Journal of Science Education and Technology, Vol. 9, No. 3, 2000

The Impact of Generating Spontaneous Descriptions on Mental Model Development

Takahisa Furuta[1,2]

The purpose of this study is to explore how people's understanding about a material is created and developed as they interact with it, and to find effective ways of promoting people developing better images, when no prior information about the device is available. Participants were classified into three groups based on the score of the tasks. Obtained verbal protocols were used to determine the contents of formed models and the factors contributing the distinction in ability between the groups. The results showed that models from good learners, who finished the tasks quickly, were coherent, centered around the notion of "balance." In contrast, remarks from poor learners suggested that they did not create a mental model. There were significant differences found in the amount of generated descriptions that characterized responses from the device using the participant's subjective expressions. The good learners introduced significantly greater number of new vocabulary words compared to the poor learners. Generating such descriptions seems like the source of power for the good learners.

KEY WORDS: Mental models; description generation; empirical learning.

INTRODUCTION

People, in many cases, have images of the device they are using. The images include components of the material, how they are configured, and how they work as a whole. Such images allow people to produce proper operation procedures, predict future events, and in case of emergency, formulate actions to deal with such an accident. For example, people have an image about the mechanism of a car. People know that a car has an engine and four wheels with tires. They also know that the engine generates power that drives the wheels, and thus the car moves. People use the image to infer that, for example, when the car will not move, there might be something wrong with the engine, which suggests them to open the hood. Images described above are called mental models in cognitive studies. Mental models are supposed to not only support our daily lives, but consist of essential part of our knowledge, therefore they are sometimes a synonym to understanding.

Mental model research has been an active area in cognitive studies, attracting researchers with various backgrounds, including industrial designers, educators, and developmental psychologists. Their studies have revealed that people rely on their own models to determine the next action to take on the material. They show that people have a lot of mental models of various materials, as diverse as handheld computers (Young, 1983), thermostats (Kempton, 1986), and the earth (Vosniadou, and Brewer, 1992). For instance, even children as young as 4 years old explicitly refer to a model of the earth and the sun to explain the day/night cycle (Vosniadou and Brewer, 1994). Also, studies have revealed that mental models are sloppy, inaccurate, and unstable (Norman, 1983). People rely on their wrong images of the mechanisms of thermostats that make them to put the temperature setting at the highest temperature when they wish to warm the room quickly (Kempton, 1986). People can for-

---

[1] College of Education, University of Gunma, Maebashi, Japan.

[2] To whom correspondence should be addressed at College of Education, University of Gunma, 4-2 Aramaki, Maebashi, Gunma, 371 Japan.

1059-0145/00/0900-0247$18.00/0

mulate a model of an imaginary artifact out of prior experiences. People can generate a procedure to blast a beam against an enemy's spaceship (Kieras and Bovair, 1984).

One of the issues that have not been much addressed so far in the literature is how models are created and developed when no prior information is available, because most of this research focuses on such issues as identifying the types of models people have, how people use models, or effective ways of teaching models. Very few of it deal with the on-line processes of mental model development. However, it occurs often for us to face a material without prior knowledge. For instance, we sometimes intentionally refuse to receive prior information. We sometimes start using a device without reading the manual, although we have no prior experience using it, except similar ones, thinking that we can come back to the manual anytime needed. Although it is evident that people have the ability of creating their own mental models out of prior experiences while interacting with it, it is not still clear how they do it.

Revealing the cognitive processes that underlie mental model formation would also benefit those who are not very good at understanding something. Because mental models are involved in our diverse cognitive activities including understanding written passages (Albrecht and O'Brien, 1993) and solving physics problems (Anzai and Yokoyama, 1984), the processes of creating them should not be restricted to device operation contexts. Rather, they should be seen as an essential part of understanding things that occur normally in our cognitive activities. Therefore, findings from this study would help students who are not good at understanding something, by showing them hints for devising their models.

Therefore, the purpose of this study is to explore how people's understanding about a material is created and developed as they interact with it, and to find effective ways of promoting people developing better images, when no prior information about the device is available. In the experiment reported here, participants were given a computer simulation of a device unknown to them and were requested to control it to achieve tasks. The participants were thus forced to form their own understanding of the device by interacting with it. This is a case that occurs normally when we face a novel event that forces us to figure out what is happening. The participants were classified into three groups based on the score of the tasks. Obtained verbal protocols were used to determine the contents of formed models. The protocols were also decomposed into categories to characterize the groups to determine the factors contributing the distinction in ability between the groups.

Note that, because there is no commonly agreed definition of mental model (Rouse and Morris, 1986), it is necessary to state the assumptions made about the nature of mental models in this paper. First, a mental model is dynamically created on the spot for the purpose of explaining behavior or predicting the future (Vosniadou and Brewer, 1992). Second, models need not be technically accurate, but they are functional, incomplete, unstable, and unscientific (Norman, 1983). Third, models are aggregations of smaller pieces of knowledge. Each such piece is fragmentary and contains less information, but the way they are organized and work together makes the aggregation fairly powerful (diSessa, 1993). Very few previous studies have made it explicit that mental models have all of these characteristics.

EXPERIMENT

Method

Material

The external appearance of the device used in the experiment is shown in Fig. 1. The obscure appearance of the console made it difficult for the participants to see the device was actually a simple water tank system simulated by a PC (Fig. 2). The tank system had two tanks with taps, connected by a pipe at the bottom. The right tank had an outlet at the bottom, which allowed water in the right tank to flow out continuously.

The two knobs controlled the valves of the taps that changed the amount of water flowing into the tanks. Each tank had the "target height" at which the participants were required to attain the water level, and the lamp colors reflected the differences between the actual and target heights of the water. Namely, when the height of the water in a tank was at the target height with a tolerance of plus or minus 5%, the lamp lit red. When the water level was below the target height within the range of 5–30%, the lamp was green. And when the water level was over the target level within the range of 5–30%, the lamp was blue. Otherwise, the lamp was black. There was no gradation in the color so that the participants could only know the range they were in.

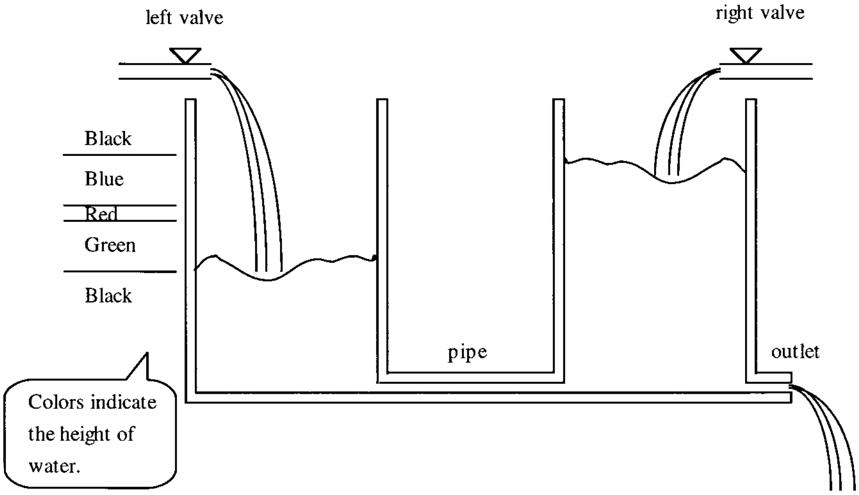

Fig. 1. The structure of the watertank system.

Because it takes time for water to rise and fall, delay in response is inevitable. Therefore, the optimal solution was to turn the knobs for certain amounts, and then keep waiting for red in the lamps, as the total amount of incoming water from the taps and the outgoing water from the outlet brings the entire system to equilibrium. However, facing with the console without knowing that it was actually a water tank system, people tended to expect quick responses in the lamps when they turned the knobs, so that they were puzzled to find the system would not respond in the first place. Therefore, the participants were forced to change the conception they had toward the device at some point, which was the reason this system was chosen for the experimental material.

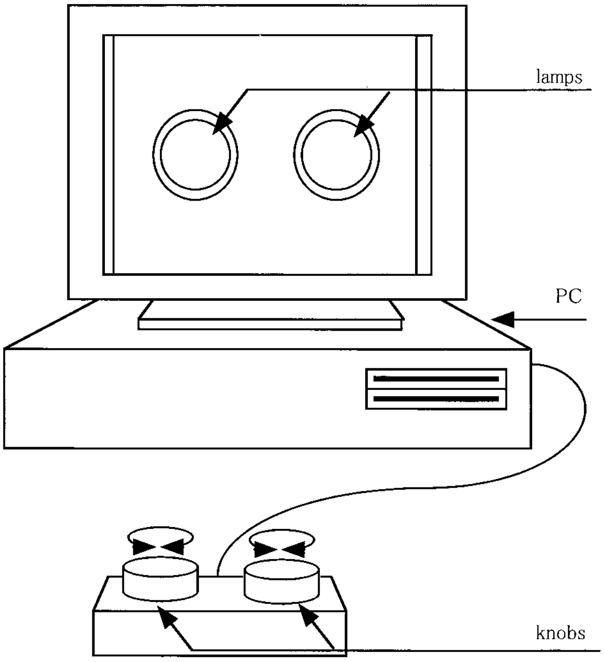

Fig. 2. The external appearance of the console.

Subjects

Eighteen male and female Japanese undergraduate and graduate students were recruited for paid participation. Each participant received 1000 yen (about US$10). None of them had experienced controlling devices like the one used in this experiment, and none had participated in a psychological experiment before. Two experimental groups (the informed group and the not-informed group) were set. The informed group was told that the target system was actually a water tank system whose inner structure was Fig. 2. This group also received information such as the relation between the knobs and the valves, and the relation between colors of the lamp and water height. In contrast, the not-informed group was not given this information, and their requests for information about the device were rejected. The informed group was set to confirm the impact of having appropriate knowledge about the target device. The not-informed group was aimed to reveal the processes of developing mental models, which demanded intensive analyses of verbal protocol. Thus, 6 participants were assigned to the informed group, and 12 were assigned to the not-informed group.

Procedure

The experiment was conducted individually. First, the participants received simple training in talk-aloud protocols. Second, they were told that their goal was to keep the both lamps red as long as possible by rotating the two knobs. Third, the participants in the informed group were given information about the system, while the not-informed group participants skipped this step. Fourth, the participants proceeded to the training task, which requested the participants to make both lamps red for at least 60 seconds out of 120 seconds for two successive trials. The participants did not have to keep the lamps red for 60 seconds in a row; all the segments of time being red during a trial were added up to determine the duration. If the participant failed within 20 trials, training was abandoned at that point. Finally, the participants moved to two transfer tasks. One of the tasks was the longer duration task, which requested the participants to make both lamps red for at least 100 seconds out of the 120 seconds. This task allowed shorter spare time, so that quicker operations were needed. The other was the no-blue task, which requested the participants to keep both lamps red for at least 60 seconds out of the 120 seconds without letting either of lamps becoming blue, which was equivalent to setting an upper limit for the water height. Therefore, contrary to the longer duration task, slower operations were preferred, so that the participants could stop turning the knobs as soon as the lamps became red. In transfer sessions, the participants switched to the other task when they met the condition once, not twice, as they were required in the training session. If the participant failed for five successive trials, they proceeded to the other task. The order of the tasks was counterbalanced between participants.

The participants talked aloud during all the tasks, and their entire problem-solving protocol was recorded for later analyses. However, due to trouble in the recording device, verbal data from one participant in the no-information group was lost, so that data from this participant was excluded.

RESULTS

Table I. Number of Trial for Criterion (S.D. in Parentheses)

	Task		
Condition	Training	No blue	Longer duration
Informed	2.5 (.55)	2.0 (.89)	1.2 (.41)
Not informed	14.0 (5.56)	4.1 (1.98)	3.8 (2.18)

Analyses went as follows. First, the scores from the informed and not-informed groups for the training and transfer tasks were compared to show the significance of mental model in the experiment. Second, the verbal protocols from the not-informed group, who started the experiment without given information, were analyzed to reveal the processes that formed mental models.

Validity of the Material

The results are shown in Table 1. Because the Informed group took significantly fewer number of trials to complete all the tasks (chi^2(1) = 11.54, P = 0.007 for the training task; chi^2(1) = 4.07, P = 0.044 for the no-blue task; chi^2(1) = 5.86, P = 0.015 for the longer-duration task[3]), it is shown that it's difficult to operate the device without prior information. This result confirms the expectation that the device is an appropriate material for studying the cognitive processes of building understanding.

Participants Classification

Because some not-informed participants achieved the transfer tasks while some totally failed, it would be reasonable to assume there were good learners and poor learners in the not-informed participants. Therefore the participants were classified into three groups based on their combined score for the two transfer tasks. Three participants with least numbers of trials belonged to the good learner group, three who failed the transfer tasks belonged to the poor learner group, and the rest (N = 5) belonged to the middle learner group. Performance measure for the three groups is shown in Table II.

---

[3] Due to the small and unbalanced numbers of participants, non-parametric statistical tests, Kruskal-Wallis tests, were performed.

Table II. Number of Trials for the Three Participant Groups (S.D. in Parentheses)

Group	Training task	Transfer tasks (combined)
Good	14.3 (4.62)	3.7 (.58)
Middle	12.0 (5.15)	7.8 (1.64)
Poor	18.3 (2.89)	10.0 (.00)

Although the three groups finished the training task at a similar number of trials ($chi^2(2) = 3.53$, $P = 0.171$), the number of trials taken in the transfer tasks were significantly different ($chi^2(2) = 8.19$, $P = 0.017$). Post-hoc Mann-Whitney U tests for comparisons between the good and the middle groups showed a marginally significant difference ($U = 1.5$, $P = 0.055$), and a significant difference between the middle and the poor groups ($U = 0.00$, $P = 0.024$). Therefore, the three groups finally gained significantly different knowledge about the target system when finished training.

Types of Formed Mental Models

It would be critical to show that the participants of the not-informed group actually formed mental models to study the processes of model formation, in the first place. Models were assessed in their concurrent verbal reports of their thought with following criteria: (1) if remarks contained explicit reference to devices, imaginary or real, to explain the participant's action or the device's response, or, (2) if remarks provided partial mechanisms of the device the participant supposed to underlie. The second criterion was set so that even fragmentary, unstable, and unsolid models could be counted.

Major findings were as follows: First, there was substantial evidence that 73% (8 of 11) of the not-informed participants formed models of their own during experiment. Therefore, the data is worth analyzed for model formation processes. Second, the models the participants formed were, as expected from previous studies, obscure, inaccurate, and unstable, and the relationship between the model and the actual system was not very clear. Third, the participants groups set in the previous section were also a good index of coherency of mental models. Namely, models from the good learners were all centered around the notion of "balance," and that they all showed high coherency by referring to imaginary devices. In contrast, remarks from the poor learners mostly consisted of factual reports that were not always consistent one another and that none of their remarks suggested they really used a model. Models from the middle group were somewhat mixed; some referred to imaginary devices, whereas others made no such reference.

The model portions of the remarks from the not-informed group are shown below, classified for the three learner groups.

Models from the Good Learners

The good learners tended to explain using analogy to some devices whose core notion was balance. Below are examples of the remarks.

> This is probably a kind of a balance. I must start slowly, not too fast.
>
> It is like controlling a wheel that has two strings attached at both ends. You pull both strings properly and the wheel becomes horizontal, and, that's when the light turns red.

The participants did not give precise accounts of the structure of the target device, but they insisted that they captured essentials about the device. Note that these balance models are not always correct because there is no physical correspondence between the water tank system and a balance or the complex wheels. However, the models are correct in terms of functional aspects because keeping a balance promotes derivation of fine tuning procedures, which requires close attention to the response and slow, careful adjustment of the input.

Models from the Middle Learners

Models from the middle learners were diverse in contents and there was no clear indication that the models shared the same notion among the participants as found in the models from the Good learners. However, some tended to use analogy, too, to explain their models. Below are examples of the remarks:

> It's like playing tag. Since this device frequently changes the angles where the knobs should be, I have to keep turning them.
>
> I feel like opening a safe. Proper combination of turning the knobs must be the key to success.

Not all the participants made remarks as explicit models as above, which suggests that they finished the task without forming a model in the experiment, which was also the case for the poor learners.

Models from the Poor Learners

None of the poor learners made explicit remarks that suggested they were using conceptual models as integrated as the above two groups. Their remarks mostly consisted of factual reports of the way they were turning the knobs and the responses of the device, such as:

> I am turning the left knob to the right (clockwise).

However, there were remarks that suggested some were using partial or weak models. Below are examples of such remarks.

> I found the lamp illuminates blue after red.
>
> The next color to red is blue.

Furthermore, some remarks suggested that some participants gained partial insights into the nature of the system. Below is an example, which was wrong.

> The system seems to be sensitive to the speed of the knob turning

Because their remarks remained fragmentary throughout the experiment and contradicted among others, it is not very clear to what extent their remarks came from the same central idea about the system.

Identifying the Elements in the Cognitive Processes

To identify the factors that contribute to forming adequate mental models, focus of analyses is place on the distinction between the good and poor learners. Subsequent analyses of protocols were similar to the manner employed elsewhere (Wineburg, 1998). First, protocols were parsed into idea units. Then protocols were systematically reviewed to create a coding scheme. Working hypotheses were developed, checked, refined, and rechecked in subsequent reviews. The codes were created to maintain theoretical significance but never stray far from the empirical data. The results indicate that there were two major differences between the groups. First, the two groups differed in the types of goal they set, and the knob strategies they employed. Second, not only were the good learners more sensitive to the response of the device, but they also produced more new vocabulary that stated interpretations to responses from the device or impressions the participants perceived. Those descriptions were in many cases subjective characterizations and they were generated during the course of operation, and not all of them were correct. I term such remarks as New Vocabulary (NV). A major role of generating NV seems to be accessing one's prior knowledge that further trigger retrieval of other prior knowledge that may constitute models. In the following sections, developmental processes of mental models are outlined in terms of goal, knob strategy, sensitivity to delay, and new vocabulary.

Goal Orientation

Goal orientation here refers to the goal of one's action, and is directly related to search control strategies. Two major goals, browse and problem solving were identified. Browse is to accumulate cases of cause and effect relations between the input (knobs) and the output (lamp colors) when one is interested in how color changes after one's action. Browse should be considered as participants' efforts to gain an overall image of the target device. In contrast, when a problem-solving goal is set, one seeks to attain a specific state of the device, such as the left lamp being green. Namely, a problem-solving goal should be considered as one's efforts to solve specific problems.

It would be reasonable to assume these two goals when one is learning to operate a device. For example, Vollmeyer, Burns, and Holyoak (1996) shows that one takes the "Vary One Thing At a Time" (VOTAT) strategy when no specific goal is presented, and one uses the difference reduction (DR) strategy when a specific goal is presented. With the VOTAT strategy, one changes only one value of all the variables one is allowed to change to test every combination of variables, so that one can identify the relation between each variable and its effect. This strategy is known to occur often when a participant's emphasis is on understanding the device itself. By contrast, DR strategy is thought an implementation of the means-ends strategy, and one tries to reduce the difference between the present situation and the goal, because they are concerned with solving the task itself. It could be said that a similar pattern of behavior of participants appeared in the results reported here.

Knob Strategies

Knob strategies concern the manner of turning the knobs, and three major strategies, angle (amount of rotation), speed, and combination were identified. Angle is to fix the knob at a certain angle and then let go of the knob, which is the same as opening or closing a faucet. Speed is to keep turning the knob at a certain speed, turning it in reverse when the knob hits an end, based on the idea that what is relevant to the device is the speed of rotation.

Combination is similar to opening a safe: the knobs are supposed to be fixed at several different angles in order, pausing for a while at each angle. For example, when the participant sets a knob at 45 degrees from the left end, then 135 degrees, and then 90 degrees, it is considered a combination strategy.

Delay in the Response

As stated in the introduction of this paper, the reason I chose this water tank system was because the delay between the input (knob) and the output (lamp) made the task difficult without adequate mental images. In other words, as soon as the participant finds the slow responsiveness of the system, s/he is more likely to be able to generate better operation procedures. Therefore, finding delay is critical for developing a model for the water tank system. However, because all the participants initially assumed the system responds quickly to their input, their finding of delay came rather later in the trials.

New Vocabulary

There are striking differences found in the number of remarks that introduced vocabulary that did not appear in the instruction. I termed them new vocabulary. Remarks shown below are examples.

> The figure on the screen looks like *a goldfish.*
>
> The figure on the screen looks like *a signal tower.*
>
> Something like *the derivative* of the knob turning seemed critical.
>
> It seemed like it was necessary to keep *the equilibrium.*

The participant who made the first instance later explained that the two lamps on the screen were goggle, which reminded her of a goldfish. Most NV were subjective and superficial in content. Some appeared among all the participants, while some were peculiar to an individual. Note that although some of the NV indicated that the participant had captured the fundamental nature of the system such as balancing, not all NV did, as ‘‘goldfish’’ indicates.

Note that I excluded ‘‘speed’’ and ‘‘angle’’ from this category, though they did not appear in the instruction given to all the participants prior to the experiment, because both learner groups mentioned them. Therefore, it would be reasonable to assume that these two were rather trivial, and that they did not contribute to forming models since even the poor learners, who appears to have failed to form models, mentioned them.

Both groups produced almost the same numbers of statements ($t\ (6) = 0.44$, $P > 0.1$), however, there were significant differences in distribution. Not only is the absolute number of NV generated by the good learners significantly greater than by the poor learners (7.0 for the good learners, 1.8 for the poor learners; $t\ (6) = 2.75$, $P < 0.05$), but proportionately, the good learners generated five times as many NV as the poor learners (11% vs. 2%).

Differences in the Distribution of the Elements

Table III shows the distribution of the above mentioned items during the training trials. In the analysis, for each participant, one total course of the training trials was evenly divided into three segments: ‘‘beginning’’ corresponds to the first one third, ‘‘middle’’ corresponds to the second one third, and ‘‘ending’’ corresponds to the last one third of the trials. I put an ‘‘x’’ in the table if all the members of that group made at least one remark relevant to the category. The reason for putting ‘‘x’’s instead of averaged figures into the cells is that the average can be calculated even if not all of the individuals actually contributed the number. An ‘‘x’’ is stronger evidence of presence of the idea among the participants.

There are several striking differences found in the table between the two groups: (1) The good learners employed the three types of knob strategies at the beginning of the trials and fewer of them in the middle, whereas the poor learners employed speed strategies at the beginning and all the three in the middle. (2) The good learners employed browse strategies in the middle of trial, which was not always the case for the poor learners. (3) The good learners correctly came up with the angle strategy at the end of the trials, whereas the poor learners relied on the

Table III. Distribution of Remarks[a]

	Good learners			Poor learners		
	Beginning	Middle	Ending	Beginning	Middle	Ending
Knob strategy						
Angle	x		x		x	
Speed	x			x	x	
Combination	x				x	x
Good orientation						
Browse	x	x		x		
Problem solving	x			x		
Responsiveness						
Slow			x		x	x
N.V.		x	x			

[a] An "x" indicates that all the members in the group mentioned the item.

combination strategy instead. (4) The poor learners made significantly fewer NV implications of these findings are discussed below.

DISCUSSION

Summary of Results

Main findings of the current study can be summarized as follows: (1) There were good and poor learners among the participants, as well as middle learners, that were distinguished by the scores for the transfer tasks. (2) The types of mental models differ significantly among the learner types. The models from the good learners were most coherent, and they shared "balance" as the core notion, whereas the models from the poor learners were least coherent, suggesting that they did not form models. (3) There were differences in the distribution of operation strategies, types of goals, and sensitivity to response of the device between the good and poor learners. The good learners relied more on browse strategies that suggest they tried to understand the device, whereas the poor learners seemed to have focused on finding a single correct solution procedure, rather than to understand the device. (4) There were significant differences in the amount of generated descriptions, or NV. The good learners produced significantly greater number of NV compared to the poor learners.

The Cognitive Processes of Mental Model Formation

In cognitive studies, it has been widely accepted that there are two types of processes involved in cognitive activities. One is top-down and the other is bottom-up. The former is also called hypotheses-driven process while the latter is called empirical learning process. As Table III indicates, the good learners used browse strategies more that the poor learners throughout the experiment, which was indicative of empirical learning processes. Along with the fact that the good learners produced more NV indicates that generating descriptions in empirical processes may help the good learners formulate mental models. Namely, the good learners sought to associate the empirical patterns to their prior knowledge, while the poor learners tried to solve the problem by finding a correct operation procedure.

The basic processes of forming models in this study can be outlined as follows. When the participant has a slight idea what the device is, s/he employs top-down processes to exploit it. One tries to confirm one's expectation that is derived from the hypothesis by detecting specific patterns in output which appears in response to one's specific input. In this case, one is more focused on the patterns in output, which are constrained by the current hypothesis. In contrast, when one runs out of hypothesis, then one can no longer rely on top-down processes because no guiding hypothesis is available, so s/he uses empirical processes to collect regularities in input and output and tries to reproduce the same output by giving the same input.

The good learners made more NV remarks, fewer remarks on the knob strategies compared to the poor learners. The good learners made fewer remarks on knob strategies, probably because their cognitive capacities were devoted to picking up and interpreting system's responses and associating them to prior knowledge that resulted in greater amount of NV. Moreover, as mentioned before, browsing type of strategies are associated with empirical learning processes. Therefore, the existence of browsing strategies in combination with NV suggests the good learners were more likely to succeed in associating data with their prior knowledge. In contrast, the poor learners spent their time testing every combination of operation in search of the correct answer, which did not necessarily require accesses to conceptual schemata. Thus, they ended up the task with procedural solving strategies.

Roles of Generating Descriptions

The explanation that a role of producing greater number of subjective NV may be to increase the chance of accessing distant part of memory has commonality with the effects of self-explanation, in which the good students who solved physics problems well produced more self-explanations while solving them, while the poor students produced significantly less self-explanations (Chi *et al.*, 1989). My speculation is that the results of her and my studies address the issue of enhancing knowledge accessibility. The good learners, who generated more NV, were likely to pick up more patterns in the responses and gave interpretations of their own, which led them to mental models that were capable of integrating these patterns. By contrast, the poor learners made either no feature pick up or no interpretation, thus they had lesser chances of improving their models, even though all of them noticed the delay in the system sooner than the Good learners.

It seems that the good learners not only took advantage of their already formed hypothetical models, but were in a position to be able to create new one. By contrast, the poor learners spent their time testing every combination of operation in search of a single correct solution, which did not necessarily require accesses to conceptual knowledge, thus they ended up the task with procedurally solving strategies, so that they were in lesser need of conceptual models. Generating descriptions of behavior seems like the source of power for the Good learners.

Educational Implications

One of the educational implications of the present results is that encouraging students to characterize what they see may help them gain deeper understanding of the material they are learning. This suggestion is based on the idea that understanding consists of two distinctive but interrelated processes of generation of fragmentary ideas and integration of them, which is fairly ubiquitous cognitive activities, as studies on other areas such as the writing processes shows (Flower and Hayes, 1981).

Because the good learners made greater numbers of NV, encouraging students to make their own characterizations and interpretations as much as possible might help them to notice different aspects of the material, compared to just letting them sit around and do nothing particular, by increasing the chance of accessing various portion of their prior knowledge.

It may be sufficient for some students to encourage generating their own descriptions. The next step necessary for students who can not go ahead by producing self-descriptions alone would be to help them organize the fragmentary ideas into a coherent set. Systems such as ConvinceMe (Ranney *et al.*, 1995), which helps users learn to increase the coherency of their belief sets about events or phenomenon through interacting with it using ratings that are calculated by a normative formula, or TEC in this case, could be a powerful tool for this purpose.

ACKNOWLEDGMENTS

The author is grateful to Michael Ranney, Andrea diSessa, and the members of the Reasoning group at University of California, Berkeley, for their helpful comments.

REFERENCES

Albrecht, J. E., and O'Brien, E. J. (1993). Updating a mental model: Maintaining both local and global coherence. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 19: 1061–1070.

Anzai, Y., and Yokoyama, T. (1984). Internal models in physics problem solving. *Cognition and Instruction* 1: 397–450.

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

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

diSessa, A. A. (1993). Toward an epistemology of physics. *Cognition and Instruction* 10: 105–225.

Eberts, R. E. and Bittianda, K. P. (1993). Preferred mental models for direct manipulation and command based interfaces. *International Journal of Man-Machine Studies* 38: 769–785.

Flower, L., and Hayes, J. R. (1981). A congnitive process theory of writing. *College Composition and Communication* 32: 365–387.

Kempton, W. (1986). Two theories of home heat control. *Cognitive Science* 10: 75–90.

Kieras, D. E., and Bovair, S. (1984). The role of a mental models in learning to operate a device. *Cognitive Science* 8: 255–273.

Ranney, M., Schank, P., and Diehl, C. (1995). Competence versus performance in critical reasoning: Reducing the gap by using Convince Me. *Psychology Teaching Review* 4: 153–166.

Rouse, W. B., and Morris, N. M. (1986). On looking into the black box: Prospects and limits in the search for mental models. *Psychological Bulletin* 100: 349–363.

Vollmeyer, R., Burns, B. D., and Holyoak, K. J. (1996). The impact of goal specificity on strategy use and the acquisition of problem solving. *Cognitive Science* 20: 75–100.

Vosniadou, S., and Brewer, W. F. (1992). Mental models of the Earth: A study of conceptual change in childhood. *Cognitive Psychology* 24: 535–585.

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

Wineburg, S. (1998). Reading Abraham Lincoln: An expert/expert study in the interpretation of historical texts. *Cognitive Scinece* 22: 319–346.

Wisniewski, E. J., and Medin, D. (1991). Harpoons and long sticks: The interaction of theory and similarity in rule induction. In Fisher, D. H., Pazzani, M. J., and Langley P. (Eds.), *Concept Formation: Knowledge and Experiences in Unsupervised Learning*, Morgan Kaufmann, San Mateo, CA.

Young, R. M. (1983). Surrogates and mappings: Two kinds of conceptual models for interactive devices. In Genter D., and Stevens, A. L. (Eds.), *Mental Models*, Lawrence Erlbaum, Hillsdale, NJ.