Journal of Science Education and Technology, Vol. 12, No. 4, December 2003 (© 2003)

Mental Models and Computer-Based Scientific Inquiry Learning: Effects of Mechanistic Cues on Adolescent Representation and Reasoning About Causal Systems

Danielle E. Kaplan[1,2] and John B. Black[1]

This research applies cognitive science to the development and study of computer-based scientific inquiry learning. A scientific inquiry software program designed in the domain of elementary hydrology was adapted for mental model reasoning research, and tested in two middle school science classes. The study explores how qualitative mechanistic cues about system factors influence mental animation of system mechanisms and reasoning about causality. Middle school groups were compared on model development, inquiry, prediction, and learning. Students provided with mechanistic cues during inquiry developed more complex models with significantly more animated explanations of how and why causality exists. When not provided with mechanistic information, students reduced the level of complexity and animation in models during inquiry. Girls started with more complex and animated models than boys and reduced the level of complexity and animation in models during inquiry, whereas boys increased the level of complexity and animation in models. Students provided with mechanistic cues had more accurate theories after inquiry than students not provided with mechanistic cues. There was a trend toward use of better inquiry strategies and more accurate prediction in girls provided with mechanistic cues. Level of animation in model descriptions was a significant predictor of developing accurate theories.

KEY WORDS: scientific-inquiry; educational-technology; cognition; reasoning; system-studies; learning.

INTRODUCTION

Einstein and Infeld (1938, cited by Zukav, 1979, in Johnson-Laird, 1983) liken our effort to know reality of trying to understand the mechanism of a closed watch. The face and moving hands are seen. The ticking is heard. But there is no way to open case. They suggest an "ingenious" act would be to "form a picture of a mechanism which could be responsible for all observed." Einstein and Infeld imply intelligent scientific discovery involves constructing a mental representation of an external mechanism to account for individual observations.

The goal of this experiment is to investigate whether providing mechanistic cues would induce greater complexity and mechanistic representation of factors and relations in mental models, leading to advances in reasoning and learning. Can mechanistic cues designed to elicit mental depiction of mechanistic qualities of system components facilitate generation of "correct" qualitative mental animations of system mechanisms? Do inquiry programs providing representational pieces of a system under investigation lead to greater visualization of more pieces in a model and better reasoning? Mechanistic cues hinting at how to represent components and causal relations between components would lead to more integrated causal models made up of subcomponent models in mental representation. Constructing a cohesive mental structure accounting for observations and illustrating explanations

[1]Department of Math, Science and Technology, Teachers College, Columbia University, Box 8, 525 W 120th Street, NY, New York, 10027.
[2]To whom correspondence should be addressed; e-mail: danielle.kaplan@columbia.edu

for observations could provide an organizational framework for making inquiries and inferences about observations.

Craik (1943) presents the notion of a mental model as an internal representation of a system. According to Johnson-Laird (1983), people rely on mental models to make deductions (Johnson-Laird, 1983; Johnson-Laird and Byrne, 1991), and reasoning could be based on semantic rather than syntactic method of formal rules. Content of premise can affect deductive performance (Wason and Johnson-Laird, 1972). Mental models account for effects of meaning, and the ability to reason in unfamiliar premises (Byrne, 1992).

de Kleer and Brown (1983) describe mental models as qualitative simulations of a complex physical device made up of simpler components and mechanisms. The structure of a static system informs one about the states of components, and the behaviors of relationships. Simulation of components and component relations allows for the generation of inferences about system relationships. Schwartz and Black emphasize imagery in mental models, describing "depictive simulations" as representations that simulate system mechanisms, from which rules can be derived (1996a; 1996b).

de Kleer and Brown (1983) differentiate between simulation as a process and as an artifact, using the word "envisioning" to refer the simulation process and "causal model" to refer to the artifact of simulation. "Envisioning" allows one to determine the function given the structure and principles, and to determine the behavior for each component, given model characteristics. The "causal model" describes functioning and the model must be developed before it can be "run" to produce a certain effect. "Running" involves developing a causal model to produce a certain effect "Running" of a mental model occurs when autonomous objects change states, thus influencing other autonomous objects (Williams *et al.*, 1983).

This line of research led to the following questions:

- How would qualitative cues about system components presented in images and text during computer-based scientific inquiry influence mental representation? Would mechanistic depiction of components lead to envisioning and running of models?
- How would qualitative cues, influencing mental representation, affect reasoning and learning? Would more animated representations advanced inquiry and learning?

Program Design

Flood Predictor is an educational and research software program, created with Macromedia Director, which supports the investigation of causal relationships in the domain of elementary hydrology (Kuhn *et al.*, 2000). Scientific reasoning activity is based upon Piagetian reasoning tasks (Inhelder and Piaget, 1958; Piaget, 1972), in which multiple factors contribute to creating an effect (i.e. a pendulum swing), and previous studies by Kuhn and Schauble (Kuhn, 1989; Schauble, 1990; Kuhn *et al.*, 1992), in which students make inquiries and predictions, and revise theories about a multivariable system.

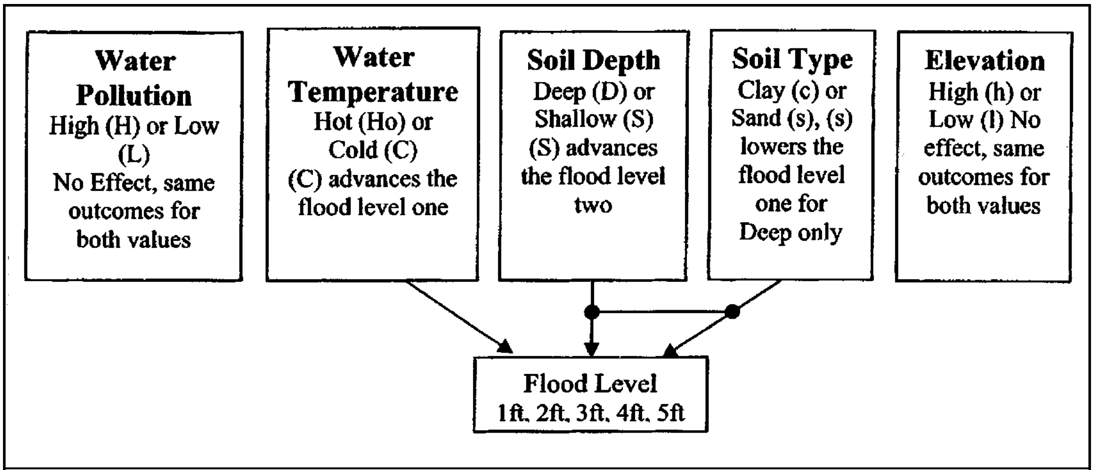
In *Flood Predictor*, participants are guided into deductive reasoning strategies while investigating causes of flooding. Anchoring instruction has been found to be effective in science education (Goldman *et al.*, 1996). Applying Goal-based Scenario constructivist design principles (Schank *et al.*, 1994), students are placed in the role of a builder working for a construction company, and assigned the job of determining how high to build stilts under houses near a group of lakes. Employees are informed that, in order to avoid flood damage and to minimize the cost of expenditure on unnecessary materials, they must determine which factors in the region will cause flooding and which will not.

Five factors are introduced as potentially causal in affecting flood level: Water Pollution (high vs. low), Water Temperature (hot vs. cold), Soil Depth (shallow vs. deep), Soil Type (clay vs. sand), Elevation (high vs. low). Three of these factors (Water Temperature, Soil Depth and Soil Type) are causal within the program scenario.

Table I summarizes the causal structure of the *Flood Predictor* system, and outcomes of specific combinations. There is one interaction in the program, between Soil Depth and Soil Type. Water Pollution and Elevation are not causal within the problem space.

Discoveries can be made by calling up records of sites by creating unique combinations of features, predicting how high flooding will rise, and making conclusions about whether features matter, do not matter, or had not found out yet. Figure 1 is a screen shot of the interface for calling up and comparing records. After each instance of examining records, students are queried about how they know that certain features

Table I. Casual Structure of Flood Problem & Outcome Combinations



Outcome Combinations (Site Flood Levels)

1ft	2ft	3ft	4ft	5ft
Ho,D,s	Ho,D,c C,D,s	C,D,c	Ho,S,c Ho,S,s	C,S,c C,S,s

matter or do not matter in causing flooding. Finally, researchers are asked to report theories. Activity within the program is tracked to assess student strategic performance and knowledge acquisition.

A supplementary "Field Report" with qualitative information about the mechanism of each factor-relation in static images and explanatory text was designed for this research and future modification of *Flood Predictor.* Mechanistic Cues are summarized in Table II. Mechanistic Cues were intended to elicit imagistic depiction of factors leading to animated illustrations of how and why factors do or do not affect flood level. Mechanistic Cues were also designed to counter "incorrect" models of causality derived

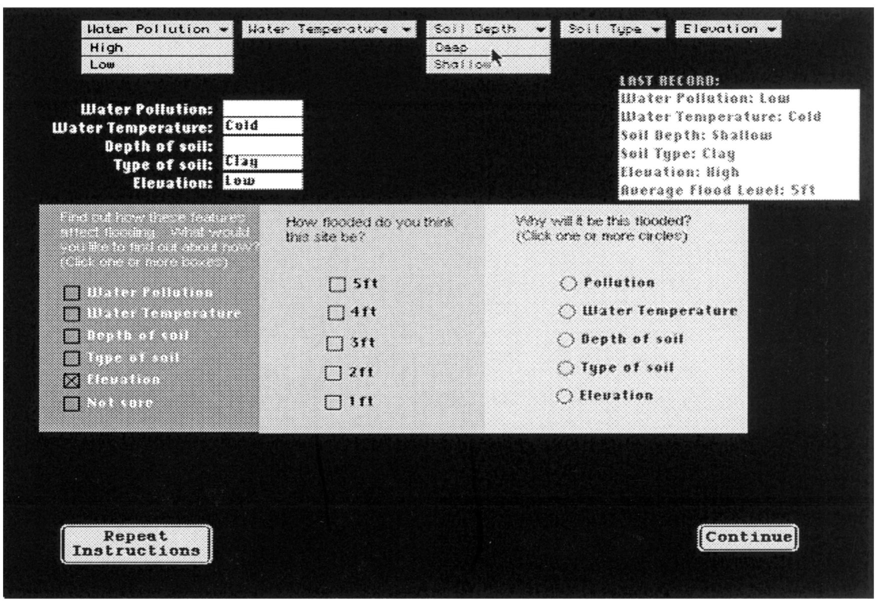


Fig. 1. Inquiry interface.

Table II. Mechanistic Cues in "Field Report" and Expected Inferences

Factor	Mechanistic Information	Inference
Water Pollution—High (H) or Low (L)	All of the pollution on these particular lakes, in high and low elevations, is soluble—it dissolves, like sugar in tea or salt in soup	Pollution will dissolve in water, whether a lot or a little, not adding to water level
Water Temperature—Hot (Ho) or Cold (C)	When really cold, it hardens like ice. When really hot, it evaporates into air	Evaporation will lead to thinking Hot causes less flooding, ice will lead to thinking Cold causes more flooding, same conclusion either way
Soil Depth—Deep (D) or Shallow (S)	Water can move easier through soil than through rock. Deep soil has more soil until hitting rock than shallow soil	In Deep soil, there is more room for water until hitting less permeable rock
Soil Type—Clay (c) or Sand (s)	Sand has bigger grains than Clay. Clay has smaller grains than sand	There is more room for water to drain/fit with Sand, so more flooding with Clay
Elevation—High (h) or Low (l)	There are hills and valleys in high and low elevations	Equal levels of flooding can occur in valleys at both high and low elevations

through analysis of flood descriptions from a previous pilot study. For example, a common "incorrect" (in the program) model of water pollution's effect was that increases in pollution would take up water space and make the water rise. So the field-report notes that all pollution in the region is soluble.

The "field report" does not give the causal mechanisms. Rather, it provides information from which the mechanistic inferences could be derived by animating the mental model. Making the correct inference would involve depicting Mechanistic Cues and require mental movement in the model. For example, a correct mechanism inference that Soil Type causes flooding because water flows easier though Sand than Clay can be inferred from evidence that Sand has larger grains and larger spaces between grains than Clay. Making this inference would involve visually depicting the Mechanistic Cues (Sand grains, Clay grains) and animating the Mental Model (water flowing through Sand grains, water flowing through Clay grains).

Study Setup

Participants were 37 student volunteers from two seventh grade science classes at a public middle school in New York City. There were 14 girls and 23 boys. The student body was diverse in ethnicity and economic income level.

All students interacted with the *Flood Predictor* program during science class. Students were randomly assigned to either a Mech group (7 Girls, 11 Boys) or a No Mech group (7 Girls, 12 Boys). The Mech group, in addition, received a supplementary "Field Report" providing Mechanistic Cues about each factor presented in *Flood Predictor.* Mechanistic Cues were intended to elicit imagistic depiction of factors, leading to Qualitative Animations of how and why factors are causal, and advances in reasoning.

Mech group participants were given Mechanistic Cues about each factor, but not given the mechanism that could be inferred from this information. Making the correct inference would require depicting the Mechanistic Cues and animating the Mental Model. For example, a correct mechanism inference that Soil Type causes flooding because water flows easier though Sand than Clay can be inferred from evidence that Sand has larger grains and larger spaces between grains than Clay. Making this inference would require one to visually depict the Mechanistic Cues (Sand grains, Clay grains) and depictively run the Mental Model (water flowing through Sand grains, water flowing through Clay grains).

Learning Activity

Each student worked individually with one of two researchers two at a time over several weeks during science class. The program was set up as one of five lab stations in the classroom. Students used headsets to hear audio guidance.

Preinvestigation System Model

Students wrote answers to the question "Describe how and why do you think flooding happens," which was typed on a blank paper and also read to them. After being told these are some factors researchers suggested might be causal (WP, WT, SD, ST, E), they responded to the question "Describe how and why do you think these factors do or do not cause flooding," which was typed on a blank paper and also read to them. It was emphasized that they should describe their model based upon everything they know.

Preinvestigation Theory Report

Upon entering the program and being presented with problem scenario, students reported whether proposed factors "matter," "don't matter" or "it depends" in causing flooding, and how sure they were about each initial factor theory: "very sure," "sure," or "not sure."

Investigation—Inquiry, Prediction, Evidence, and Inference

Each participant investigated four self-selected flood sites, predicted the flood outcomes for each of the four self-selected sites, and indicated whether factors matter or not in causing flooding. In each inquiry, the participants checked which factor(s) they intended to find out about and called up a record by selecting one of two values for each of the five factors (Water Pollution, Water Temperature, . . .). Inquiries were made in comparisons between new and previous records (after the first inquiry). Relationships were investigated by varying one or more factors between records and comparing flood outcomes. Three controlled experiments investigating individual effects could be made between the four self-selected records. For each record selection, students made a prediction about site flood level, and indicated which factor(s) would be responsible for the predicted flood level.

Students in the "Mech" group, in addition to receiving evidence about covariation in the "change" or "no change" in flood level between sites investigated, received the Field Report. The report, designed to lead to visualization about how and why factors are or are not related to flooding, was presented as observations recorded by researchers in the field.

Participants reported whether they found out if each factor "matters," "doesn't matter," or "hadn't found out." If one claimed to find out a factor "matters" or "doesn't matter" in causing flooding, they indicated how they know what they know, from "the records," "other information," "the records and other information," or "I just know."

Postinvestigation Theory Report

After inquiring about four sites, participants checked whether factors "matter," "do not matter" or "it depends" in causing flooding, which factor values would cause more flooding (i.e., Clay vs. Sand causes more flooding), and how sure they were about each factor theory: "very sure," "sure," or "not sure."

Postinvestigation System Models

Immediately after using the program, students again responded to the question "Describe how and why do you think these factors do or do not cause flooding?" in writing on paper. It was emphasized that they should describe their model based upon everything they know.

Debriefing

During class presentations of lab experiments at the end of the school year, the researcher described the experiment to the students, mentioning that just as they were trying to find out what causes flooding, this researcher was investigating how receiving different types of information would influence the way they investigated the causes of flooding. It was explained that some of them, in addition to finding out site flood levels associated with various sets of environmental factor values, were also given a "Field Report" which suggested how and why factors might cause flooding.

The researcher facilitated a discussion about strategy use in investigating whether factors matter in causing flooding. Some students pointed out they wanted to see how all the factors at once or all factors that mattered created an effect, which is why they varied more than one factor at a time. Only a portion of these students expressed on understanding that one would not know which factor had created the effect. Several students communicated an understanding that in order to investigate an individual effect one has to isolate factors of interest, by "keeping other factors the same," to be sure that only the factor varied could had caused a change. Students discussed the differences between the sample sites in the program world and the larger scale real world, mentioning that in other places they might find certain effects. Students were encouraged to continue to investigate their models of flooding in the real world through their upcoming classroom lab experiments and in their daily lives.

Data Analyses and Results

All but mental model measures were calculated directly from digital data, which reported each action and answer choice response made by each student while investigating in the computer program, or calculated from answer choices in paper-based assessments by the researcher and research assistant. Gender was incorporated into analyses.

The following data was analyzed:

- Complexity in Model Descriptions
- Qualitative Animation in Model Descriptions
- Inquiry
- Prediction Accuracy
- Inferences during investigation
- Post-investigation Theory Accuracy

Initial Flood Models

Models of how flooding happens fell into two distinct groups: clogged toilets and sewage blockage. There were also general categories of models of the relationships between suggested factors and flood level. Causal models of water pollution generally described how the water level would go up if pollution were poured into it. A noncausal model described how pollution would float so it would not make the water rise. A few students offered more than one model by adding a potentially intervening factor. One student described an interactive effect: The hot water would melt the pollution, which would make the water level change. Causal models of water temperature suggested hot water might evaporate. There were many causal models of soil depth that suggested there would be more room for water with deep soil. Some interpreted this as a flood. Causal models of soil type included water rushing through sand at the seashore. Most causal models of elevation described water falling or streaming down. Most students initially referred to evidence from events and places they had seen or heard about, such as a hurricane at their grandfather's house in Puerto Rico, television and newspaper reports, and digging holes at a trip to the beach with their family.

Model Complexity

The research assistant coded responses to questions "How and why you think these factors do or do not cause flooding?" using a coding rubric developed by the researcher, based upon Jonassen's (1995) suggestions for measuring mental models: namely considering coherence, integration, fidelity, imagery, complexity, transferability, and inferential ability in assessing mental models. Two participant Mental Model Complexity descriptions were independently coded by both the researcher and research assistant to check for inter-coder reliability, and were found to be coded exactly the same by both coders.

Mental Model Complexity scores were computed by coding and counting the following elements in Mental Model descriptions Preinvestigation and Postinvestigation: number of Components, number of Non-Qualitative Animations, number of Qualitative Animations, highest number of connected Links between components, number of Interactions, highest number of component States, Fidelity of component sub-models, and number of Images.

The number of components was calculated as the number of elements related to flooding in participant responses to how and why flooding occurs. Non-qualitative animations were the number of reports that one component affects another component without an explanation of how or why it is causal. Qualitative animations were the number of reports of how and why one component affects another component. Highest number of connected links was coded as the greatest number of connected links made between components (i.e., "Hot Water Temperature causes Water Pollution to melt which makes the flood go down" was coded as 2 Links. "Soil Type causes flooding" was coded as one link. If this were the entire response, the highest number of connected Links would be 2). Interaction was a report that the effect of one factor depends upon the value of another factor. Highest number of component states was coded as the greatest number of states described for a factor (i.e., "Clay causes flood to go up" was coded as 1 state, whereas "Soil Type causes flooding" was coded as 0 states.). Fidelity was coded as the level of correctness of causal inference and mechanism (1 point for correct causal inference, a second point for correct explanation of how and why component is causal).

Mental Model Complexity Causal factors and Mental Model Complexity NonCausal factors Preinvestigation and Postinvestigation scores were computed by adding the following: number of Components, Non-Qualitative Animations, Qualitative Animations, highest number of connected Links between components, Interactions, States, Fidelity, and Pictures about Causal factors and about NonCausal factors at preinvestigation and postinvestigation. There is no outer limit to the possible level of the Mental Model Complexity score.

Providing Mechanistic Cues about system components was expected to encourage qualitative mental depiction of system components, leading to development of more complex and more mechanistic mental models. Higher Mental Model Complexity scores in the Mech group than the No Mech group would suggest Mechanistic Cues instigates greater mental model complexity development.

The data in Table III show when providing Mechanistic Cues, participants had greater Complexity in postinvestigation mental model descriptions (33.9) than participants not provided with Mechanistic Cues (27.5). Students in the Mech group had greater complexity in Mental Model descriptions specifically about Causal factors (22.9) than No Mech participants (17.9). Students not provided with Mechanistic Cues reduced the level of complexity in model descriptions from 31 (SD = 7.0) to 27.5 (SD = 6.7). Girls started off with slightly more complex models, 32.8 (SD = 13.1), than boys, 29.7 (SD = 8.5), and decreased the level of complexity in models during investigation from 32.8 (SD = 13.1) to 30.9 (SD = 7.0), whereas boys barely changed the level of complexity in models during investigation, from 29.7 (SD = 8.5) to 30.1 (SD = 12.5).

ANCOVAs with Mech as a factor controlling for preinvestigations showed differences between the Mech group and the No Mech group in level of Mental Model Complexity Total and Causal factors were statistically reliable at the 0.05 level. When provided with Mechanistic Cues, participants had greater Mental Model Complexity scores about NonCausal factors (10.7) than when not provided with Mechanistic Cues (9.4), but the difference was not statistically significant at the alpha 0.1 level. As expected, providing Mechanistic Cues was associated with development of greater mental model complexity, especially about causal factors.

Qualitative Animation of System Mechanisms in Models

Qualitative Animations were calculated as the total number of Qualitative Animations preinvestigation and postinvestigation, descriptions about how and why components are causally related, on preinvestigation and postinvestigation. Typical Qualitative Animations in Mental Model descriptions were “Hot water evaporates and disappears making the water go down” and “Sand has bigger holes for water to go through so water won’t be so high.” Qualitative Animations Causal and Qualitative Animations NonCausal Preinvestigation and Postinvestigation were calculated as the number Qualitative Animations about Causal factors and the number of Qualitative Animations about NonCausal factors at preinvestigation and postinvestigation. There is no outer limit to the possible number of Qualitative Animations in mental model descriptions. More Qualitative Animations in Mech group mental model descriptions than No Mech group mental model descriptions would suggest providing Mechanistic Cues led to more animated mental simulations of the mechanism of a system.

The data in Table IV show when provided with Mechanistic Cues participants had higher levels of animation in Mental Model descriptions (3.6) than when not provided with Mechanistic Cues (1.8). Mech

Table III. Mean Complexity Levels in Post-Inquiry Mental Model Descriptions by Group and Gender

Gender	Group	Mean	*N*	Std. Error
Girls	No-Mech	30.43	7	1.97
	Mech	31.29	7	3.37
Boys	No-Mech	25.75	12	2.04
	Mech	35.64	11	4.58
Total	No-Mech	27.47	19	1.54
	Mech	33.94	18	3.06

Table IV. Qualitative Animations Total in Post-Inquiry Mental Model Descriptions by Group and Gender

Gender	Group	*N*	Mean	Std. Error
Girls	No-Mech	7	1.43	0.61
	Mech	7	3.14	0.59
Boys	No-Mech	12	2.00	0.56
	Mech	11	3.90	0.86
Total	No-Mech	19	1.79	0.42
	Mech	18	3.61	0.57

group participants specifically had more animations about Causal factors (2.8) than No Mech group participants (1.3). Students not provided with Mechanistic Cues reduced the level of animation in model descriptions from 2.9 (SD = 2.1) to 1.8 (SD = 1.8). Girls started off with more Qualitative Animations (3.64) than boys (2.61) and decreased the level of Qualitative Animation in models from 3.64 (SD = 4.27) to 2.29 (SD = 1.17), whereas boys increased the level of Qualitative Animation in models during investigation from 2.61 (SD = 2.06) to 2.91 (SD = 2.56).

ANCOVAs with Mech as a factor, controlling for preinvestigations, showed differences between the Mech group and the No Mech group in Qualitative Animations Total and Qualitative Animations about Causal factors were statistically reliable at the alpha 0.05 level. When provided with Mechanistic Cues, participants had more Qualitative Animations about NonCausal factors (0.8) than when not provided with mechanistic cues (0.5), but the difference was not statistically reliable at the 0.05 level. As expected, providing Mechanistic Cues in images and text was associated with more qualitatively mechanistic models, especially about causal factors.

Table V. Controlled Comparisons Frequencies and Means by Group and Gender

	Mechanistic Cues	
	No Mech	Mech
Gender		
All		
0 Controlled Comparisons	15	11
1 Controlled Comparisons	2	5
2 Controlled Comparisons	2	2
Total Controlled Comparisons	6	9
n	19	18
Mean	0.32	0.50
Std. Error	0.15	0.17
Girls		
0 Controlled Comparisons	7	5
1 Controlled Comparisons	0	1
2 Controlled Comparisons	0	1
Total Controlled Comparisons	0	3
n	7	7
Mean	0.00	0.43
Std. Error	0.00	0.30
Boys		
0 Controlled Comparisons	8	6
1 Controlled Comparisons	2	4
2 Controlled Comparisons	2	1
Total Controlled Comparisons	6	6
n	12	11
Mean	0.50	0.54
Std. Error	0.23	0.21

Inquiry

Inquiry was calculated as the number of times students made a controlled comparison, holding all but one factor of interest constant, out of a total of three across the four records. The data in Table V show when providing Mechanistic Cues participants made more Controlled Comparisons in inquiry (0.5) than when not providing Mechanistic Cues (0.3). There were six controlled comparisons made in the No Mech group and nine in the Mech group. Seven out of 18 Mech group participants made at least one Controlled Comparison, whereas only 4 out of 19 participants in the No Mech group made at least one Controlled Comparison. ANOVA with Mech as a factor showed the difference between Mech and No Mech groups was not statistically significant at alpha 0.05.

Prediction

Total Misprediction was calculated as the total level of discrepancy between the site flood level predicted by the participant and the true site flood level in the program, across four record instances. The data in Table VI show providing mechanistic cues did not lead to less Misprediction when analyzing the total group (5.3, out of the highest possible level of 16) than not providing Mechanistic Cues (5.2). However, when considering gender as a factor, an interactive effect was revealed. Providing mechanistic cues led to more accurate predictions for girls, but mechanistic cues led to less accurate prediction for boys. ANOVA with Mech and Gender as factors show differences were not statistically significant at the alpha 0.05.

Table VI. Mean Misprediction by Group and Gender

Gender	Group	*N*	Mean	Std. Error
Girls	No-Mech	7	5.71	1.23
	Mech	7	4.86	0.94
Boys	No-Mech	12	4.92	0.69
	Mech	11	5.54	0.62
Total	No-Mech	19	5.21	0.61
	Mech	18	5.28	0.52

Table VII. Mean Correct Inferences Total During Inquiry by Group and Gender

Gender	Group	*N*	Mean	Std. Error
Girls	No-Mech	7	8.29	1.27
	Mech	7	12.86	1.33
Boys	No-Mech	12	8.50	1.01
	Mech	11	10.82	1.15
Total	No-Mech	19	8.42	0.77
	Mech	18	11.61	0.88

Table VIII. Post-inquiry Theory Accuracy Total Means by Group and Gender

Gender	Group	Mean	*N*	Std. Error
Girls	No-Mech	1.29	7	0.52
	Mech	2.57	7	0.61
Boys	No-Mech	2.25	12	0.27
	Mech	2.64	11	0.43
Total	No-Mech	1.89	19	0.27
	Mech	2.61	18	0.34

Inference

Correct Inferences was calculated as the total number of correct inferences about causality during inquiry activity, out of a total possible score of 20 correct inferences. The data in Table VII show when provided Mechanistic Cues, participants made more Correct Inferences (11.6) than when not provided Mechanistic Cues (8.4). This was true for both boys and girls, though the difference between Mech and No Mech students in correct inferences was greater for girls. An ANOVA, with Mech (No Mech group, Mech group) as a factor, on Correct Inferences, revealed the difference between Mech and No Mech groups in the number of Correct Inferences was statistically significant at the alpha 0.05 level. Providing Mechanistic Cues led to more correct inferences.

Theory Accuracy

Theory Accuracy was assessed before and after inquiry. Level of Theory Accuracy Total Preinvestigation and Postinvestigation were calculated as the number of Accurate Theories about causality on preinvestigation and postinvestigation. An Accurate Theory was recognized as having correctly indicated what factor settings are causal and the direction of causality. A total of five Accurate Theories, three Correct Causal Theories, and two Correct NonCausal Theories are possible. A score of 5 indicates all factor theories about causality in the *Flood* system were correct. Correct Causal factor theories and Correct NonCausal factor theories were calculated as the number of Accurate Theories about Causal factors and the number of Accurate Theories about NonCausal factors at preinvestigation and postinvestigation.

The data in Table VIII show when providing Mechanistic Cues, participants had more Accurate Theories Total (2.6) than when not providing Mechanistic Cues (1.9). Specifically, Mech participants make more Accurate Theories about NonCausal factors (1.0) than No Mech participants (0.5), and slightly more Accurate Theories about Causal factors (1.6) than No Mech participants (1.4). ANCOVAs with Mech as a factor controlling for preinvestigations showed differences between the Mech group and the No Mech group in Accurate Theories Total (at the alpha 0.095 level), and specifically about NonCausal factors (at the alpha 0.05 level), were statistically significant. Providing Mechanistic Cues led to better knowledge acquisition.

Regression analyses showed postinvestigation Mental Model Complexity about Causal factors was a marginally significant predictor of the number of Accurate Theories about Causal factors (at the alpha 0.08 level), controlling for Accurate Theories preinvestigation, Mental Model Complexity preinvestigation, and the number of Controlled Comparisons. Postinvestigation Mental Model Complexity about Total factors and NonCausal factors were not significant predictors of Accurate Theories Total or about NonCausal factors, controlling for Mental Model preinvestigation (Total factors, NonCausal factors), Theory preinvestigation (Total factors, NonCausal factors), and the number of Controlled Comparisons. Development of mental model complexity was associated with better knowledge acquisition about causal factors, but the association was not as significant as expected.

Regression analyses show postinvestigation Qualitative Animations in mental model descriptions about Causal factors were a significant predictor (at the alpha 0.05 level) of the number of Accurate Theories about Causal factors. Postinvestigation Qualitative Animations Total factors and NonCausal factors were not significant predictors of Accurate Theories Total or Accurate Theories about NonCausal factors, controlling for Qualitative Animations preinvestigation (Total factors, NonCausal factors), Theory preinvestigation (Total factors, NonCausal factors) and the number of Controlled Comparisons.

Increases in Qualitative Animations, explanations about how and why factors matter, were associated with better knowledge acquisition about causal factors.

Discussion and Implications

Mechanistic cues in inquiry led to more complex and animated mental reprepresentations, with more illustrations of how and why factors are or are not causally related. Providing Mechanistic cues resulted in higher levels of complexity in post-inquiry model descriptions than not providing cues. Mechanism cues incited enhanced depiction of system components, allowing for more animated mental simulation.

Providing mechanism cues led to more correct inferences during inquiry and greater knowledge acquisition, especially in girls. Mech group participants made more correct inferences and had more accurate postinvestigation theories. Level of complexity and animation in mental model descriptions were significant predictors of formulating more accurate theories.

Providing mechanism cues in computer-based inquiry led to greater use of the controlled comparison inquiry strategy and greater prediction accuracy in girls than not providing cues. The mental model presents hypotheses for focussed inquiry. Having a more complex and animated mental model perhaps allowed for more focussed initial inquiries because fewer exploratory tests of effects were needed before determining hypotheses. Another explanation is that thinking about the how and why of causality is taxing on working memory, and encourages variation of only one component state at a time.

A mental model accurately depicting components and implicit causal attributes allows for generation of correct mechanisms, thus leading to correct inferences and more accurate prediction. Mental models depicting an accurate mechanism explain covariation, perhaps leading to more correct theoretical conclusions. Mental animation allows for prediction in the mental model before prediction in the external system, perhaps explaining greater accuracy in program predictions.

Researchers and educators should consider how information format in instructional and assessment materials influences student representation and reasoning performance. Helping students mentally depict systems under investigation led to better inquiry and learning in girls. Static imagery with qualitative motion cues stimulated visualization and running of mental models. Both boys and girls provided with mechanistic cues developed more complex and animated representations than students not provided with cues. Enhanced representation was especially advantageous for reasoning in girls. Girls improved inquiry strategies, predictions and knowledge acquisition as a result of complexity and animation stimulated by mechanism cues.

REFERENCES

Byrne, R. M. (1992). The model theory of deduction. In Rogers, Y., Rutherford, A., and Bibby, P. (Eds.), *Models in the Mind: Theory, Perspective, and Application*, Academic Press, San Diego, CA.

Craik, K. (1943). *The Nature of Explanation.* Cambridge University Press, Cambridge.

de Kleer, J., and Brown, J. S. (1983). Assumptions and ambiguities in mechanistic mental models. In Gentner, D., and Stevens, A. L. (Eds.), *Mental Models*, Erlbaum, Hillsdale, NJ.

Gentner, D., and Stevens, A. L. (Eds.) (1983). *Mental Models.* Erlbaum, Hillsdale, NJ.

Goldman, S. R., Petrosino, A. J., Sherwood, R. D., Garrison, S., Hickey, D., Bransford, J. D., and Pellegrino, J. W. (1996). Anchoring science instruction in multimedia learning environments. In Vosniadou, S., De Corte, E., Glaser, R., and Mandl, H. *International Perspectives on the Design of Technology-Supported Learning Environments,* Erlbaum, Mahwah, NJ.

Inhelder, B., and Piaget J. (1958). *The Growth of Logical Thinking From Childhood to Adolescence.* Basic Books. (Original work published 1955), New York.

Johnson-Laird, P. N. (1983). *Mental Models: Toward a Cognitive Science of Language, Inference, and Consciousness.* Harvard University Press, Cambridge, MA.

Johnson-Laird, P. N. (1998). Imagery, visualization, and thinking. In Hochberg, J. (Ed.), *Perception and Cognition at Century's End.* Academic Press, San Diego, CA.

Johnson-Laird, P. N., and Byrne, R. M. (1991). *Deduction.* Erlbaum, Hillsdale, NJ.

Jonassen, D. H. (1995, October). *Operationalizing Mental Models: Strategies for Assessing Mental Models to Support Meaningful Learning and Design Supportive Learning Environments*, Computer Support for Collaborative Learning 95, Indiana University Bloomington, IN.

Hochberg, J. (Ed.) (1998). *Perception and Cognition at Century's End*, Academic Press, San Diego, CA.

Kuhn, D. (1989). Children and adults as intuitive scientists. *Psychological Review* 96: 674–689.

Kuhn, D., Schauble, L., and Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction* 9(4): 285–327.

Kuhn, D., Black, J., Keselman, A., and Kaplan, D. (2000). The development of cognitive shills to support inquiry learning. *Cognition and Instruction* 18: 495–523.

Piaget, J. (1972). *The Child's Conception of Physical Causality*. Adams, and Totawa, NJ. First published in English by Routledge and Kegan Paul, London, 1930.

Rogers, Y., Rutherford, A., and P. Bibby (Eds.), *Models in the Mind: Theory, Perspective, and Application.* Academic Press, San Diego, CA.

Rumelhart, D., and Norman, D. (1975). The active structural network. In Norman, D. A., Rumelhart D. E., and the LNR Research Group. *Explorations in Cognition*, Freeman, San Francisco, CA.

Schank, R. C., Fano, A., Bell, B., and Jona, M. (1994). The design of goal-based scenarios. *The Journal of the Learning Sciences.*

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology* 49: 31–57.

Schwartz, D. L., and Black, J. B. (1996a). Analog imagery in mental model reasoning: Depictive models. *Cognitive Psychology* 30: 154–219.

Schwartz, D. L., and Black, J. B. (1996b). Shuttling between depictive models and abstract rules. *Cognitive Science* 20: 457–497.

Wason, P. C., and Johnson-Laird, P. N. (1972). *Psychology of Reasoning: Structure and Content.* Harvard University Press, Cambridge, MA.

Williams, M. D., Hollan, J. D., and Stevens, A. L. (1983). Human reasoning about a simple physical system. In Gentner, D., and Stevens, A. L. (Eds.), *Mental Models*, Erlbaum, Hillsdale, NJ.

Zukav, G. (1979). *The Dancing Wu Li Masters: An Overview of the New Physics,* Morrow, New York.