of information (such as the omitted information about the function), but moreover, when they do explain at any given location, there was hardly any consensus in what each student said. This pattern of inconsistency in *when* an explanation inference is generated and *what* is explained at each line of text, questions the assumption that self-explaining only serves the purpose of inferring new knowledge that is missing from the point of view of a normative model. Instead, self-explaining seems to serve another purpose, a purpose that is tailored to the individual student him/herself. These unexpected findings, along with the individual differences finding in the differential number of SEI students generate, converge on the conjecture that self-explaining may be more than a process of inferring missing knowledge to fill gaps. The notion to be proposed here presumes that a student comes to the learning situation with some kind of preliminary mental model that can be incomplete and incorrect in many ways (thus perhaps explaining the source of the 25% incorrect self-epxlanations). In this alternative perspective, self-explaining can be conceived of also as processes of revising one's existing mental model of the learning materials. Such a perspective can explain why some students would generate more SEI than others (because the amount of self-explaining depends on their need to revise their mental model, which then depends on the frequency with which they perceive conflicts between their mental models and the text model). Thus, a revision view requires that we re-interpret self-explanations in the context of the student's existing evolving knowledge (or mental model) about the sentences that s/he is reading, rather than interpret self-explanations in the context of the correct scientific knowledge (the normative model) that a text embodies. That is, self-explaining is a process that students engage in for the purpose of customizing inferences to their own need. Hence, the alternative view to be proposed here is that self-explaining also serves the purpose of revision so that the processes of self-explaining allow the individual student to repair his/her own representation. Such a view is still consistent with the more broader constructive knowledge-building view of learning (Chan, Burtis, & Bereiter, 1997).

## Undertaking Revision: The Imperfect Mental Model View

The unexpected findings cited in the previous section make more sense in the context of an imperfect mental model view. Such a view assumes that self-explaining is the process of revising (and updating[2]?) one's own mental model, which is imperfect in some ways. Therefore, it makes sense that the majority of the students would not generate a semantically similar explanation at any given sentence location, because each student may hold a naive model that may be unique in some ways, so that each student is really customizing his/her self-explanations to his/her own mental model. Similarly, it also now makes sense that there would be no consensus in when an explanation would be generated, because presumably one would repair one's mental model only when a "conflict" is perceived. This revision interpretation also accounts for individual differences in the amount of SE students generate: presumably they generate a SE when they perceive a conflict between their mental model and the text model (what the text describes). Hence, although there is no empirical support up to this point to unambiguously assert that mental model revision is one of the process of self-explaining, the imperfect mental model framework at least accounts for all these heretofore unexplained findings coherently.

The imperfect mental model view assumes that students come to a learning situation with

[2] It may be wise to refrain from using the term "updating" here, because it is used in the comprehension literature in a slightly different way than repairing/revising here. Updating has been used in the sense of foregrounding or highlighting a specific part of one's mental model. Thus, updating corresponds to "the current status of objects or events described by the text" (McNamara, Miller, & Bransford, 1991, p. 496). There is no reference in the updating literature to any conflicts between the existing mental model and the text sentences.

different imperfect pre-existing mental models, or build imperfect ones in the course of learning. It also requires that we develop a method to assess what the status of a student's initial mental model is prior to learning, which is a much more complex task than a text analysis. Fortunately, a method for capturing a student's mental model has already been developed and discussed in Chi et al. (1994) and is briefly redescribed in Appendix A. Such an assessment method reveals that students do have pre-existing mental models, and the majority of them are flawed. For the domain of the human circulatory system, we now know that students' initial mental models fall into six types, as shown in Figure A3, with 50% of the initial models being of the single loop type (more details later).

### How to Operationalize Mental Model Gaps and Conflicts (or Violations)

To claim that self-explaining is a process of revision requires that we make assumptions about *when* revisions are required. In order to specify when the process of repair is needed, we first need to clarify how mental models can be imperfect, in what ways do imperfect mental models correspond to or conflict with incomplete text model, and what is the implication of a correspondence or a conflict for the learning processes.

First, a mental model can be imperfect by having gaps of missing knowledge. Such gaps may or may not correspond directly to omissions in the text. (The term "omissions" will be used primarily to refer to information missing from a text passage, and the term "gaps" will be used to refer to knowledge missing from one's mental model.) When gaps do correspond to omissions, no conflicts between the mental model and the text model exist. The tacit assumption of the earlier work (Chi et al., 1994) was that gaps correspond to omissions. In order to account for the prompted or enforced self-explanation effect, we need to make an additional assumption that students are not often aware of omissions (except perhaps for "explanatorily" glaring ones) and/or gaps in their knowledge, unless they are explicitly told to reflect on their understanding. These two assumptions (that gaps correspond to omissions and students are not aware of gaps and omissions) explain why instruction to self-explain is beneficial. That is, one way to interpret the enforced self-explanation effect is that students are basically encouraged, through prompting, to induce the omitted information and gaps of knowledge.

Notice that the issue of whether or not mental model gaps correspond to text omissions, is tangential to the issue of whether or not a mental model is fragmented. Fragmented mental models may be defined to be ones that are missing connections. In much of the text comprehension research, it is often assumed that the goal of a constructive strategy is to make the representation more coherent by connecting ideas in the text (Graesser, Singer, & Trabasso, 1994). One could conceive of making connections as generating inferences that integrate knowledge. In this sense, the concern in the comprehension literature of making a representation more coherent is analogous to the benefit of self-explaining, when gaps correspond to omissions.

Gaps, on the other hand, need not correspond to omissions in the text. In this case, the text explicitly provides the information missing from a mental model. This means that sentences containing information that fill gaps of missing knowledge can simply be assimilated without much self explaining. For example, a sentence (such as S22) which states that <u>In each side of the heart blood flows from the atrium to the ventricles,</u> may simply be encoded directly without much self-explaining because such a sentence may simply fill a gap of missing knowledge. Consequently, whether or not gaps correspond to omissions, no conflicts occur.

Before exploring circumstances under which conflicts occur, we should first discuss what is not a conflict between a mental model and a text model. (Henceforth, the term “conflict” will be used in its generic sense, and the term “violation” will be used to refer to the specific type of conflict discussed here.) A violation cannot be merely a “contradiction”. A contradiction occurs when a sentence conveys some information that the mental model already has represented but incorrectly. Suppose a student thought the size of the heart is about ten inches in diameter and the text sentence says it’s seven inches, then this would constitute a contradiction, not necessarily a violation. A contradiction can usually be corrected without much self-explaining, and the more often a piece of incorrect knowledge is contradicted, the more likely it will be corrected (de Leeuw, 1993). So, the issue boils down to how do we determine whether a text sentence violates a piece of knowledge or merely contradicts it. One way to discriminate between violations and contradictions is to assess the “embeddedness” of a piece of knowledge. A piece of knowledge (i.e., a belief) should be held with greater perseverance if it is highly embedded in a network of beliefs with multiple consequent nodes following that belief (de Leeuw, 1993).

A violation is also not a “disagreement” between two coherent views.  That is, one can recognize that one holds a particular view that conflicts with a text view or a partner’s view, without necessarily feeling the need to revise one’s view.  In this case, a disagreement is between two opposing views that are already learned and held in place.

Finally, a violation is not a “discrepant outcome”.  For example, in many physics experiments, students can clearly see that their predictions (let’s say of where the ball will land after being dropped by a moving airplane) are discrepant with where the ball actually landed.  Such a discrepancy between the ball’s actual and predicted landing location may create a recognition that something is wrong, but the students generally have no idea what is wrong with their mental model as to produce an incorrect prediction. That is, the students have no idea what aspects of the correct text model violate their mental models.

To define a violation requires a second way of conceiving of imperfect mental models.  That is, conceiving of imperfect mental models as ones having gaps assumes that the mental model is correct globally, with incorrect and/or missing knowledge at the local level.  Alternatively, a mental model can be imperfect not only because it has gaps, but in the sense that it is flawed.  A flawed mental model can be defined as one that gives opposing or different predictions from the scientifically correct text model.  A flawed mental model can nevertheless be consistent and coherent, in the sense that it can make the same consistent prediction over time, and the model also has internal consistency.  For example, in the single loop model, lungs are merely conceived of as another organ to which blood has to be supplied.  In such a flawed model, one would not predict that blood goes to the lungs to receive oxygen and get rid of carbon dioxide. Instead, in such a single loop model, the heart is the source of oxygenation, the heart acts as a unitary component, and blood goes to the lungs (as it does to other parts of the body) strictly for the purpose of supplying lungs with blood.  Such a flawed model is distinctly different from the scientifically correct double loop model, because they would make different predictions about the need for blood to go to the lungs.

In the context of a flawed mental model, a violation requires a recognition that a piece of knowledge is violated in some *causal way* , in the sense that this flawed belief has some implication for additional consequences, so that a repair is not merely the case of replacing an isolated incorrect belief with the correct one. In this paper, a violation will be defined as a conflict between a text sentence and a belief that is embedded in a flawed mental model. Thus, a text sentence (S18) that says <u>The right side pumps blood to the lungs and the left side pumps blood to other parts of the body</u> should violate the single loop model, because such a model would not predict that blood needs to go to the lungs for oxygen. This means that a student can recognize that a text sentence violates her belief if she somehow *takes the effort to propagate* the conflict so that she recognizes the consequences of changing that belief. Such effort can be undertaken by self-explaining. Thus, it is possible that, without taking the effort to propagate a conflict, a given sentence that violates a student's belief may be conceived of only as a contradiction.

When violations occur, different learning processes occur depending on whether a student acknowledges the violation, misinterprets it, or rejects it. The process of repair (or *accommodation?)* occurs only when the student recognizes and acknowledges that a violation exists. When it is misinterpreted, then the student can assimilate the conflicting information, usually incorrectly. Typically, students misinterpret a violation to be consistent with their models. For example, the student (RC) below misinterpreted S18 by self-explaining:

> SE#9: "The right side of your heart gives blood to your lungs,
> and the other side, the left side, gives blood to every other
> part of your body."

The use of the word "gives" in this student's explanation clearly conforms to the single loop model in that he has misinterpreted the sentence to mean that blood goes to the lungs merely to supply lungs with blood, and not to get oxygen. Thus, even though this sentence should have violated the student's mental model, no violation was perceived.

Sometimes a sentence that should violate a student's mental model may be misinterpreted as a contradiction or filling a gap. For the same S18 that stated the right side of the heart pumps blood to the lungs and the left side pumps blood to other parts of the body, one student (JP), whose initial naive model is also the single loop without lungs, treated the information in that sentence not as a violation of her model but as contradiction of what she already knew. She explained:

> "I never knew that before. Um, I always thought they all do the same thing.
> Um, nothing else."

This SE shows that the student treated the heart as a unitary component without specialized functions in different parts of it. Moreover, the student treated the information in the sentence not as a violation of her mental model in that blood should circulate to the lungs to pick up oxygen, but rather, she treated the information as conveying new knowledge about the heart as having decomposable compartments with different functions.

From the examples just presented, it is clear that once we have captured a student's initial naive model, and even though we can in principle determine whether a given sentence would or would not violate a student's mental model, it is still not obvious that a student would necessarily interpret the sentence as such. What in principle should be a violation can be interpreted by the student as actually being compatible with their mental model, as shown earlier, but requiring merely direct assimilation or correction. Thus, even though a serious violation existed, a student may not interpret it that way.

### How to Characterize Mental Model Repair

Even though a way of defining violations has been laid out, it is nevertheless problematic to determine when violations are detected by the students. That is, even though a post-hoc analysis can easily show that self-explanations are usually generated in order to resolve a discrepancy in understanding, it is difficult to determine accurately when violations are detected. Three complexities exist. First, in order to predict more accurately when a violation exists, the researcher has to track the changes in the mental model as it *evolves* while the student reads and learns from the text. Second, even if the researcher can accurately predict when a violation should exist in the context of an evolving mental model, a student may not notice it (such as when a sentence is misinterpreted or parsed incorrectly). Finally, even if a violation is recognized at the time that it was introduced, it may not be resolved until much later when more sentences are read.  This means that we cannot accurately predict when a violation is detected, thereby we cannot predict when self-explanations would be generated to resolve it.

Given all these difficulties in tracking when a violation occurs and is perceived, another way to support the claim that self-explanations are in part mental model repairs (in addition to inference generation) is to characteristize the kind of self-explanations that can be observed on occasions when a violation is detected as well as occasions when no violation is detected.  Table 1 lists four characteristics that can be captured.  First and foremost, because repairs should be undertaken when violations are detected, the detection of violations can be characterized by a great deal of uncertainties such as "ums" and pauses.  Second, there should be evidence of monitoring statements of failure to understand.  Third, the concept of repair also means that revisions of the mental model sometimes need to be strengthened.  Repetition in self-explanations may be taken to characterize strengthening of one's mental model, whereas it is difficult to makes sense of the purpose of repetition in an inference generating framework.   Finally, when violations are detected and repairs are undertaken, self-explanations should be lengthy and effortful.  However, if the incoming new information is consistent with one's mental model so that no violations are detected (but only gaps), then presumably the new information can be assimilated readily or ignored, reflecting short confirmatory self-explanations.  In the next section, an example will illustrate how it might make more sense to consider self-explaining as a process of revision.

Insert Table 1 about here

Note that the proposal is not that each of the above characteristic absolutely differentiates a revision view from an inference view; but rather, that there will be a differentiation in the characteristic of a self-explanation, depending on whether a conflict is detected or not, at the time the self-explanation was articulated.  There is no a priori reason on the other hand, to expect differential characteristics of self-explanations, such as long versus short ones, or effortful versus confirmatory ones, viewed from an inference-generating perspective.

### A Microgenetic Analysis of a Single Case from a Second Biology Study

When the text was used as a context for interpreting self-explanations, it was fairly straightforward for Chi et al. (1994) to claim that self-explanations were inferences. To make such a claim, one only had to know what information the text already provides and thereby conclude what information is omitted from the text, such as the function of a component. To then claim that such missing knowledge is inferred, one simply had to verify in the coding that such knowledge is present in the self-explanations. Thus, the coding scheme was fairly straightforward. Notice that our coding scheme could only support or reject what we expected to find: that is, our coding scheme tested our hypothesis of an inference-generating view in the same way that conventional experimental methods test hypotheses. (See Chi, 1997b for a discussion of this kind of coding method.) That is, our original incomplete text view dictated our coding scheme: We coded each proposition within a self-explanation (in the biology study), or each idea unit (in the physics study), independently of one another but in the context of the information provided in the text sentences. What we had not done was to code each self-explanation in the context of the student's evolving mental model. Only then can we differentiate and discern whether self-explaining is a mechanism of generating inferences or making repairs. Below, I attempt such an analysis in a single case for the purpose of illustrating how revision can be conceived of differently from inferencing.

In order to re-analyze SEs in the context of a self-repair (versus an inference-generation) view, we need a new set of SE data because the SE protocols of the previous biology study's were contaminated in the following way. In that study (Chi et al., 1994), after we had captured the students' initial mental models, the students were further required to answer a set of pre-test questions. These questions often contained premises that conveyed new information, allowing the students to change their initial models in some way. This means that we were no longer completely certain what initial mental models they had coming into the learning phase. Thus, in order to make sure that we had a stable initial mental model prior to the reading and learning phase of the study, a second set of SE data were collected. This second biology study was improved in a number of ways for the purpose of capturing students' mental models more accurately. First, we deleted the question-answering part of the pre-test and kept only the definition of terms and drawing of the blood path tasks. Second, because we wanted to collect self-explanation data that would allow us to capture the mental models more accurately, it was necessary to probe more deeply in order to understand what the students were explaining; whereas the original biology study forbade the experimenter to deliver additional probes because the original interest focused on the effect of prompting for self-explanations. Consequently, in this study, if a student self-explained something that was not clear to the experimenter, the experimenter would then probe the student for further clarification. Under no circumstances, however, was the experimenter supposed to lead the student in any way, or provide additional information, or scaffold the student. This study involved six eighth-grade students studying the same circulatory passage.

Flawed mental models. If self-explaining is the process of self-repair, and if self-repair occurs primarily when a mental model conflicts with a text model, then we need to first establish that some naive models are flawed and thereby are distinctly different from the correct scientific model. The single loop model (the most common naive mental model, discussed in Appendix A) appears to be flawed because it generates different predictions from the scientifically correct double loop model. Figure 5 depicts an idealized portrayal of a single loop with no lungs model. In such a model, blood carrying oxygen flows to all parts of the body, either simultaneously or sequentially from one part of the body to another part, then returns to the heart. The implicit assumption is that the heart oxygenates the blood. Lungs play no obvious role in circulation, even though all students know that oxygen, as inhaled in air, enters the lungs. We have simply delineated all parts of the body as "the body" in the figure. Variations among students can occur in the details of the features, such as whether they know that the vessel leaving the heart is called the artery, whether they know that blood in the artery is oxygenated (depicted by the ++ signs), and so on. A single loop with no lungs also means that the student can talk about the lungs in the protocols, but the lungs do not serve a vital aspect of the circulation.

Insert Figure 5 about here

Three pieces of information about the double loop model (as shown in Figure 6) violate the single loop model (Figure 5).

Insert Figure 6 about here

These three pieces of conflicting information are that: (1) the lungs are a component of the circulatory system, (2) lungs are the site of oxygenation, and (3) blood going to the lungs and returning to the heart constitutes a separate second loop in the circulatory system. The first piece of conflicting information, that lungs are an important component of the circulatory system, was first introduced in S18 of the text, and then again indirectly in S32 and S33. The second piece of conflicting information, that lungs are the site of oxygenation, was first introduced in S33. The last piece of conflicting information, that the heart-lung path and the lung-heart path constitute the second loop in the circulatory system, is partially presented in S34. In order to see whether self-explaining can be interpreted as a process of repair, the next section analyzes the characteristics of the SEs generated at these four loci of conflict (S18, S32, S33, and S34), in the context of a student's evolving mental model.

Four self-explanations at points of conflicts. In the abbreviated passage that was used to describe the circulatory system, only six sentences have been presented prior to S18. These six sentences primarily introduced the circulatory system, the heart, and its chambers. These six sentences, as well as the sentences intervening between S18 and S32, are shown in Table 2.

Insert Table 2 about here

It should be pointed out at this point that the student's protocols to be analyzed here were chosen randomly among the six students in this study. I did not analyze each of the six students' protocols to pick the best one. Instead, I picked a student who has an easily identified and robust naive model of a single loop with no lungs because that was the most common initial naive model among the 24 students of the 1994 study, for whom we had captured their initial mental models (see Appendix A, Figure A3). It may be fortuitous, but the first student whose self-explanations I analyzed illustrated the concept of revision quite clearly.

In order to walk the reader through the processes of repair in the protocols, Figure 7, the top row, depicts iconically what information is provided by the text sentences. The heading for each sentence also gives a sense of the critical information conveyed by that sentence. Thus, S18 conveys the idea that the lungs are a component of the circulatory system; S32 discusses the contraction of the right ventricle; S33 clarifies that the site of oxygen exchange is in the lungs; and S34 specifies the destination of the returning blood as being the left atrium.

Insert Figure 7 about here

The bottom row of Figure 7 depicts what the student articulated in the self-explanations after reading each of the four sentences. Similarly, the heading of each SE summarizes the content of the SE. In SE18, the student misinterprets the information and conceives of the lungs merely as another destination; in SE32, this misconception is resolved and the student anticipates the lungs as the site of oxygen/carbon dioxide exchange; S33 reinforces the notion of the heart-lung link, and S34 summarizes the second loop. Even though I assume that the revision of a mental model is a cumulative process (that is, that the entire mental model is retained in memory), the depiction does not carry forward what was constructed at each step of the way in order to contrast more sharply what information was conveyed in the text sentence and what knowledge is explained in the SE. To be conservative, I assume that what the student articulated is what is activated in memory. The first and last column depict the initial and final mental models, based on the student’s drawings and interpretation of its accompanying protocols in the pre-test and post-test. The entire protocols of the four sentences and the self-explanations are shown in Table 3.

Insert Table 3 about here

Now, we can read the SEs and see to what extent they characterize processes of repair. Note that the following analysis attempts to interpret the entire SE uttered after reading a sentence, in contrast to previous analyses: In the physics study, only the portion of each SE that contained physics-relevant comments were identified and coded; in the biology study, only SE inferences were coded.

The student's initial mental model captured from the pre-test data of definition of terms is shown as a single loop, in the first column of Figure 7. The depiction is idealized in that she said blood went to the head, the toes, the hands, all of which we have consolidated in the figure as "the body." At pre-test, this student (AFM) did not know that the lungs were involved in the circulatory system (such as that oxygen is in the blood, and that blood distributes oxygen), although she did know that lungs contain oxygen. (This is a common piece of knowledge among children of this age because they all know that the air people breathe in goes to the lungs, and we breathe in oxygen. Notice that the students mentioned lungs only because the instruction required them to discuss the role of lungs.) I have therefore depicted the lungs independently without them being linked to the circulatory system. This interpretation of her initial mental model on the basis of the definition of terms data is also consistent with her drawing of the circulatory system requested at the pre-test. So, her model needs all three "repairs," corresponding to the three potential violations that have been identified: She needs to add lungs to the circulatory path, she needs to attribute lungs with the function of injecting oxygen and extracting carbon dioxide, and she needs to add the pattern of blood flow to, and returning from, the lungs.

Her first chance to incorporate the lungs in the blood path is at S18, in which it mentions that the right side of the heart pumps blood to the lungs in the context of the location to which the left side pumps blood. Thus, the meaning of the purpose of lungs in this sentence is ambiguous, as shown:

S18: The right side pumps blood to the lungs,

> <u>and the left side pumps blood to other parts of the body.</u>

This piece of information will be referred to as the Heart-to-Lung link (depicted in the lower right-hand corner of Figure 6 as the H-->L link); it is basically the first half of the "other loop," where the blood path is directed toward the heart. What she said at this point was:

> SE#18: "Just that um, the right side is primarily for the lungs
> and the left side is to the rest of the body."

At first glance, the fact that she confirmatorily rephrases that the right side is primarily for the lungs, seemed to have suggested that this piece of new information was easily assimilated, even though it obviously violated her single loop model. However, being able to read the subsequent self-explanations suggests that she has misinterpreted this sentence to mean that lungs are just another destination to which blood has to be delivered, and for some reason blood from the right side of the heart is reserved primarily for that specific destination—the lungs. Thus, I offer the interpretation that at this point, she is thinking of the lungs not just yet as a component of the circulatory system, but merely as another destination to which blood has to travel. This interpretation is consistent with what she said during the pre-test, when she described blood as going to the head, the toes, the fingers, and so on. Thus, naturally, lungs would simply be another destination for blood to go. (This interpretation is the same as the one given for SE#9, where the student talked about "giving" blood to the lungs). This interpretation, if correct, means that she had no conflict at this point because she viewed the text information as already consistent with her view but with a slightly different emphasis. Thus, her explanation at this point is short and basically agreed with the sentence (characteristic four of Table 1) and emphasized the fact that the right side is *primarily* for the lungs. It will become apparent that this interpretation is correct. This interpretation of her self-explanation is depicted in Column 2 of Figure 7, bottom row, whereby the lungs are embedded in "the body," as if lungs are just a component of the body.

The next occasion at which she encounters the possibility that lungs play a role in the circulatory system is S32, in which the information in S32 reinforces the notion that blood circulates to the lungs, except it is presented as background information, as:

> S32: The muscles of the right ventricle contract and force blood
> through the right semilunar valve and into vessels leading to the lungs.

Her self-explanation at this point is long, uncertain, effortful, and contains six episodes. First, she monitors her confusion (even though this sentence should not be any more confusing than other sentences):

> (1) "(pause) Um, the sentence is a little confusing. It's like,
> it's almost wordy, sort of, I guess. Or something.
> Reading it once, I don't understand."

After this confusion, the experimenter prompted the student with (experimenter's prompt is contained in square brackets and occurs only when the student seems confused and inarticulate):

> [Can you try to explain, you know, put it in better words? Or words you might use?]

The second episode of her SE32 focuses on acknowledging (by paraphrasing) the new, foregrounded information in the sentence about muscles contracting (up to this point, the heart has only been referred to as pumping, never as a muscle contracting), so she said:

> (2) "(pause) I mean, I guess I understand now. I just, I can't think.
> I don't know, but kind of a muscle contraction that pushed the blood, um,
> through the valve and into vessels, but I don't know."

After this episode of uncertainty, the experimenter again prompted with:

> [Does the sentence make sense with your understanding of where blood is flowing in the circulatory system?]
>
> "(Pause) Mmm, yeah."
>
> [So what is happening here, basically?]

After which the student summarized with:

> (3) "Um, blood is just flowing from the ventricle into vessels and *going, um, to the lungs*, um, ok."

Notice that her summary is about the destination of blood flow, which is not the gist of S32 at all. Her summary excludes the details and inferences about contraction, the force that pushes blood, and the route of flow through the right semilunar valve. Instead, she mentioned the background information of going to the lungs. I interpret this to mean that this aspect of her mental model was incorrect and had to be repaired.

The fourth episode is her next unprompted SE, which in effect resulted from her noticing that this sentence reminded her of S18:

> (4) "I guess, then I was remembering the thing about the left side is for the rest of the body, and the *right side is mostly for the lungs*, but well I guess I don't know."

This fourth episode of SE32 confirms my original interpretation that she had misinterpreted Sentence 18, so she is now trying to resolve her misinterpretation of what S18 said about the lungs, and whether she ought to have treated the lungs just as another destination or body parts. The experimenter prompted again:

> [Well what, what are you thinking there?]

The fifth episode is her revision of her previous incorrect view that the lungs was just another destination, so here she says:

> (5) “Well either, at first, I was just thinking that, you know,
> it was just *providing the lungs with blood*
> which didn’t make much sense.”

So my interpretation of her SE back at S18 is now supported by her remark above. With this new insight (that it didn’t make sense that blood flowing to the lungs was just going there to supply lungs with blood, but rather, it’s going there for another purpose), she predicts in the next episode that the lungs might be the site of oxygenation, but waivers, continues to acknowledge conflict, and finally anticipates that going to the lungs may serve a different purpose:

> (6) “but that possibly, I don’t know, but maybe
> it’s *going there [the lungs] to receive oxygen* or something.
> I don’t know.”

Basically, what she has learned here, as opposed to S18, is that blood flows to the lungs for another reason besides delivering blood (and oxygen). It goes there to get oxygenated. So at this point, her mental model has incorporated two of the key pieces of conflicting information: that lungs are a component and they are the site of oxygenation. The representation of her knowledge learned at the SE of S32 is shown in Figure 7, Column 3, bottom row.

There are numerous characteristics to notice about this long SE at S32. First, she was extremely uncertain about what was going on, claiming that the sentence was confusing and didn’t make sense, when in fact, the sentence itself was no more confusing than any other sentence. One evidence of her uncertainty was the number of times she said “I guess,” “I don’t know,” “maybe,” “not make much sense.” There were a total of 14 of these uncertainty comments, and at least 6 “ums.” Smith and Clark (1993) show that “um” should be interpreted as a longer delay before answering a question, indicating that the subject is aware that he or she does not know the correct answer (but is still hedging for time). The only interpretation possible for her uncertainty is that it arose from a conflict between the information conveyed in the sentence, and her erroneous mental model of attributing the lungs merely as a blood destination rather than as a site for the purpose of oxygenation. Hence, in this SE of S32, it is clear that a conflict has been detected. Thus, this entire SE exhibits uncertainty (characteristic one of Table 1) when a violation is detected. If this sentence is confusing, as she claimed, the confusion arose from this violation, and not from the information about muscle contraction and the right semilunar valve.

Second, even though the foregrounded information in the sentence dealt with the ideas of contraction of the muscles and blood flowing through the semilunar valve, such knowledge was not as problematic for her because these details provided no violation of her existing mental model, so that they can be readily incorporated. This means that no prolonged self-repair process is needed, even though there is a knowledge gap. So she acknowledged this new knowledge briefly (in episode two), but focused the remaining episodes on the backgrounded information that blood went to the lungs. Thus, here is an example that illustrates how gaps in a mental model can be readily filled without much fanfare, if no violations exist.

Third, in four of the six episodes in SE32, she puzzled over the possibility that blood flows to the lungs. Moreover, she continued to mention this conflicting piece of knowledge (that blood goes to the lungs) four times (in episodes three, four, five, and six; these have been italicized). The four repetitions of this conflicting but backgrounded information presented in the sentence do not make sense in the context of the view that self-explaining is a process of generating inferences to fill gaps of missing information, because the knowledge that was repeated was already explicitly stated in the text: It is difficult to reconcile why one would need to repeat a text statement four times. Yet, the repetition makes complete sense in the context of repairing one's mental model (characteristic three of Table 1), which did not initially incorporate this piece of knowledge in the right way. Hence, repetition makes more sense as a process of making the repair and/or reinforcing one's repair than a process of inferencing.

In sum, the features to emphasize about SE32 at this point are that the student finally noticed a violation between her mental model and the text information. This detection created a conflict for which she tried to resolve, resulting in the long and tortuous explanation. Her SE here was exceedingly long (characteristic four of Table 1), not because the experimenter probed her for additional clarification (she was probed additionally precisely because she exhibited confusion, as compared with her SEs after reading S18, S33, and S34), but because she had to resolve her conflict. This SE is particularly telling for contrasting it with an inference, in that, although the sentence was really about contraction and semilunar valve, her explanation focused on the backgrounded information about going to the lungs. (An inference-generating perspective would have predicted that she elaborated on the new information given in the text, and in this case, it would be the foregrounded information.) In fact, after her initial claim of not understanding, and the experimenter probed her, she did say explicitly that "I guess I understand now" (in episode two), obviously because she was referring to understanding the new information about muscle contraction and the valve (which was referred to in episode two). However, the differential way that various pieces of information in this one sentence is handled confirms the interpretation that a violation was detected. Finally, although such agony makes complete sense in this context, one would not have predicted, on the basis of a text analysis, that this sentence would elicit such a long and uncertain self-explanation. That is, this sentence is not technically harder to understand than the other three sentences, nor could one claim that this sentence omitted any particularly relevant  information than the other three sentences, so that there is no *a priori* reason to expect S32 to evoke such a reaction.

The fact that lungs are the site of oxygenation was not actually described until S33, which says

S33: In the lungs, carbon dioxide leaves the circulating blood and oxygen enters it.

Because this knowledge had already been anticipated by the student in her previous SE (episode six of SE32), her current SE merely confirmed this piece of information, which no longer provides a conflict:

SE#33: "OK, well then, so there, I guess so the blood *goes to the lungs* to um, get rid of carbon dioxide and have more oxygen put in it."

Thus, when the information conveyed in a sentence not only does not violate one's mental model, but confirms it, the student expresses this with a simple confirmatory statement, such as "so there." The student also took another opportunity to reinforce the notion that blood goes to the lungs.

S34 provides explicit information about the other half of the second loop, namely, the lungs-to-heart link and the location of the returning blood:

S34: The oxygenated blood returns to the left atrium of the heart.

At this point, the student's SE says:

SE#34: "So, the blood leaves the heart, goes to the lungs, and comes back."

Again, notice the characteristic of this very important SE. First, the information provided in S34 that was new to her had to do with the completion of the second loop. However, because encoding this half of the second loop (the lungs-heart link) does not conflict with her model at this point, the assimilation could be done readily, so her self-explanation incorporated this lungs-heart link. Her SE at this point basically summarized the complete second loop, that blood goes to the lungs from the heart and returns to the heart, as if she is seeing this loop by her mind's eye in her mental model. It seemed independent of the detailed content of the information conveyed in this S34: She ignored the information about blood returning to the left atrium, nor did she differentiate whether it was oxygenated or deoxygenated blood. Instead, she articulated the most important aspect of what she has learned: the presence of a second loop.

Figure 8 summarize this microgenetic analysis by showing the frequency of utterances of repair characteristics one, two and three, plus self-explanations for the four sentences at which we have predicted that a conflict might be perceived. In reality, violation was recognized only at S32 , so that there was an abundant evidence of repair, such as pauses, repetitions, and monitoring statements of failure to understand. What is revealing is that these characteristics of repair statements give rise to a pattern of differential long and short self-explanations, as were the cases shown in Figure 4. Presumably, when violations are detected, long and arduous explanations would be manifested, whereas when new information is consistent with one's mental model, then it can be either assimilated (if gaps exist), agreed with (it it matches), or ignored, giving rise to a pattern of differential long and relatively short and confirmatory self-explanations.

Insert Figure 8 about here

Although only four characteristics of repair were specified *a priori*, the analysis itself revealed additional characteristics that could also be interpreted as consistent with a repair (and/or the successful detection of violations) interpretation. The most salient one is the focus on the backgrounded information rather than the foregrounded information.

It should be noted that this repair process, at the microscopic level, includes adding links and features that were not there originally (such as the relationship between the heart and the lungs, that lungs are the site of oxygenation); integrating links such as the case of integrating the heart-to-lungs link with the lungs-to-heart link. Repair can also include deletion or removal of links and relationships that we have seen in the evolution of other students' mental models. Thus, at a microscopic level, the repair processes per se are elementary operators such as addition, deletion, concatenation, or generalizing of features, operators that are well understood by cognitive scientists. At a macroscopic level, however, repair can be described as processes that are triggered by perceived violations and have the characteristics shown in Table 1. Thus, our original coding scheme of identifying self-explanations as individual pieces of knowledge inferences did not permit us to conclude that self-explaining included the process of mental model revision. It is only when we re-interpreted the self-explanations in the context of the student's evolving mental model, could we then see that these self-explanations are not necessarily inferences that serve the purpose of filling gaps created by an incomplete text, but rather that some SEs, especially those at points of conflicts, are really manipulations that serve the goal of repairing an evolving understanding. This interpretation also suggests that the reference of self-explaining is the mental model, and not the external text.

### Evidence Consistent with the Repair Interpretation

The interpretation that self-explaining is in part a process of repairing one's own mental model, rather than strictly a process of inferring missing information, was based on the analysis of a single subject's self-explanations in the context of her evolving mental model. This level of detailed analysis was necessary in order to portray the differences between a repair versus an inferring process. Analyses at this level of detail unfortunately also preclude the analysis of multiple subjects. However, the conclusion about the mechanism of self-explaining as a repair process is consistent with several additional results that are either findings in the literature or the result of more systematic analyses. In this section, three sets of findings are cited that have been difficult to explain strictly in the context of an inference-generation interpretation, but make sense in the context of a repair interpretation.

First, if self-explaining is the process of inducing missing information, then presumably if the researcher added this missing information, a text would be more comprehensible, and one could learn from such a text better. For instance, students should be able to learn better from reading an elaborated versus an unelaborated text, because the elaborations are generated on the basis of what the researchers think is optimal information that should be supplied. However, the results of this kind of study are not always consistent: Sometimes elaborations inserted into the text help and other times they do not (Recker & Pirolli,1995; see discussion in Reder, Charney, & Morgan, 1986); and whether they help or not depends on the readers' background knowledge (McNamara et al., 1996). Therefore, it seems nearly impossible for an external agent (such as the researcher) to improve a text passage (by adding elaborations and inferences) so that a passage is uniformly comprehensible for all readers. Again, the lack of consistency in the benefit of inserting the missing information also suggests that one cannot think of self-explaining strictly as a mechanism of inferring knowledge to fill omissions in the text. The inconsistency can be understood in the context of a repair point of view. That is, the elaborations provided by a researcher, on the basis of a normative text analysis, may not always be the repairs that any given student needs in order to revise his/her own model.

Second, Webb (1989) has found that generating explanations is generally more facilitative to learning than receiving explanations from others. Although this result is consistent with the constructive view in general, the result makes even more sense in the context of repair, in that the students can repair their own mental models more effectively than have an external agent repair their models for them. That is, more generally, didactic instruction may be less effective than self-explanations precisely because such instruction is not tailored to repairing the individual student's mental model (Chi, 1996), largely because teachers and tutors cannot accurately diagnose what mental model students have (Chi,1997c). Thus, explanations are most useful when they are generated by oneself, because they serve the purpose of repairs.

Third, our results (Chi et al., 1989) on monitoring accuracy are consistent with the repair view. Conceiving of self-explaining as the process of self-repairing assumes that self-repairing is more likely to be undertaken if a conflict is detected by the learner. Monitoring statements sometimes reveal detection of conflicts. In the physics study, 39% of the protocols were coded as monitoring statements. These statements reflect self-assessment of understanding, such as "I can see now how they did it," or failure to understand, such as "Why is mg sin theta negative?" The results showed that the poor learners exhibited awareness of comprehension failure only 15% of the time, whereas the good learners acknowledged comprehension failures 46% of the time. This means that the poor learners thought they understood most of the time when in fact they did not, implying that they did not detect any conflicts between their understanding and what the text says. The relevant finding for the present discussion is that, regardless of how frequently a student detects comprehension failures, once they did detect them, it was generally (73% of the times) followed by episodes of self-explaining (Chi et al., 1989). Thus, when students realized that they did not understand, they must have perceived a conflict between their own mental model and the text information, and such conflicts elicited self-explanation, which is the process of trying to resolve the conflict. At the time, we offered no explanation for this pattern, that is, why detection of comprehension failures would lead to self-explaining, but now, with the self-repair view, one could say that self-repairing requires the successful detection of comprehension failure, or some conflict between what the student thinks is going on versus what the text is presenting.

If we assume monitoring to be the process of comparing one's mental model with the text information, then the preceding interpretation suggests that the source of individual differences in self-explaining might be the frequency with which students monitor their understanding. In the physics data, if we compute the frequency of monitoring per sentence (rather the amount of monitoring statements overall, as reported in the preceding paragraph), it is apparent that the high explainers did monitor their comprehension more frequently (across 37% of the sentences), than the low explainers (23% of the sentences). However, the frequency of monitoring does not necessarily imply that conflicts will be detected, although there is no question that the more often students monitor their comprehension, the more likely they are to notice conflicts (the correlation between monitoring of understanding statements and monitoring of misunderstanding is .79). This is because conflict detection may depend on a number of other factors, such as the status of one's mental model and correct interpretation or parsing of the text sentences.

In sum, although the proposal that self-explaining is at times a process of revision was based on analyzing the characteristics of a single student's self-explanations, these characteristics seem to fit a view of repairing an imperfect mental model, better than a view of inferring knowledge to fill gaps and omissions from an imperfect text. This view was proposed as a result of skepticism arising from some unexpected findings (such as no consistency in when and what self-explanations are articulated, and the existence of individual differences in the number of self-explanations generated). Moreover, the repair view receives additional support when we project certain characteristics requisite of self-repairing and not inferencing, such as the repetition of information that is already explicitly stated in the sentences (our previous coding of self-explanations as inferences did not count these repetitions, because these repetitions did not contain new information, thus were not coded as SEIs). These characteristics give rise to a distribution of differential (long and short) self-explanations. Additionally, in the process of analysis, other detailed differences also surfaced to support a repair view, such as focusing on the backgrounded rather than the foregrounded information. Moreover, the repair view seems to be consistent with evidence in the literature (such as no uniform benefit is derived from inserting inferences into a text, or that self-explaining is more effective than receiving explanations), as well as the analyses of our own monitoring results. Thus, conceiving of self-explaining as a mechanism of repairing one's mental model, in addition to a mechanism of inferring information that is missing from a text, seems more accurate and complete. Finally, the self-repair view allows us to make sense of and tie in the monitoring results and postulate a potential explanation for individual differences in the amount of SEs students generate. That is, if we conceive of monitoring as a process of comparison between the mental model and the external text, then it makes sense that a greater frequency of monitoring would lead to more opportunities to detect conflicts, which in turn would lead to self-repairs.

## Summary and Discussion

This final section begins with a summary of the mechanism that enables self-explaining to promote learning and shows how it can account for the most critical piece of finding, namely that there are individual differences in the number of SEs students produce, which then determine how well they learn. Several other related issues are also addressed.

## **The Process of Self-repair**

Originally, self-explaining had been portrayed as primarily the mechanism of inferencing, both inferencing from prior knowledge (either commonsense knowledge or domain-relevant knowledge), inferencing from integrating two or more pieces of information from the text, or inferencing by relating the new information in the text with prior knowledge. Moreover, the inferences supplied were seen as serving the purpose of inducing missing information that was omitted in the text sentences. However, the inference-generation view fails to explain individual differences in the amount of SEs generated, corresponding to the amount of learning gains observed. The presence of this individual differences finding, not accountable by ability or prior knowledge differences, poses a problem for understanding the kind of intervention that might be most effective to optimize learning gains. An additional mechanism for self-explaining is then proposed. Extended self-explaining is now conceived of as a process of repairing one’s representation, usually when a conflict is detected. Thus, individual diffeences arise from differences in the mental models students bring to the learning situation and/or the ones they construct. Moreover, even if two students have the same mental model at a global level, their mental models may have different gaps, and they may differentially notice whether or not text sentences violate their mental model. The proposal of a repair mechanism was first instantiated with a detailed analysis of a single student’s self-explanations after reading four critical sentences (critical to changing the initially incorrect mental model the student had). Then, this interpretation was shown to be consistent with some of the findings in the literature at large, as well as some of our own findings that were heretofore unexplainable (such as the lack of consistency in terms of when and what SEs are generated). Most importantly, this repair view allows us to tie in the monitoring results and postulate an additional source of individual differences as differences in the frequency with which students monitor their comprehension.

What do we gain by viewing self-explaining as a process of repair in addition to a more straightforward process of inferring missing information? First, such a view explains the differential amounts of self-explaining students undertake, even when prompted. That is, it was always mysterious why some students explained and learned more (the high explainers) and others explained less, even though both groups were prompted to explain, and had similar ability and background knowledge.  The explanation I had offered before, that the high explainers learned more because they explained more frequently, or that they provided deeper explanations, was circular. The current view is that spontaneous and lengthy self-explaining occurs when conflicts are detected.  However, whether or not conflicts are noticed in the first place depends on a number of  other independent factors, such as the status of student’s existing mental model, and how a given sentence is interpreted in the context of it.  For example, in the example presented earlier, S18 was misinterpreted by the student, thinking that lungs were just another body destination to which blood has to travel, therefore, the student did not perceive a conflict.   Thus, various factors (other than ability and prior domain-relevant knowledge) can account for the differential accuracy with which students detect a conflict, thereby self-explain and learn. Because detection of conflict generally encourages the process of resolution, prompted high explainers (in the biology study) may simply be those students who have accurately detected conflicts more frequently, and spontaneous high explainers (in the physics study) may simply be those students who monitored their comprehension more frequently and suceeded in detection.

Second, viewing self-explaining as a process of self-repair, requiring an accurate detection of conflicts, also suggests that prompting may achieve two additional benefits, besides encouraging students to generate omissions.  It may be a way of encouraging students to compare their own mental model with the incoming information (akin to the process of reflection, Collins, Brown, & Newman, 1989), thereby giving rise to more opportunities to notice conflicts.  Prompting may also encourage students to propagate a contradiction (to see its consequences), causing a realizaation that the contradiction is really a violation.

Finally, this mental model perspective makes complete sense of the monitoring results.

## **What Happens with Incorrect Self-explanations?**

Self-explaining has been a non-intuitive learning activity because it was always assumed that a novice cannot learn without the guidance of an expert.  That is, if a novice learner does not have the appropriate domain knowledge in the first place, what happens if his/her self-explanations are incorrect?  It turns out that across our studies, about 25% of self-explanations are incorrect. However, the basic pattern of results remains the same whether or not these incorrect ones are included or excluded from the data analyses. Why are they harmless or do they in fact facilitate learning?  One could not construct a sensible explanation of why incorrect SEs are harmless in the context of an inference-generating view only, but it makes complete sense in the context of a self-repair view.  The harmlessness of incorrect SEs makes sense if we do not conceive of them always as inferences of omitted information.  Such a conception should predict that they are harmful, analogous to the way supplying imprecise elaborations are harmful, as shown in the elaboration research (Stein & Bransford, 1979).

However, if self-explaining is conceived of at times as a process of repair, then one could even entertain the idea that generating incorrect self-explanations might actually promote rather than depress learning.  That is,   generating an incorrect piece of knowledge will more likely than not ultimately create a conflict, because the text is likely to refute it, either directly or indirectly (in our data, about 26 out of 31 times, or 84% of the time, incorrect SEs are refuted by subsequent text sentences).  Given that the detection of misunderstanding (as indicated by monitoring of comprehension failures) inevitably leads to self-explaining episodes (73% of the times, Chi et al., 1989), this would suggest that generating an incorrect piece of knowledge simply creates an opportunity for conflicts and misundertanding to occur because the incorrect self-explanation will be challenged by the correct information from the text. Given that the detection of a misunderstanding then leads to self-explaining episodes of trying to resolve it (Chi et al., 1989), this suggets that generating incorrect self-explanations can actually promote learning.

If this hypothesis is correct, then one would predict that students would be more likely to learn a piece of knowledge if it conflicted with prior knowledge.  In VanLehn’s (in press) reanalysis of the (Chi et al., 1989) physics data, he found eleven target pieces of knowledge that students had to learn that were missing from the textbook.  For each target, VanLehn judged whether the students had a belief that conflicted with the target.  For five of the eleven targets, there were clear conflicting prior beliefs, and these five were significantly more likely to be learned than the six target rules without prior conflicting beliefs, suggesting that the presence of conflicts (with the possibility that they would be detected), does promote learning.

## Comparison to other Constructive Activities

Is self-explaining unique or different from other types of learning activities?  In this section, six other constructive activities that students (instead of teachers) can engage in while learning from a source such as a textbook (by reading), a lecture (by listening), or a model (by observing), will be considered.  These six activities are 1) self-questioning (or posing questions to oneself); 2) explaining or posing questions to others; 3) asking questions (for information that one does not know the answers to); 4) answering questions posed by others; 5) summarizing and note-taking; and 6) drawing. All six learning activities have been explored in the literature and shown to be effective at promoting learning to varying degrees.

The first activity, self-questioning or posing questions to oneself that one does not know the answers to, also showed superior comprehension improvement when undertaken while listening to class lectures (King, 1991). In self-questioning, students are taught how to generate questions, usually using question stems such as “How is ...related to ...?” or “Explain why...” or “What do you think would happen if...” Hence, one conjecture of why students learn from self-questioning is that in the process of trying to answer questions they have generated using such question stems, they may experience feedback about what knowledge they fail to understand, and such feedback can alert them to the need of resolving their misunderstanding.  This interpretation suggests that self-questioning is effective because it leads to some resolution of misunderstanding while the student is trying to answer the questions.  Notice that self-questioning is very different from the kind of monitoring questions that students utter in the process of self-explaining, such as  “Why is the knot the body?” (in reference to the knot versus the block in Figure 1) or “Why does blood go to the lungs?” (in reference to the single loop model in Figure 5).  These latter types of questions are much more content-driven and do not fit neatly into question stems: they are really a form of query that points out conflicts between their mental models and the text’s content model.  However, one could say that self-questioning is effective because it is an initiating event that ultimately leads to the detection of misunderstanding in trying to answer it.

Explaining to others and posing questions to others seem different from self-explaining and self-questioning in that both self-explaining and self-questioning may be conceived of as an activity that manipulates and questions one's mental model, whereas explanations and questions posed to others may be based on information contained in the content of the text. Because explaining and posing questions to others is a constructive activity, it should promote learning gains to some degree. In fact, explaining to others is superior to receiving explanations, as Webb (1989) had shown. It seems logical that the kind of questions students pose to others is derived from information in the text, and not necessarily misunderstanding in their own mental models, although the latter possibility cannot be ruled out. There are no studies, to my knowledge, that contrast self-questioning or self-explaining with questioning and explaining to others. Explaining to others obviously can also result in mental model revisions.

The third activity, asking questions of others, presumably for information that one is lacking, can be a very effective learning strategy. Not only is it a constructive activity, but presumably one asks questions about conflicts and misunderstanding in one's own representation. The opportunity to ask questions can be a reasonable explanation for the improved learning in a tutoring context as compared to a classroom context. According to Graesser's (1993) estimate, the opportunity students have to ask questions in the classroom setting is 0.11 questions per hour (including both deep and shallow questions), whereas the opportunity to ask questions goes up significantly in a one-to-one tutoring context, to 8 deep questions per hour. To the extent that the explanations provided to the students' questions are salient and appropriate, then seeking these answers would help the students resolve conflicts and misunderstanding in their representations.

The fourth activity, answering questions, such as questions posed by a teacher or by a peer, is a constructive learning activity. To that extent, one would expect some learning gains. However, learning gains from answering questions are typically small (Hamilton, 1985; Redfield & Rousseau, 1981), unless the students are encouraged to generate more explanatory answers (King, 1990, Exp. 2). The small gains in answering questions may result from the fact that answering questions obviously serves the purpose of others' goal, rather than the goal of revising one's own misunderstanding. Answering questions is also less effective than asking questions (Davey & McBride, 1986).

The fifth activity is summarizing. A summary by definition, is like an abridged version of a text passage, which can be produced only after the students have read a longish passage. Summarizing seems to serve the purpose of the text rather than oneself, because a summary, by definition, discourages the students from integrating the information in the external source, such as the text, with one's own knowledge. A summary that contains prior knowledge is usually considered to be inaccurate because it contains "intrusions." Not surprisingly then, summarizing is less facilitating to retention of the lecture content after a one-week delay than self-questioning (King, 1992).

Although note-taking can be a distinctly different activity from summarizing, it is similar to summarizing in that students often take notes by selecting key sentences from the text (as if they merely delete certain sentences from a text, much like the way students write summaries using the copy-and-delete strategy, Brown & Day, 1983). Again, being a constructive activity, one would predict that note-taking would be an effective learning activity. It does facilitate retention of text information after a one-week delay, for instance, more so than re-reading the text (Dyer, Riley, & Yekovich, 1979). Note-taking is less effective, however, for retention of lecture content than self-questioning (King, 1992), which would be consistent with the self-repair prediction because self-questioning, like self-explaining, serves the learner's own goal.

The last type of activity, drawing, of either concept maps (Novak, 1993) or diagrams, are constructive activities that occur in a spatial medium. Drawing also seems to serve one's own purpose, so that the self-repair view would predict that it is an effective learning activity. However, drawing is limited in one peculiar way but is superior in another way. The limitation is that one cannot make drawings to be as detailed and rich as one could in talking (such as self-questioning or self-explaining). Nor can drawing be done in as rapid a way as talking in self-explaining or self-questioning. The advantage of drawing is that one can look at one's drawing and make further inferences from it. For instance, drawing diagrams while solving geometry problems is particularly useful because further insights can be gained about relations depicted in the diagrams. Similarly, in a concept map, certain relations among nodes may be more transparent after a concept map is drawn. Because of the limitation of a spatial rather than a verbal medium, I would predict that it does facilitate learning (as in the case of concept mapping, Novak, 1990), but the effect is sometimes small (Heinze-Fry & Novak, 1990).

The purpose of the foregoing exercise is to compare the relative effectiveness of each of these learning activities, in contrast to self-explaining. Although all these activities are constructive ones and promote learning to varying degrees, it is obvious that they embody different processes. Gross predictions were made about their relative effectiveness in terms of whether an activity involves self-repair or not. Clearly, such comparison statements can only paint broad strokes, because the effectiveness of any given activity depends also on the quality of the product. For example, answering questions becomes a more effective activity when the students are asked to provide more elaborated answers (Woloshyn, Willoughby, Wood, & Pressley, 1990). Likewise, summaries may be more effective at enhancing learning if they were produced by condensation rules such as invention and integration, rather than copy-delete rules (Brown & Day, 1983). However, it seems clear that the activities that are most facilitating to students are ones that serve the students' own goal of trying to understand in ways that allow them to question and repair their mental representations. In this sense, self-questioning might force the student to realize, in the process of trying to answer the questions, that they do not understand. Although self-questioning seems to differ from the kind of questions that usually occurs when the student experiences conflicts (that is, when there is a mismatch between the mental model and the content model), such as “Why is the knot the body?” or “Why is blood going to the lungs?” (perhaps these should be called “monitoring questions”), nevertheless posing questions to oneself can be a useful initiating event that can trigger self-explaining episodes.

## **What Type of Intervention is Effective?**

Assuming that students come into an instructional context (either in a classroom or in a one-on-one tutoring situation) with some preconceived notions of the content, they will thus have some existing mental representations that are incomplete and distinct from the scientific or correct conception. In such cases, a mental model repair view also dictates what types of intervention might be more appropriate. It suggests that it would not be productive to supply all the inferences missing from a text, and expect this type of manipulation to benefit all students, as confirmed by the studies of McNamara et al. (1996), because these inferences may not necessarily correspond to gaps in students' mental representations. Rather, such a view suggests that the students themselves are in the best position to detect what gaps and conflicts exist so they may fix them, consistent with the commonsense view that didactic instruction is less effective than constructive learning. In fact, several pieces of evidence now suggest that it is difficult for an external agent, such as a teacher or a tutor, to know precisely what the students' existing mental representations are, in that they cannot accurately diagnose what mental models students have and what conflicts they perceive (Putnam, 1987; Chi, 1996; 1997c; Staub, 1997 personal communication). Consequently, the students themselves are in the best position to know what and how to modify and revise their representation, in light of the correct knowledge presented to them.

Assuming this interpretation is true, a more promising intervention might be a kind of prompting that encourages students to detect inconsistencies and violations between their mental models and the normative model, rather than focusing on the most optimal ways of conveying the correct information. A way to give them opportunities to detect conflicts is to ask them to reflect, to self-question, or to self-explain (that is, prompting to self-explain may inadvertenly also prompt them to reflect). This is consistent with Collins et al.'s (1989) notion of reflection. Reflective learning is an activity of reflecting on one's choice of solution steps by observing one's own problem-solving steps/trace, and comparing them with the expert's trace. Thus, reflection is a comparison process, albeit discussed in the literature primarily in the context of a procedural task. Similarly, it might be equally useful to prompt students to reflect in a conceptual domain, whereby they compare their mental representations with the external information. Thus, a useful intervention might be some kind of prompt that can promote reflection.

## Two Caveats

Two very important caveats need to be emphasized at the conclusion of this paper. First, the self-repair view assumes that a student's naive mental model for a given domain and the scientific model are "compatible," in that the naive model can be repaired in a cumulative way so that it can evolve eventually into the correct scientific model. It is under such circumstances that conflicts can lead to self-repairing, as it occurs through self-explaining. There are other domains and concepts, for which the students' initial mental models are completely inappropriate, so that no amount of revisions can accumulate and lead ultimately to the correct scientific model, even if the students perceive conflicts. In such cases, instead of revision, the students should "abandon"[3] their initial mental models and initiate the construction of a new mental model. For example, if students initially conceived of natural selection as a causal event rather than the accumulation of multiple, independent, simultaneous, and uniform, subevents, then no amount of revision of their initial model will result in the correct scientific model (Chi, 1997a; Ferrari & Chi, in press; Chi, in press). In such cases, it may be necessary to provide a metaframework (of a self-organizing vs an event-like process) for the student to construct interpretation of new information, rather than rely on the recognition of conflicts, since conflicts in this case occurs at the level of the metaframework. We have preliminary data showing that providing such a metaframework facilitates understanding of domains for which the naive conceptions and the scientific conceptions are not compatible (Slotta & Chi, 1996).

A second important caveat is that the notion of conflict-driven learning leading to repair in a conceptual domain is reminiscent of the notion of failure-driven learning when expectations are violated (Schank, 1986) or impasse-driven learning leading to repair in a procedural domain (VanLehn, 1988). However, it is not yet clear whether conflicts, impasses and failures are isomorphic. For example, impasses in a procedural domain are reached when a solver can no longer find any rules to apply to reach the goal, whereas conflicts are detected when sentences violate a

[3] "Abandon" does not implicate forgetting or eliminating from one's memory. Abandon simply means that the student should not build upon the initially retrieved mental model, but instead, construct a new mental model, or build from another one.

mental model. The extent to which the mechanism of mental model revision corresponds to the mechanism of buggy-rule repair or gap-filling (VanLehn, in press) remains a deep issue that cannot be resolved in this paper.

In conclusion, self-explaining seems to be an effective domain-general learning activity. If psychologists and educators had heeded Ben Franklin's wise remarks, we would not have had to waste our time studying learning (as measured by remembering and forgetting) in the context of telling and teaching. We should have known that we needed to focus on involvement, a form of which is self-explaining, in order to achieve learning. However, perhaps Franklin could have gone one step further, and added, "Challenge me and I understand."

## References

Brown, A. L., & Day, J. D. (1983). Macrorules for summarizing texts: The development of expertise. *Journal of Verbal Learning and Verbal Behavior, 22*(1), 1-14.

Chan, C., Burtis, J., & Bereiter, C. (1997). Knowledge building as a mediator of conflict in conceptual change. *Cognition and Instruction*, 15(1), 1-40

Chi, M.T.H. (1978). Knowledge structures and memory development. In R. Siegler (Ed.), *Children's thinking: What develops?* (pp.73-96). Hillsdale, NJ: Erlbaum.

Chi, M.T.H. (1996). Constructing self-explanations and scaffolded explanations in tutoring. *Applied Cognitive Psychology,* 10, S33-S49.

Chi, M.T.H. (1997a). Creativity: Shifting across ontological categories flexibly. In T.B. Ward, S.M. mith, & J. Vaid (Eds.), *Conceptual Structures and Processes: Emergence, Discovery and Change* (pp. 209-234). Washington, DC: American Psychological Association.

Chi, M.T.H. (1997b). Quantifying qualitative analyses of verbal data: A practical guide. *Journal of the Learning Sciences,* 6(3), 271-315.

Chi, M. T. H (1997c, October). Self-construction and co-construction of explanations during tutoring. Final report, Spencer Foundation.

Chi, M.T.H. (In press). Understanding of complex, abstract, and dynamic concepts. In *Encyclopedia of Psychology,* APA and Oxford University Press.

Chi, M. T. H., & Bassok, M. (1989). Learning from examples via self-explanations. In L. B. Resnick (Ed.), *Knowing, learning, and instruction: Essays in honor of Robert Glaser* (pp. 251-282). Hillsdale, NJ: Erlbaum.

Chi, M. T. H., Bassok, M., Lewis, M., Reimann, P., & Glaser, R. (1989). Self-explanations: How tudents study and use examples in learning to solve problems. *Cognitive Science, 13,* 145-182.

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

Chi, M. T. H., Feltovich, P., & Glaser, R. (1981). Categorization and representation of physics problems by experts and novices. *Cognitive Science, 5,* 121-152.

Chi, M. T. H., & Koeske, R. (1983). Network representation of a child's dinosaur knowledge. *Developmental Psychology, 19,* 29-39.

Chi, M.T.H., & VanLehn, K.A. (1991). The content of physics self-explanations. *Journal of the Learning Sciences, 1*, 69-105.

Cobb, P. (1994). Where is the mind? Constructivist and sociocultural perspectives on mathematical development. *Educational Researcher*, 23, pp. 13-20.

Collins, A., Brown, J. S., & Newman, S. (1989). Cognitive apprenticeship: Teaching the craft of reading, writing and mathematics. In L. B. Resnick, (Ed.), *Cognition and instruction: issues and agendas* (pp. 453-491). Hillsdale, NJ: Lawrence Erlbaum Associates.

Davey, B., & McBride, S. (1986). Effects of question-generation training on reading comprehension. *Journal of Educational Psychology*, *78*(4), 256-262.

Dyer, J. W., Riley, J., & Yekovich, F. R. (1979). An analysis of three study skills: Notetaking, summarizing, and rereading. *Journal of Educational Research, 73*(1), 3-7.

Ericsson, K. A., & Simon, H. (1993). *Protocol Analysis* (revised edition). Cambridge, MA: MIT Press.

Ferguson-Hessler, M.G.M., & de Jong, T. (1990). Studying physics texts: Differences in study processes between good and poor performers. *Cognition and Instruction, 7* , 41-54.

Ferrari, M. & Chi, M.T.H. (in press). The nature of naive explanations of natural selection. To appear in a special issue on “Conceptual Development in Science Education,” *International Journal of Science Education.*

Graesser, A.C. (1993). Dialogue patterns and feedback mechanisms during naturalistic tutoring. In *Proceedings of the Fifteenth Annual Conference of The Cognitive Science Society*. (pp.126-130). Hillsdale, NJ: Erlbaum.

Graesser, A.C., Singer, M., & Trabasso, T. (1994). Constructing inferences during narrative text comprehension. *Psychological Review*, 101, 371-395.

Halliday, D., & Resnick, R. (1981). *Fundamentals of physics*. New York: Wiley & Sons.

Hamilton, R. J. (1985). A framework for the evaluation of the effectiveness of adjunct questions and objectives. *Review of Educational Research, 55*(1), 47-85.

Heinze-Fry, J. A., & Novak, J. D. (1990). Concept mapping brings long-term movement toward meaningful learning. *Science Education, 74*(4), 461-472.

Hempel, C. (1965). *Aspects of Scientific Explanation*. New York: Free Press.

King, A. (1990). Guiding knowledge construction in the classroom: Effects of teaching children how to question and how to explain. *American Educational Research Journal, 31*(2), 338-368.

King, A. (1991). Improving lecture comprehension: Effects of a metacognitive strategy. *Applied Cognitive Psychology, 5*(4), 331-346.

King, A. (1992). Comparison of self-questioning, summarizing, and notetaking-review as strategies for learning from lectures. *American Educational Research Journal, 29*(2), 303-323.

Kintsch, W., & van Dijk, T. A. (1978). Towards a model of text comprehension and production. *Psychological Review, 85,* 363-394.

Kintsch, W., & Vipond, D. (1979). Reading comprehension and readability in educational practice and psychological theory. In L. G. Nilsson (Ed.), *Perspectives of memory research* (pp. 325-366). Hillsdale, NJ: Erlbaum.

McNamara, D. S., Kintsch, E., Songer, N. B., & Kintsch, W. (1996). Are good texts always better? Interactions of text coherence, background knowledge, and levels of understanding in learning from text. *Cognition and Instruction, 14*(1), 1-43.

McNamara, T.P., Miller, D.L., & Bransford, J.D. (1991) Mental models and reading comprehension. In R. Barr, M.L. Kamil, P.B. Mosenthal & P.D. Pearson (Eds.), *Handbook of Reading Research,* Vol. 2 (pp. 490-511). new York: Longman.

Novak, J. D. (1990). Concept maps and Vee diagrams: Two metacognitive tools to facilitate meaningful learning. *Instructional Science, 19*(1), 29-52.

Novak, J. D. (1993). Human constructivism: A unification of psychological and epistemological phenomena in meaning making. *International Journal of Personal Construct Psychology,* *6*(2), 167-193.

Palincsar, A.S., & Brown, A.L. (1984). Reciprocal teaching of comprehension-fostering and monitoring activities. *Cognition and Instruction*, 1, 117-175.

Papert, S. (1991). Situating constructivism. In I. Harel & S. Papert (Eds.), *Constructivism.* Norwood, NJ: Ablex.

Pirolli, P., & Recker, M. (1994) Learning strategies and transfer in the domain of programming. *Cognition and Instruction,* 12, 235-275.

Putnam, R. T. (1987). Structuring and adjusting content for students: A study of live and simulated tutoring of addition. *American Educational Research Journal, 24*(1), 13-48.

Recker, M. & Pirolli, P. (1995). Modeling individual differences in students' learning strategies. *Journal of the Learning Sciences, 4*(1), 1-38.

Reder, L. M., Charney, D. H., & Morgan, K. I. (1986). The role of elaborations in learning a skill from an instructional text. *Memory & Cognition, 14,* 64-78.

Redfield, D. L., & Rousseau, E. W. (1981). A meta-analysis of experimental research on teacher questioning behavior. *Review of Educational Research, 51*(2), 237-45.

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

Schank, R.C. (1986). *Explanation Patterns: Understanding Mechanically and Creatively.* Hillsdale, NJ: Lawrence Erlbaum Associates.

Siegler, R.S. (1995). How does cognitive change occur: A microgenetic study of number conservation. *Cognitive Psychology, 28*(3), 225-273.

Slamecka, N. J. & Graf, P. (1978). The generation effect: Delineation of a phenomenon. *Journal of Experimental Psychology: Human Learning and Memory*, *4*(6), 592-604.

Slotta, J.D., & Chi, M.T.H. (1996). Understanding constraint-based processes: A precursor to conceptual change in physics. In G.W. Cottrell (Ed.), *Proceedings of the Eighteenth Annual Conference of the Cognitive Science Society* (pp. 306-311). Mahwah, NJ: Erlbaum.

Smith, V. L. & Clark, H. H. (1993). On the course of answering questions. *Journal of Memory and*

*Language, 32*(1), 25-38.

Stein, B. S. & Bransford, J. D. (1979). Constraints on effective elaboration: Effects of precision and subject generation. *Journal of Verbal Learning and Verbal Behavior, 18*, 769-777.

Teasley, S.D. (1995). The role of talk in children's collaboration. *Developmental Psychology,* 31, 207-220.

VanLehn, K. (1988). Toward a theory of impasse-driven learning. In H. Mandl & A. Lesgold (Eds.), *Learning Issues for Intelligent Tutoring Systems.*(pp. 19-41). New York: NY: Springer.

VanLehn, K. (in press). Rule learning events in the acquisition of a complex skill: An evaluation of Cascade. *Journal of the Learning Sciences.*

von Glasersfeld, E. (1984). An introduction to radical constructivism. In P. Watzlawick (Ed.), *The invented reality* (pp. 17-40). New York: Norton.

von Glasersfeld, E. (1989). Cognition, construction of knowledge, and teaching. *Synthese,* 80, 121-140.

Wathen, S.H. (1997). Collaborative versus individualistic learning and the role of explanations. Ph.D. Thesis, University of Pittsburgh.

Webb, N. M. (1989). Peer interaction and learning in small groups. In Webb, N. (Ed.) Peer interaction, problem-solving, and cognition: Multidisciplinary perspectives. [Special Issue] *International Journal of Education Research,* 13, 21-39.

Woloshyn, V. E., Willoughby, T., Wood, E., & Pressley, M. (1990). Elaborative interrogation facilitates adult learning of factual paragraphs. *Journal of Educational Psychology, 82*(3), 513-524.

## Appendix A: Ways of Capturing and Validating Students' Mental Models

In Chi et al. (1994) we developed a method to capture one aspect of the representation of the circulatory system, namely, the representation of the blood flow pattern (henceforth referred to as the mental model). In order to know what initial mental model of the circulatory system the students already have as they embark on the reading task, their initial mental models of the circulatory system were captured from the protocols collected from the tasks administered in the pre-test. In the pre-test, students were asked to define terms such as the atrium, the heart, valves, veins, and so forth (23 terms in all). They were also asked to draw the path of blood flow throughout the body, given an outline of the body with the requirement of including the heart, lungs, brain, feet and hands in the path. We can piece together a mental model of the connections of the components and the flow of blood on the basis of what they say. So for instance, from the definition given for the term *artery:*

> "Artery is a general term for all tubes that are from the heart and they
> carry the clean blood from the heart to all the body...it [the body] always needs
> clean blood and the blood travels through the arteries and when it's used,
> it travels back up in the veins to go back to the heart, the heart cleans it again,
> ummm, replenishes it with oxygen, umm, and then it goes again to all the
> parts of the body"

we can tell, in combination with the explanations given while drawing the blood path, that this student had the preliminary model as shown in Figure A1. This preliminary assessment of the student's mental model then underwent continuous revision and updating as we integrated additional definitions, such as the one made for *heart:*

> "The blood goes in at the upper right chamber and it then goes down to the
> downward right chamber, then it goes to the downward left chamber then
> it goes to the upward chamber then it goes out of the heart. And each
> chamber is divided by a valve that makes sure the blood goes in one direction."

These additional statements allowed us to add more details to our depiction of the student's model so that it now looks more enriched, as shown in Figure A2. After several iterations through all the protocols students generated from defining the 23 terms and drawing the blood path in the pre-test, each student was credited with a composite initial mental model.

Insert Figure A1 and A2 about here

Although a great deal of variations existed among the students' models, we were able to discern six general types, as shown in Figure A3. These six types ranged from the least accurate to the most accurate: (1) no loop, (2) ebb and flow, (3) single loop, (4) single loop with lungs, (5) double loop-1, and (6) double loop-2. The mental model depicted in Figure A4 is primarily a single loop model. From the protocols of the 24 students in the Chi et al. (1994) study (10 from the control group and 14 from the prompted group), the majority of the students (12 out of 24) had the single loop model as their initial model. Figure A3 also shows the number of students (out of 24) that had each type of models.

Insert Figure A3 about here

This initial identification of the student's mental model was further validated in the following four ways. First, our depiction of the blood path from our analyses of the protocols should be consistent with the student's own sketch of the blood path. Figure A4 shows what the student (NH) drew along with the following explanations:

> “Um, it starts at the heart...first it goes up to the brain, then...
> It goes back down to the arms, back up the arms, down to the feet,
> then back up...to your heart and then the process repeats itself...
> And I don’t think the lungs have anything to do with where the blood goes.
> I think the lungs are just there because they’re like right near the heart.”

Insert Figure A4 about here

Basically, this is a single loop in that the blood is going to all parts of the body, then returns to the heart, where the student thinks blood gets replenished with oxygen. Thus, the heart is the site of oxygenation and it is not clear what role the student thinks the lungs play.

A second way to validate our analysis of the student’s mental model was to predict, on the basis of the student’s identified initial mental model, either which of the questions in the pre-test they can answer correctly or how they would answer such questions. Thus, a subset of questions can be selected that is tailored to, and diagnostic of, each individual student’s initial model. For example, a diagnostic pre-test question for a single loop model is the following:

> “Does blood change in any way as it passes through the heart?”

A student with a single loop model would answer that the blood does change (in that it becomes oxygenated) because such an answer would be consistent with the incorrect assumption of the single loop model that the heart is the site of oxygenation, as shown below:

> “Yes. It passes through the heart, um, the stuff that needs oxygen gets oxygen,
> and then it goes through the body, and then when it comes back, it needs oxygen
> again, and it gives it more oxygen. Then it leaves the heart again. So yes,

it changes."

A third way to validate our analyses of students' mental models is to define a set of critical features for each of the six types of models. For example, some of the critical features of the single loop with lungs model are that heart pumps blood to the lungs and that lungs play a role in the oxygenation of blood. Each student's protocols were then re-read in order to ascertain that each of the critical features of the model with which the student was credited was indeed mentioned.

Finally, to further show that there is some validity and reliability in the way we have captured the students' mental model, we have shown that their mental models improved substantially from the initial to the final state. As reported in Chi et al. (1994), all of the four high explainers reached the most sophisticated double ;oop-2 model at the post-test, whereas only one out of the four low explainers did. A similar contrast is obtained for the prompted versus the unprompted students. Eight out of the nine prompted students reached the Double Loop-2 models, whereas only two of the nine unprompted students did. Hence, the mental model analysis seems valid given that the models' improvement from pre- to post-test corresponded to the students' improvement in their answers to the post-test questions.

In sum, the four sets of measures serve as different forms of validation so that we were fairly confident that our method did capture the mental models with which we have attributed to the students.

**Table 1**: **Characteristics of Repair**

1) Pauses and uncertainty statements such as “ums” at points of conflicts.

2) Monitoring statements of failure to understand.

3) Repetitions of the same self-explanations.

4) Effortful and lengthy self-explanations in contrast to short confirmatory ones.

**Table 2: Sentences Used in the Circulatory Passage**

**The Circulatory System:**

1) Human life depends on the distribution of oxygen, hormones, and nutrients to cells in all parts of the body and on the removal of carbon dioxide and other wastes.
2) These tasks are partially carried out by the circulatory system, which consists of the heart, an intricate network of blood vessels, and blood.
3) The blood moving through the vessels serves as the transport medium for oxygen, nutrients, and other substances.

**The Heart:**

4) The heart is a muscular organ that pumps blood through the body.
7) The typical adult heart beats 72 times and pumps about 5.5 liters of blood per minute.

**Structure:**

13) The heart consists of cardiac muscles, nervous tissue, and connective tissue.
17) The septum divides the heart lengthwise into two sides.
18) The right side pumps blood to the lungs, and the left side pumps blood to other parts of the body.
19) Each side of the heart is divided into an upper and a lower chamber.
20) Each upper chamber is called an atrium.
21) Each lower chamber is called a ventricle.
22) In each side of the heart, blood flows from the atrium to the ventricle.
23) One-way valves separate these chambers and prevent blood from moving in the wrong direction.
24) The atrioventricular (a-v) valves separate the atria from the ventricles.
25) The a-v valve on the right side is the tricuspid valve, and the a-v valve on the left is the bicuspid valve.
26) Blood also flows out of the ventricles.
27) Two semilunar (s-l) valves separate the ventricles from the large vessels through which blood flows out of the heart.
28) Each of the valves consists of flaps of tissue that open as blood is pumped out of the ventricle.

**Circulation of the Heart:**

30) Blood returning to the heart, which has a high concentration of carbon dioxide and a low concentration of oxygen, enters the right atrium.
31) The atrium pumps it through the tricuspid valve, into the right ventricle.
32) The muscles of the right ventricle contract and force the blood through the right semilunar valve and into vessels leading to the lungs.
33) In the lungs, carbon dioxide leaves the circulating blood and oxygen and enters it.
34) The oxygenated blood returns to the left atrium of the heart.

**Table 3: Self-explanations of the Four Sentences**

S18: The right side pumps blood to the lungs, and the left side pumps blood to the other parts of the body.

“Just that um, the right side is primarily for the lungs and the left side is to the rest of the body.”

S32: The muscles of the right ventricle contract and force blood through the right semilunar valve and into vessels leading to the lungs.

(1) “(pause) Um, the sentence is a little confusing. It’s like, it’s almost wordy, sort of, I guess. Or something. Reading it once, I don’t understand.”

[Can you try to explain, you know, put it in better words? Or words you might use?]

(2) “(pause) I mean, I guess I understand now. I just, I can’t think. I don’t know, but kind of a muscle contraction that pushed the blood, um, through the valve and into vessels, but I don’t know.”

[Does the sentence make sense with your understanding of where blood is flowing in the circulatory system?]

“(pause) Mmm, yeah.” [So what is happening here, basically?]

(3) “Um, blood is just flowing from the ventricle into vessels and going, um, to the lungs, um, ok.”

(4) “I guess, then I was remembering the thing about the left side is for the rest of the body, and the right side is mostly for the lungs, but well I guess I don’t know.”

[Well what, what are you thinking there?]

(5) “Well either, at first, I was just thinking that, you know, it was just providing the lungs with blood which didn’t make much sense.”

(6) “but that possibly, I don’t know, but maybe it’s going there [the lungs] to receive oxygen or something. I don’t know.”

S33: In the lungs, carbon dioxide leaves the circulating blood and oxygen enters it.

“OK, well then, so there, I guess so the blood goes to the lungs to um, get rid of carbon dioxide and have more oxygen put in it.”

S34: The oxygenated blood returns to the left atrium of the heart.

“So, the blood leaves the heart, goes to the lungs, and comes back.”

## Figure Captions

Figure 1. Diagram of the problem with three strings connected by a knot with a hanging block.

Figure 2. Design of the biology and the physics studies.

Figure 3. Differences between the biology and the physics studies.

Figure 4. Distribution of spontaneous self-explanations across the sentences of three worked-out examples for four high explainers for Example 5 (4A) and Example 8 (4B), four low explainers (4C, 4D), for the best and the worst learner's enforced self-explanations (4E, 4F).

Figure 5. Idealized single loop model.

Figure 6. Idealized double loop model.

Figure 7. Headings and iconic depictions of the content of S18, S32, S33, and S34 in the top row, with corresponding self-explanations generated by the students in the bottom row. The first and last columns depict the student's initial and final mental models.

Figure 8. Frequency of repair characteristics across four sentences.

Figure A1. Capturing a student's mental model of a single loop (taken from Figure 7, Chi et al., 1994).

Figure A2. Capturing a student's evolving and enriched mental model (taken from Figure 8, Chi et al., 1994)

Figure A3. Six general types of student mental models and the proportion of students having each type.

Figure A4. Blood path drawn by a student.