

Model-Based Teaching and Learning With BioLogica™: What Do They Learn? How Do They Learn? How Do We Know?

**Barbara C. Buckley,^{1,5} Janice D. Gobert,¹ Ann C. H. Kindfield,² Paul Horwitz,¹
Robert F. Tinker,¹ Bobbi Gerlits,¹ Uri Wilensky,³ Chris Dede,⁴ and John Willett⁴**

This paper describes part of a project called Modeling Across the Curriculum which is a large-scale research study in 15 schools across the United States. The specific data presented and discussed here in this paper is based on BioLogica, a hypermodel, interactive environment for learning genetics, which was implemented in multiple classes in eight high schools. BioLogica activities, data logging, and assessments were refined across this series of implementations. All students took a genetics content knowledge pre- and posttests. Traces of students' actions and responses to computer-based tasks were electronically collected (via a "log file" function) and systematically analyzed. An intensive 3-day field test involving 24 middle school students served to refine methods and create narrative profiles of students' learning experiences, outcomes, and interactions with BioLogica. We report on one high school implementation and the field test as self-contained studies to document the changes and the outcomes at different phases of development. A discussion of design changes concludes this paper.

KEY WORDS: genetics; model-based learning; interactive environments; data logging; technology-enhanced assessment.

With support from NSF, the Concord Consortium developed an interactive, computer-based learning environment, BioLogica, that is designed to support students in high school classrooms to build a deep understanding of core concepts in Mendelian genetics. The pedagogical challenges are numerous. What do we mean by deep understanding? How can we help them develop deep understanding? How do we know when they've done so? This paper focuses on the learning that takes place when students use BioLogica, an interactive genetics curriculum, in their high

school classrooms. It presents the model of learning we use, what this looks like in practice, how we determine the nature and extent of student learning when BioLogica is used in high school classrooms, and what we've learned about all of the above.

MODEL-BASED LEARNING

The theory we employ is an elaboration and extension of Model-Based Teaching and Learning (MBTL) set forth in a special issue of the International Journal of Science Education (Gobert and Buckley, 2000). The tenets of model-based learning are based on the presupposition that understanding requires the construction of mental models of the phenomena under study, and that all subsequent problem-solving, inferencing, or reasoning are done by means of manipulating or "running" these mental models (Johnson-Laird, 1983). We view mental models as internal, cognitive representations used in

¹The Concord Consortium, 10 Concord Crossing, Suite 300, Concord, Massachusetts 01742.

²Educational Designs Unlimited, Inc., Hillsborough, New Jersey.

³School of Education and Social Policy, Northwestern University, Chicago, Illinois.

⁴Graduate School of Education, Harvard University, Cambridge, Massachusetts.

⁵To whom correspondence should be addressed; e-mail: bbuckley@concord.org

reasoning of many kinds (Brewer, 1987; Rouse and Morris, 1986). Mental models, like prior knowledge, influence our perceptions of phenomena and our understanding of information. Interactions with phenomena and representations, in turn, influence our mental models (Gentner and Stevens, 1983; Johnson-Laird, 1983).

Before proceeding, it is important to define the different types of models we use in addition to the notion of models as mental representations of phenomena. Our starting point is Norman's (1983) differentiation of models related to a target system or phenomenon that is to be represented or modeled. Norman distinguishes not only the learners' mental models but also the scientist's and designer's conceptual models of the system, as well as the researchers' conceptualization of the learner's mental models.

The hypermodels used in this project are another type of model that we add to the framework of models described by Norman (1983). Hypermodels are conceptual models realized as computer models embedded in interactive curricula. These are external representations (as opposed to mental models, which are internal representations) with which the learner interacts, and in doing so, constructs and/or revises his/her mental model. Representations are considered models only when they represent structural, dynamic, and/or causal aspects of the target model; that is, they are not just visualizations or diagrams of phenomena.

With these definitions in mind, we define model-based learning as a dynamic, recursive process of learning by constructing mental models of the phe-

nomenon under study. It involves the formation, testing, and subsequent reinforcement, revision, or rejection of mental models. This is analogous to hypothesis development and testing seen among scientists (Clement, 1989). See Fig. 1.

In response to the demands of a task, a learner draws on prior knowledge and new information during model-formation to construct a mental model of some phenomenon (Buckley, 1992, 2000; Kozma *et al.*, 1992). The learner's prior knowledge may include a partial mental model of the phenomenon or naive models that are incompatible with the scientifically accepted model. If the learner's model is used successfully to accomplish the task, the model is reinforced (Clement, 1989) and may eventually become a precompiled, stable model (Vosniadou and Brewer, 1992) from which students are capable of making inferences, etc. However, if inconsistencies and/or deficiencies in the model are noted (Bransford *et al.*, 1986), the learner may reject the model and form a new one, or revise the initial model (Chinn and Brewer, 1993; Clement, 1989; Schauble *et al.*, 1991). Model revision may involve modifying parts of the existing model or elaborating the model by adding to or combining existing models. Embedding a model in a larger model is an example of elaboration (deKleer and Brown, 1983; Monaghan and Clement, 1994; Stewart and Hafner, 1991).

Metacognitive processes such as selecting which perceptual cues to attend to, directing interactions with representations and phenomena during learning and inquiry, and monitoring the results of those

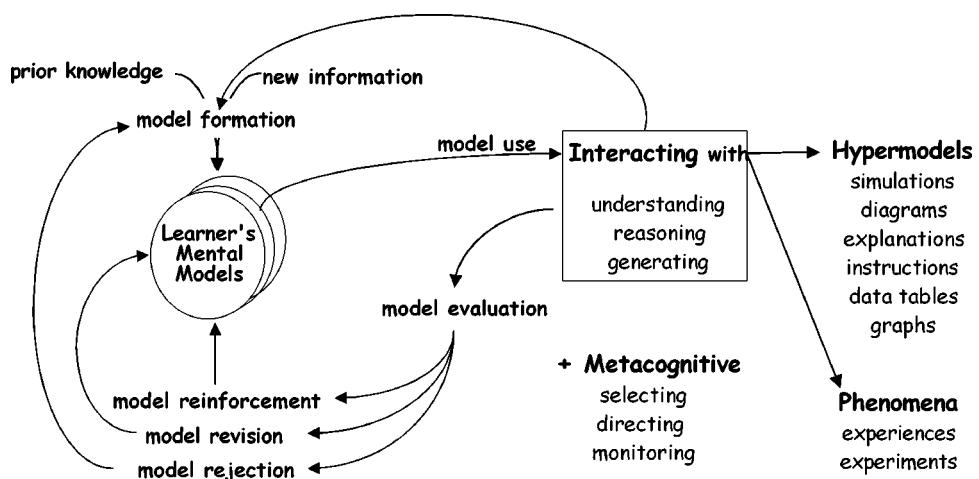


Fig. 1. Model-based learning framework.

interactions and the evaluation of one's model also play an important part in model-based learning as do reasoning processes such as hypothesis generation from the model, testing that hypothesis, and interpreting the data that are among higher order inquiry skills.

Thus, the mental model evolves through multiple recursions as it is made increasingly complex and, hopefully, more accurate (Johnson-Laird, 1983; White and Frederiksen, 1998).

MODEL-BASED TEACHING WITH THE BioLogica HYPERMODEL

Hypermodels are software environments that allow learners to interact with a manipulable model of some phenomenon in a domain (Horwitz, 1995; Horwitz and Christie, 1999). Hypermodels are controlled by short programs, or activity scripts, that mediate a learner's interactions with the model.

The core of the BioLogica hypermodel (see Fig. 2) is a model of transmission genetics that represents the genetic mechanisms of inheritance as the parts and processes that take place at multiple levels of biological organization. It combines the Mendelian model of genetics and the molecular and cellular mechanisms that produce the phenomena Mendel observed. The learner interacts with the model through manipulable representations appropriate to the task posed by the activity script. Because the manipulable representations operate on the core model, changing any one of them can affect each of the others, as appropriate. For example, altering an organism's genotype affects not only that organism, but also any of its offspring that happen to inherit the altered gene. The

activity scripts are used to orchestrate instructional activities built around a series of challenges.

Research has shown that students have difficulty interpreting and reasoning with external models and representations that scaffolding is needed to support students' learning (e.g., Gobert and Buckley, 2003; Gobert and Clement, 1999; Kindfield, 1993/1994; Larkin, 1989; Larkin and Simon, 1987; Lowe, 1993). The pedagogical elements outlined in Table I scaffold the learner's interaction with the challenge, the manipulable representations, and ultimately with the model of transmission genetics. Each of the elements serves both pedagogical purposes and research purposes. For example, embedded questions are intended to facilitate a deeper engagement on the part of the learner by posing questions that require more attention to the details of the activity. They are also intended to elicit responses from which we can infer the state of the learner's mental models. Narratives and instructions model the kind of reasoning a geneticist might pursue in solving a challenge. Activity scripts also monitor a learner's actions so that constructive feedback can be given as appropriate. Explanations and solutions are provided only after the learner has had an opportunity to grapple with the challenge. Each activity also creates a log file that records students' actions and responses and can be processed to yield information useful to teachers, students and researchers.

RESEARCH DESIGN AND METHODS

Our research is a large-scale design study conducted in demographically diverse classrooms to

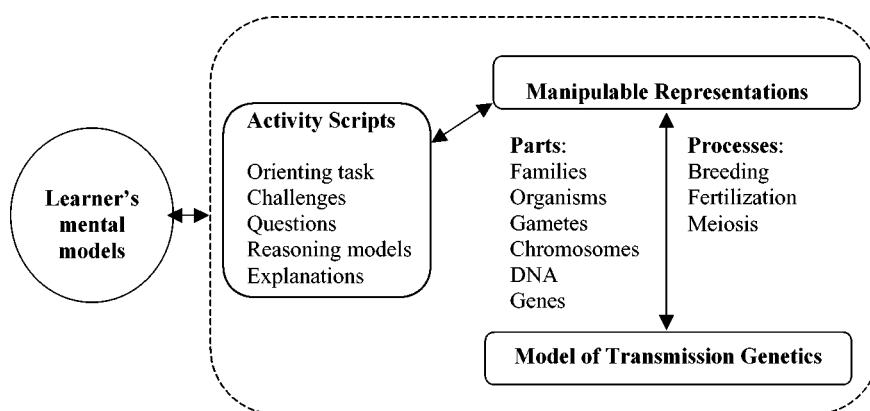


Fig. 2. BioLogica hypermodel.

Table I. Pedagogical Elements in Activity Scripts

Element	Description	Location
Advance organizer or orienting task	Briefly describes expected performance in terms of focus question, content and actions	Headlines in login screen
Narratives	The narrative into which puzzles are woven	Elaborated in introduction screen Begun in introduction and woven throughout the activity
Tasks and puzzles	Require cause-to-effect and effect-to-cause reasoning, within and between generations	The heart of all activities
Representational assistance	Links diagrams with vocabulary, highlights structures involved in genetics	When new representation or tool is introduced
Reasoning models	Walks learner through reasoning strategies	After initial exploration of concept and representations
Explanations	Summarizes concepts and models good explanations	After completion of task/puzzle
Embedded questions	About representations, models, and reasoning About the pieces of the model central to the activity	Throughout activity
Feedback on actions and responses	Monitors responses and problem solving actions	More frequent initially then fades away before assessment questions
Parallel phenomena exposure	Scenarios and questions about genetics in other species and in humans	After in-depth exploration and experimentation with dragon genetics
Assessment questions	General concepts Activity specific	End of activity
Reflective questions	What they knew What they learned What they wonder about now	End of activity

inform the development of BioLogica. The activities, data logging, and assessments were refined across a series of implementations. Data were collected in multiple classes from one middle school and eight high schools. An intensive 3-day field test involving 24 middle school students served to refine methods and create narrative profiles of students' learning experiences, outcomes, and interactions with BioLogica. From this massive amount of data, we will focus on two studies conducted at various times in the development of BioLogica to examine both development issues and research findings. Since BioLogica activities, the instruments used to assess learning, and data logging capabilities changed over the course of the year, we report on one high school implementation and the field test as self-contained studies. To provide a frame for these studies we begin by describing the development of activities, assessments, and data logging.

Development of Activities

Prior research with GenScope (Hickey *et al.*, 1998b, 1999, 2003; Hickey and Kindfield, 1999; Horwitz *et al.*, 1998; Kindfield and Hickey, 1999) guided the initial development of BioLogica activities, which are built around a series of puzzles or

challenges. BioLogica activities began to be used in classrooms in November 2000. Revision of these activities was informed by classroom observations in three schools, prior research by the authors (Buckley, 2000; Buckley and Boulter, 2000; Gobert, 2000), and by the model-based teaching and learning framework described earlier. During the first phase of testing (11/00 through 2/01) BioLogica was used in three high schools, the number of activities was increased from 8 to 12, the first data logging was implemented, and a Teacher's Guide was created. See Appendix A for a complete description of BioLogica activities.

Development of Pre- and Posttests

Like the activities, the pre- and posttests evolved over the academic year 2000–2001. In the initial implementation of BioLogica at site "MR" the pre- and posttests were identical. The instrument consisted of 70 NewWorm items originally developed for the GenScope project (Hickey *et al.*, 1998a). Identical pre- and posttests were administered at site "F," using a new, shorter, version of NewWorm from Hickey and Kindfield with the addition of four multiple choice items consisting of published items from MCAS and NY Regents exams. We learned that

most students and many teachers considered even the 35-item test too long. In subsequent implementations items were deleted from the test if the item tested a concept (such as crossing-over) that was not addressed in any of the BioLogica activities used in that implementation. The pretest items were a subset of the posttest items. Additional multiple choice items were included to better assess how students performed on standardized test items and a few multiple choice questions were included that addressed common naive conceptions. Because the NewWorm test assumes an understanding of genetics terminology, but doesn't test for that understanding independent of the problem solving required for most NewWorm items, we introduced open-ended conceptual questions beginning with implementations in April 2001.

We examined all items on the pre- and posttests in three ways: (1) using the matrix of reasoning patterns identified by Hickey, Wolfe, and Kindfield (1998b), (2) using a model-based learning perspective (Buckley and Boulter, 1999, 2000), and (3) identifying the genetics concept tested. The reasoning matrix has two dimensions: between versus within generations, and cause-to-effect versus effect-to-cause. Hickey *et al.* (1998b) demonstrated that questions involving cause-to-effect were easier for students than effect-to-cause and that reasoning within a generation was easier than reasoning across generations. Drawing on the model-based learning framework, we analyzed items using a method developed for analyzing representations (Buckley and Boulter, 1999, 2000). This analysis focused primarily on the models of parts (structures) and processes needed to answer the question.

Development of Data Logging

First, what do we mean by data logging? When students use BioLogica activities, the activity scripts that guide and control students' interactions with the underlying genetic models generate log files. The specific entries in a given log file are generated either automatically or by design. The automatic entries are used to capture start and stop times for an activity or a given challenge within the activity and to capture learners' responses to questions posed by the activity script. The designed entries can be tailored to answer specific research questions and provide assessment data for teachers, students, and researchers. Embedded questions were added to activities implemented

in April 2001 and later. In addition, most activity scripts capture learners' actions, such as how many crosses were made or what alleles were changed, as they explore and experiment with the representations of the underlying genetic model. Thus, the content of log files varies not only by activity but also by user.

The use of data logging increased throughout the year. During the first use of BioLogica in November 2000 only 5 activities generated log files. Of the 54 logs generated by 13 students only 57% were usable; that is, the other 43% were empty, lacking even a time stamp. This is caused either by the user quitting the activity immediately after launch or by a program crash. By May 2001 all 13 activities generated log files eventually achieving a yield rate of 94%.

IMPLEMENTATION STUDIES

At "L" high school, three teachers taught both control classes and experimental classes and three additional teachers used BioLogica with a range of biology classes. These implementations are reported in Study 1. We used the data collected during the intensive field test conducted at site "MV" to develop methods for analyzing log files. That work is reported in Study 2.

Study 1: "L" High School

During this implementation in May 2001, four teachers used BioLogica daily for 1 week with a range of classes. In addition, one of the teachers taught two control classes of ninth grade students matched with her two experimental classes.

BioLogica activities were more stable at this point in development, but still not totally bug free. There were two changes in activities compared to previous implementations: *Sex-Linkage* and *Sex-Linkage II* were combined and revised into one activity and *Invisible Dragons*, a game that requires reasoning from effect-to-cause between generations, was added. The data logging generated accurate start and end times. Although we weren't able to match log files to individual students or classes, we were able to average across all the logs for this implementation. Table II summarizes the logs collected at "L" high school for all classes. We could not determine with precision what proportion of students completed each activity.

Table II. Logs Collected at “L” High School

Implementation site	L 5/01
Activities logged (<i>N</i>)	13
Students (<i>N</i>)	120
Logs generated (<i>N</i>)	2128
Valid logs (<i>N</i>)	1566
Yield (%)	74
Introduction logs (<i>N</i>)	273
Rules logs (<i>N</i>)	266
Meiosis logs (<i>N</i>)	283
Inheritance logs (<i>N</i>)	130
Monohybrid logs (<i>N</i>)	204
Mutations logs (<i>N</i>)	98
Mutations 2 logs (<i>N</i>)	36
Horns dilemma logs (<i>N</i>)	32
Dihybrid logs (<i>N</i>)	26
Sex-linkage logs (<i>N</i>)	52
Scales logs (<i>N</i>)	35
Plates logs (<i>N</i>)	25
Invisible dragons logs (<i>N</i>)	106

However, we can estimate which activities were completed and which were not completed by most of the students. We estimate that the first five activities were completed by all students. Similarly, it is likely that most of the students did not complete activities after Monohybrid. Even Invisible Dragons with 106 logs may be the result of a small number of students using the activity multiple times. See Study 2 for more detailed information on processing and analyzing log files.

We used this information in conjunction with item analysis of the posttest to identify which items on the posttest to exclude from analysis. We examined the mean score on each item across the entire data set from “L.” We looked at each item that had a mean of less than 50% of the possible points and considered whether it tested a concept that most of the students had explored with BioLogica. If it tested understanding of a concept that was included in the first five activities, we did not exclude the item. Conversely, if the items with low mean scores were not

covered in the first five activities we excluded the item from further analysis. By mistake, an older version of the pretest was administered that included only 23 NewWorm items. The posttest consisted of 32 NewWorm items, 13 multiple choice items that covered misconceptions and standardized test items, and 17 open-end concept questions for a total of 106 points. We excluded 19 items, which resulted in a possible maximum score of 62 points. Therefore, the analysis reported in this study is based on what we term “relevant” test items. We report first the study of the ninth grade classes, then the school wide results.

Study 1A: Ninth Grade Control and Experimental Classes

The two control classes received the teacher’s usual introduction to genetics during the week that the two experimental classes used BioLogica. One of the experimental classes received an introductory lesson prior to using BioLogica but the other did not.

Analysis of Pretest. A *t* test was computed in order to test for pretest differences among the four classes at “L” high school. There was a statistically significant difference found between the four groups on the pretest (see Table III). Post hoc contrasts yielded that the differences found were between the “red” group and the other three groups, “yellow,” “green,” and “blue.”

Analysis of Posttest. We pooled the two control groups and the two experimental groups since it was the same teacher and the same level of students in all four groups. The pretest score was used as a covariate in order to take into account the pretest differences. An Analysis of Variance (ANOVA) was computed in order to determine whether there were any statistically significant differences on the total posttest score between the experimental and

Table III. Comparison of Pretest Means for Ninth Grade Classes

Class	Group	N	Pretest			<i>F</i> statistic	Significance (<i>p</i> value)
			Total number of points	Mean	SD		
Blue	Control	16	40	12.3	2.9		
Yellow	Control	20	40	13.1	3.5	4.14	0.009 ^a
Green	Experimental	22	40	13.6	3.7		
Red	Experimental	14	40	9.0	6.1		

^aSignificant at *p* < 0.05 α .

Table IV. Comparison of Control and Experimental Groups on Posttest Mean With Pretest Covariate

Group	N	Pretest		Posttest			F statistic	Significance (p value)
		Total number of points	Mean	Total number of points	Mean	SD		
Control	36	40	12.7	62	25.7	11.0	6.269	0.003 ^a
Experimental	36	40	11.8	62	31.1	10.6		

^aSignificant at $p < 0.05\alpha$.

control groups. For this ANOVA the dependent variable was total posttest score, the independent variable was group (control vs. experimental), and the pretest score was used as a covariate. There was a statistically significant difference found between the control and experimental groups on the posttest with the experimental group scoring higher. See Table IV.

Analysis of Selected Posttest Items. ANOVAs were computed in a series of analyses in which we selected items from the entire posttest and pooled them according to the concept that they were assessing. The pooled items assessed students' understanding of: inheritance, dominance, sex linkage, genotype and phenotype, monohybrid, dihybrid, and pedigree. For each of these analyses, the pretest scores were used as a covariate to allow us a more accurate measure of posttest differences between the control and experimental groups.

There was a statistically significant difference found between the control group and the experimental group on two concepts: monohybrid and dihybrid. For both concepts the experimental group scored higher than the control group. A summary of results on monohybrid and dihybrid can be found Table V. No statistically significant difference was found on inheritance, dominance, sex linkage, genotype and phenotype, or pedigree. See Appendix B for a summary table of this data.

ANOVAs were computed in a series of analyses in which we selected items from the posttest and pooled them according to the type of reason-

ing that they were assessing. The pooled items assessed students on seven different types of reasoning: cause-to-effect within generations (cew), effect-to-cause within generations (ecw), cause-to-effect between generations (ceb), effect-to-cause between generations (ecb), Punnett square, structure, and process. For each of these analyses, the pretest scores were used as a covariate to allow us a more accurate measure of posttest differences between the control and experimental groups.

There was a statistically significant difference found between the control group and the experimental group on six of the seven types of reasoning: cause-to-effect within generations (cew), cause-to-effect between generations (ceb), effect-to-cause between generations (ecb), Punnett squares, structure, and process. The experimental group means were higher than the control group means on the following types of reasoning: cause-to-effect within generations (cew), cause-to-effect between generations (ceb), Punnett squares, structure, and process. The control group means were higher than the experimental group means on effect-to-cause between generations (ecb). It should be noted, however, that the Levene's Test of Equality of Error Variances (which tests whether the error variances of the two groups are equal, was significant for cause-to-effect within generations, and effect-to-cause between generations. Thus, we must interpret the results regarding the two groups' performance on these scores with caution. These results, therefore, should be interpreted with caution. These results are summarized in Table VI. No statistically significant difference was found at $p = 0.05\alpha$ on effect-to-cause within generations. See Appendix B for a summary table of the data for "L" High School.

Table V. Concepts With Statistically Significant Differences Between the Control and Experimental Groups

Dependent variable	Mean	F statistic	Significance (p value)
Monhybrid		5.798	0.005 ^a
Control	7.444		
Experimental	9.167		
Dihybrid		4.261	0.018 ^a
Control	1.250		
Experimental	1.694		

^aSignificant at $p < 0.05$.

Study 1B: Analysis of All Class Levels at "L"

Although the ninth grade was the only class level with controls, three other teachers used BioLogica with their biology students. Analysis of the pre- and

Table VI. Types of Reasoning With Statistically Significant Differences Between Control and Experimental Groups

Dependent variable	Mean	Levene's test	F statistic	Significance (<i>p</i> value)
Cause-to-effect within generations (cew)		0.019 ^a	5.734	0.005 ^a
Control	4.472			
Experimental	5.083			
Cause-to-effect between generations (ceb)		0.904	5.951	0.004 ^a
Control	5.958			
Experimental	7.056			
Effect-to-cause between generations (ecb)		0.011 ^a	3.572	0.033 ^a
Control	5.792			
Experimental	5.694			
Punnett square		0.987	4.280	0.018 ^a
Control	5.250			
Experimental	5.653			
Structure		0.192	3.985	0.023 ^a
Control	2.917			
Experimental	4.139			
Process		0.059	7.420	0.001 ^a
Control	0.667			
Experimental	1.194			

^aSignificant at *p* < 0.05.

posttest scores for those classes are compared in Table VII.

Post hoc contrasts (see Table VIII) reveal that the honors biology class significantly outperformed the other classes on the posttest and that the business biology class significantly underperformed the other classes. It is interesting that the ninth graders did as well as the college prep biology.

Study 2: Intensive 3-Day Field Test at “MV”

The primary purpose of the intensive field test was to find bugs in the software. A secondary purpose was to investigate the usefulness of log files for helping us understand what and how students learn when using BioLogica. Log files were intended from the start as a research tool to help us assess (a) what learners understand, (b) how that understanding changes, and (c) how they use BioLogica. We’re trying to develop a

fine-grained understanding of how different students learn with BioLogica. However, we also want to identify variables that can be quantified and used as covariates in analyzing pre- and posttest gains. In this study we looked for ways to design and process log files so that they provide the data needed in a format that can be analyzed with less effort.

The intensive field test was chosen for this study because it provided a small but diverse population that completed nearly all of the BioLogica activities under relatively controlled circumstances. We reasoned that because it was not a classroom situation, learning gains could be attributed more fully to BioLogica use. Therefore, connections between log files and learning should be more direct and less influenced by other classroom variables.

In the intensive field test conducted at site “MV” in April 2001, 24 middle school and high school students were paid to use all the BioLogica activities over a 3-day period and to take the pre- and posttests and

Table VII. Pre- and Posttest Scores for Different Class Levels

Type of class	N	Pretest		Posttest			F statistic	Significance (<i>p</i> value)
		Total number of points	Mean	Total number of points	Mean	SD		
Ninth grade	36	40	11.8	62	31.1	10.6	26.941	0.000 ^{a,b}
Business biology	6	40	8.8	62	17.1	5.5		
College prep biology	49	40	13.1	62	31.3	12.6		
Honors biology	30	40	21.6	62	47.3	9.4		

^aSignificant at *p* < 0.05^a.

^bLevene’s Test for Equality of Error Variances is significant at *p* < 0.05^a.

Table VIII. Tukey HSD Post Hoc Contrasts Across Class Types

Type of class (mean)	Type of class	Mean	SD	Significance (<i>p</i> value)
Ninth grade (31.1)	Business biology	17.1	5.5	0.025 ^a
	College prep biology	31.3	12.6	1.000
	Honors biology	47.3	9.4	0.000 ^a
Business biology (17.1)	Ninth grade	31.1	10.6	0.025 ^a
	College prep biology	31.3	12.6	0.018 ^a
	Honors biology	47.3	9.4	0.000 ^a
College prep biology (31.3)	Ninth grade	31.1	10.6	1.000
	Business biology	17.1	5.5	0.018 ^a
	Honors biology	47.3	9.4	0.000 ^a
Honors biology (47.3)	Ninth grade	31.1	10.6	0.000 ^a
	Business biology	17.1	5.5	0.000 ^a
	College prep biology	31.3	12.6	0.000 ^a

^aSignificant at *p* < 0.05^a.

surveys. Students used computers in two rooms. In one room there were 10 students working individually on iBooks. In the other room 14 students worked on iMacs; 10 students worked in five pairs and 4 students worked individually. They used an earlier version of the same set of activities used in the “L” high school implementation. The Invisible Dragons activity was used for the first time on the last day of the field test.

A total of 387 log files were generated, of which 306 or 79% were usable files. Files that contain no date or data beyond the user name and the activity name are unusable. They are caused either by the user deciding not to run the activity or by the activity script crashing.

An excerpt from an unprocessed log file is shown in Table IX. The user’s name has been changed for privacy reasons. Each log contains the following kinds of data: user name, date and time, question and answer, and actions (basically mouse clicks), automatically identified with XML tags such as <user> or <date>. The XML tags enable the logs to be processed for different purposes.

Analyzing Log Files for BioLogica Use

The objective in analyzing log files was to identify and characterize variables we could use to quantify and compare the use of BioLogica across students. What kinds of data are useful in explaining the results of the pre- and posttests? Given the variability across implementations, one of the first questions is Which of the activities did the students complete? Related process questions are How much time did learners spend on an activity?; How mindful was their interaction with the activity?; What challenges did they accomplish?; How easily?; Which questions did they

Table IX. Selections From an Unprocessed Log File Generated by the Introduction Activity

- <log>
<user>julia</user>
- <question>
<date>2001.04.17.21.46.01 04/ 17/ 01 21:46:01</date>
Good job! You’ve created your first dragon. How would you describe it? Does anything surprise you about this dragon?
Type your description in the box below.
<answer>this dragon has no arms or wings and also no fire.</answer>
</question>
- <question>
<date>2001.04.17.21.46.18 04/ 17/ 01 21:46:18</date>
Quite a variety of dragons here! Dragons apparently come in many different colors. What OTHER differences do you notice? Type your answer in the box below.
<answer>There are many types of dragon. I pictured dragons to look like the first male on my list.</answer>
</question>
- <question>
<date>2001.04.17.21.57.40 04/ 17/ 01 21:57:40</date>
What did you notice as you examined their chromosomes?
Type your answer in the box below.
<answer>You could have the same chromosomes but if you had the wrong mixer you could kill the dragon.</answer>
</question>
- <action>
<date>2001.04.17.22.03.17 04/ 17/ 01 22:03:17</date>
Matched comparison dragon after 28tries.
</action>
- <question>
<date>2001.04.17.22.20.19 04/ 17/ 01 22:20:19</date>
Explain why the dragons you created in this activity look different from one another.
<answer>they have different types of chromosomes that make different cells.</answer>
</question>
</log>

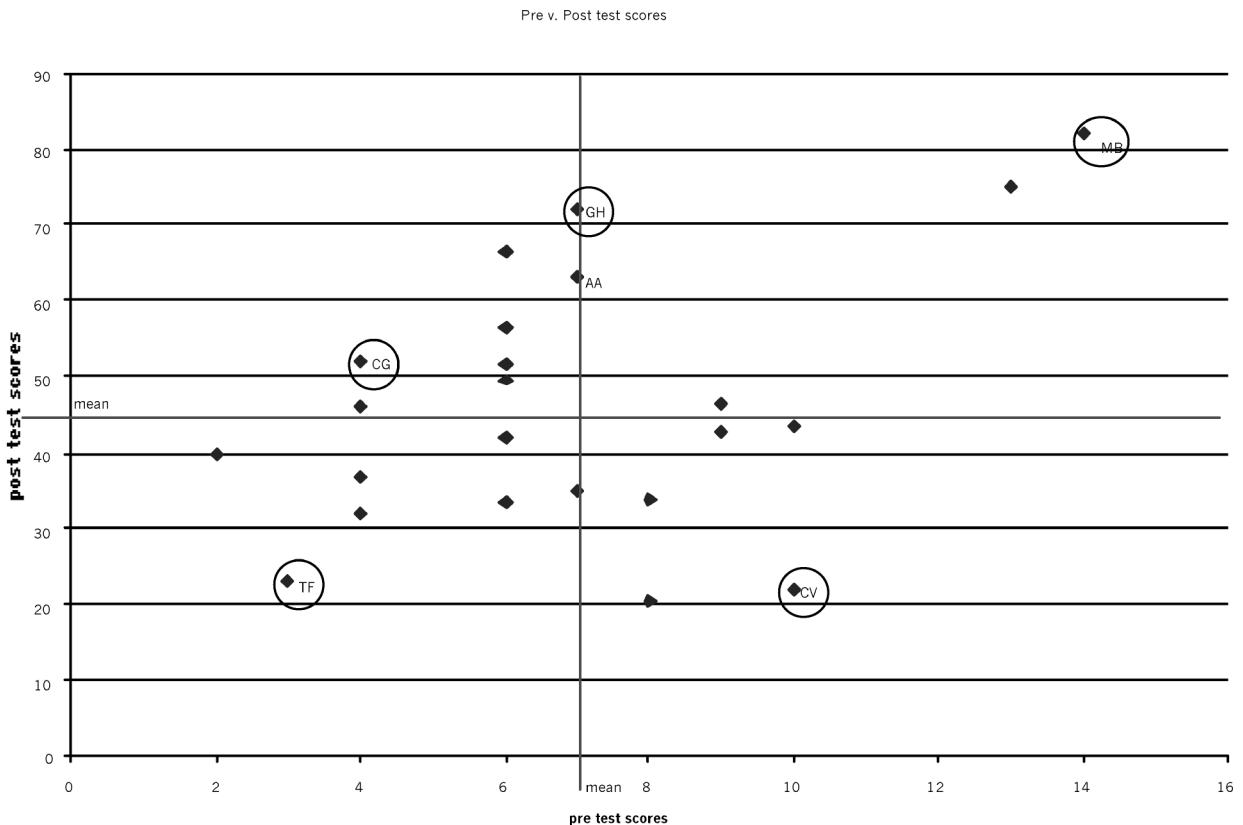


Fig. 3. Scatter plot of posttest scores against pretest scores for field test participants.

attempt?; Were their answers correct?; How thoughtful were their answers? We cannot yet answer these questions with a reasonable degree of reliability or effort.

Even from the brief excerpt in Table IX, it is clear how time-consuming is to make sense of unprocessed log files. Therefore, we chose to analyze the logs of a small subset of learners who represented different levels of prior knowledge and performance on the posttest.

We began by analyzing the pre- and posttest scores to identify the subset of learners. We plotted the posttest scores against the pretest scores (see Fig. 3).

We chose to examine the log files for five students, labeled TF, CG, CV, GH, and MB. MB was chosen for high scores on both the pretest and the posttest. GH was chosen to represent a successful student with an average pretest score. TF was chosen because he came in with a low pretest score and didn't perform very well on the posttest. CG and CV were chosen as a result of observations made while they were working through BioLogica activities. CG is dyslexic and reads

with difficulty, but worked diligently on activities. CV scored well on the pretest but was observed mindlessly clicking through, rather than reasoning through, an activity.

The next challenge was identifying the logs of these students. Students could type in any name or variation of that name when they logged in to use an activity. In this study we were able, with reasonable yield and effort, to connect the log files to individual learners.

To determine which activities the learner used, we created a computer program to strip out the XML tags and generate a file of tab-delimited text that could be imported into a database or spreadsheet, which could then be sorted. Table X shows the activities used by student GH in the order in which they were used, when, and for how long. There were also four empty logs indicating bugs or aborted launches.

The initial approach to analyzing the log files was to read through the logs of one student at a time in order to develop a sense of that student's path through BioLogica, what challenges and questions

Table X. Activities Used by GH

Activity	Date	Time	Length of time
Introduction	4/17/01	21:33:53	0:31
Rules	4/17/01	22:19:44	0:53
Rules	4/17/01	23:06:57	0:01
Meiosis	4/18/01	1:04:17	0:18
Inheritance	4/18/01	1:32:31	0:03
Monohybrid	4/18/01	1:37:40	0:14
Mutations	4/18/01	1:53:28	0:27
Mutations2	4/18/01	2:25:38	0:20
HornsDilemma	4/18/01	21:21:08	0:03
HornsDilemma	4/18/01	21:26:00	0:09
Dihybrid	4/18/01	21:44:23	0:17
Plates	4/18/01	23:17:18	0:05
Sex-linkage	4/18/01	23:25:27	0:21
<i>Total</i>			3:42

were difficult or easy, and how engaged the student seemed to be. By the time we had read through the logs for all five students, we began to see patterns of data that might be relevant and quantifiable. In addition to the total-time, candidates for further analysis included the proportion of questions answered, the proportion of multiple choice questions answered correctly, how many attempts it took to solve a challenge, how much time it took to solve the challenge, the length of time each explanation was displayed. To date we have been able to investigate only a few of these possibilities.

We created another computer program to process the logs to calculate how many questions the user encountered and answered in an activity and how many words were in their answers. We learned that the data on number-of-questions-encountered and the number-of-words-per-answer were a better measure of the state of development of the activities than of the learner's engagement with the activity.

We continued to search for some measure of the quality of a learner's interaction with the activity. We knew from classroom observations that time spent with an activity does not automatically equate to time engaged with an activity. Similarly, interaction with the activity in terms of actions taken or number of mouse clicks does not necessarily equate to mindful interaction. We conceptualized this quality measure as an index of interaction and operationalized it as the length of time a user spends with an activity divided by the number of actions taken by the user. A low index of interaction indicates many actions in a short period of time, while a high index of interaction indicates a few actions in a long period of time. Crashes also affect the index of interaction. If a user has been

through part of the activity before, the tendency is to quickly click through to reach the new portion of the activity.

We learned that our data logging is not yet systematic enough to produce meaningful numbers for index of interaction. The median index of interaction varied over the 11 logged activities from 0.08 to 2.0. This is primarily a reflection of the variation across activities in terms of what gets logged, how many challenges and questions were posed, etc. For a given activity there does appear to be a range of productive interaction with outliers at both extremes. One extreme is the very low index of interaction calculated for CV who clicked mindlessly through the activities, playing BioLogica as if it were a game. The other extreme, a very high index of interaction, was calculated for CG, a dyslexic but diligent learner, who took a very long time, even with assistance, to read the instructions and information in BioLogica, but who could then reason through the activities very well. Because of the length of time required for each activity, this learner did not complete all the activities. Since learners may use different numbers of activities, an index of interaction averaged across all activities cannot be used for comparison across learners.

Analyzing Log Files for Learner's Models and Understanding

We were able to make the log files easier to read with text harvests, like that shown in Table XI. Text harvests grew out of requests from teachers for a way of assessing students' work with BioLogica. A report generator strips out all the XML tags and other entries leaving only the questions in an activity and the answers entered by the learner.

Text harvests make it easier to assess a learner's models and understanding of genetics. They are much easier to read than log files and can be coded and scored just like pre- and posttests. However, they don't enable us to infer much about learners' learning or reasoning strategies. We need to consider the learner's performance on the challenges as well. Taken together, the pre- and posttests, text harvests, and log files can be used to describe a learner's path through BioLogica. An example using the data generated by GH is included in Appendix C. The log files of different students can be compared for a given activity.

For this paper, the findings of Study 2 relate to what we've learned about creating and analyzing log

Table XI. Text Harvest From Processed Log File

Introduction roberto <04/ 17/ 01 22:26:40>	
Q:	Quite a variety of dragons here! Dragons apparently come in many different colors. What OTHER differences do you notice? Type your answer in the box below.
A:	Some dragons have wings, some have arms, legs, both, or none at all. Others have different shaped tails, some have horns or none, some breathe fire or not, and they have different body shapes.
Q:	What did you notice as you examined their chromosomes? Type your answer in the box below.
A:	The dominant alleles usually gave them the characteristics that they have.
Q:	In particular, how do the chromosomes of male and female dragons differ? Type your answer in the box below.
A:	The male and female chromosomes respond differently to the dominant allele.
Q:	Why do you think you can't match both dragons to the comparison dragon? Type your answer in the box below.
A:	I don't think the female dragon has the genes for the yellow color dragon.
Q:	What do you think 'phenotype' means? Tell us in your own words and give us an example.
A:	Phenotype is the physical characteristics of a person who gets it from the genes. An example would be a girl with wavy brown hair, big brown eyes, short stature, wing shaped eyebrows, heart shaped face, etc...
Q:	What is the connection between genotype and phenotype? That is, how does genotype relate to phenotype? Tell us in your own words and give us an example.
A:	They relate to each other because they involve chromosomes and alleles. The genes from the genotypes give the physical characteristics of someone, this is the phenotype. Ex. TT this genotype gives the phenotype tall genes to someone.
Q:	Explain why the dragons you created in this activity look different from one another.
A:	They have different colors, shapes, sizes, legs, horns, tails, wings.
Q:	Where are a dragon's genes located? Click all that apply.
A:	chromosomes alleles DNA.
Q:	What did you know about phenotype and genotype before you started this activity?
A:	Phenotype are the physical characteristics Genotype are the letters of the genes.
Q:	What do you think you learned while working with this introductory activity?
A:	About the different male and female characteristics.
Q:	What question(s) do you currently have about phenotype and genotype?
A:	none at the moment.

files. We present these as an annotated wish list of requirements for data logging.

- Consistent user IDs

In order to use log files as evidence of process and reasoning, each file must be attributable to individual learners. A log-in

procedure that requires students to use pull down menus to select their name from the class roster and enter a password is now used.

- Systematic logging of and scoring rubrics for student performance on challenges

This will require analysis of existing activity scripts and the logs they generate to develop the fine-grained data needed and ways of postprocessing that data for easier analysis.

- ID tags for questions embedded in scripts

Each question embedded in a script should have an identifier that enables a person or program to pull all the responses to that question from the database.

- Precode questions and tasks

Each question, action, and task could be associated with one or more data analysis codes to facilitate content and statistical analysis.

- Fine-grained time stamps and calculations of time spent working on pedagogical elements

This also requires consistent use of pedagogical elements within scripts in order to calculate time spent on activities such as reading explanations or manipulating alleles.

The task of mining useful data from log files continues as we try to develop ways of generating logs that are more systematically related to research questions and variables and ways of analyzing logs in less labor intensive ways that are useful to teachers and students as well as researchers.

DISCUSSION

In the studies presented we have tried to illustrate how fruitful it can be to conduct classroom-based research with interactive curricula. We have also illustrated how difficult such studies can be. Conducting the kinds of studies presented in this paper requires the systematic development of a series of tasks that engage learners, scaffolding that helps them learn how to accomplish the tasks, ongoing assessments to help us understand what and how students are learning, and how to use technology to support and facilitate all of the above.

Study 1: "L" High School, May 2001

The ninth grade students in the experimental groups used a more stable version of BioLogica than in previous implementations, also, scaffolding had been put into about half of the activities at this point in development. Thus, the findings that the experimental group outperformed control group was welcomed! When we examine the kinds of items on which they outperformed the control group, we see that it is primarily on the easier type of reasoning from cause-to-effect which is a good starting point. We don't know from the data logs how many of the activities students actually completed because of the difficulty of matching logs to individual students. The experimental groups also outperformed the control groups on the structure and process items. This is encouraging because parts and processes are what students manipulate in many BioLogica activities. They also are critical aspects of model-based learning (Buckley, 2000; Buckley and Boulter, 2000; Gobert, 2000; Gobert and Buckley, 2000). Their performance on Punnett squares is also encouraging because of the frequency with which Punnett squares are used as reasoning tools in BioLogica activities.

This paper has focused on the cognitive aspects of model-based learning, but there are individual and classroom factors influencing whether and to what extent a learner may actually engage in model-based learning. The presumed motivation for engaging in model-based learning is a desire to understand. Motivation theorists conceptualize this desire as an intrinsic love of "learning for learning's sake" (Ames, 1992; Nicholls, 1989). It is also described as intrinsic motivation (Deci and Ryan, 1985; Lepper and Chabay, 1985; Lepper and Malone, 1987), a learning orientation rather than a performance orientation (Dweck, 1986), or an intentional learning stance (Bereiter and Scardamalia, 1989). Extrinsic motivation arising from teacher assessment is also a factor in students' engagement in classroom activities whether computer-based or not.

We acknowledge that both intrinsic motivation and the extent to which extrinsic factors may influence motivation and learning differ not only by student but also by the context and culture in which the student learns (Christie, 2002). Christie's work on students' perceptions of learning (Christie, 1999, 2001, 2002) has shown promising indications that motivational constructs also vary within domain and between tasks, according to the modality, e.g., text-based versus computer-based representations. Thus, students'

achievement cognitions (Christie, 2001, 2002), their understanding of the nature of models (Gobert *et al.*, 2002; Gobert and Discenna, 1997) and of science (Schommer, 1993) may influence not only their motivation to engage in learning but also the strategies they employ in response to the task (Songer and Linn, 1991), and ultimately, the outcomes.

Lastly, there are a host of pragmatic factors such as the availability of computers and classroom time that constrain how much and when students use BioLogica and whether it is integrated into the rest of the biology curriculum.

Future Research and Development

The results of these studies have informed the templates for the activities created for the Modeling Across the Curriculum (MAC) project, an IERI-funded project at the Concord Consortium. The MAC project is a large-scale 3-year longitudinal study of the learning of high school students as they use our computer-based activities to develop and revise their mental models of domain phenomena in physical science, biology, and chemistry. MAC is developing computer-based learning activities for each of these areas. The development of the BioLogica and Dynamica (physics) activities are taking place at The Concord Consortium; the development of the Connected Chemistry activities is taking place at Northwestern University in collaboration with Uri Wilensky. These activities support model-based learning through scaffolded interactions with manipulable models of domain phenomena, thus providing content, tools, and contexts.

Now in its third year, MAC has (a) developed computer-based learning activities that scaffold learners' interactions with interactive, manipulable models of domain phenomena and that record and react to the learners' interactions; (b) developed and validated online pre- and posttests aligned to the learning objectives and activities; (c) developed a technological infrastructure to deliver MAC learning activities to classrooms and collect data when students use MAC activities; (d) developed a web portal that provides access to reports of students' use of activities to teachers, students, and researchers; and (e) developed Classroom Communiquès, web forms filed by teachers via the web portal to provide data about integration of MAC activities into the taught curriculum. MAC learning activities are currently in use in more than a dozen schools across the country.

There is considerable research begging to be done on assessment using interactive items (National Research Council, 2002). We're not talking about adaptive testing but rather about using interactive items to assess a learner's problem solving or model-building skills in a hypermodel environment such as BioLogica. We question how well traditional test items assess the state of learners' models of and reasoning about phenomena and plan to investigate the affordances that the hypermodel environment brings to this question.

APPENDIX A: DESCRIPTION OF BioLogica ACTIVITIES

(1) Introduction—What Do Dragons Look Like and Why?

Introduction guides the user through BioLogica's representations of chromosomes, genes, and alleles and stresses the connection between genotype and phenotype. As learners use pulldown menus in the Chromosome View to change allele combinations, they see changes in the Organism View of BioLogica's dragons.

Key Concepts

- Genotype determines phenotype
- Definition of traits, genotype, phenotype, chromosomes, genes and alleles

(2) Rules—How Do Genes Affect Appearance?

The Rules activity is divided into three subactivities that can be invoked independently via a menu selection. These are (1) traits, which deals with the four autosomal traits of dragons genetics—horns, wings, number of legs, and shape of tail; (2) firebreathing, which introduces a sex-linked trait; and (3) colors, which are polygenic and pleiotropic traits. One of the color genes contains a recessive lethal allele. Traits and firebreathing are essential introductions to dominance, recessive, incomplete dominance, and sex-linkage. At this point in its development, BioLogica does not build on the colors subactivity.

Key Concepts

- Genotype determines phenotype

- Particular allele combination produces particular trait
- Dominance/recessive/sex-linked inheritance

(3) Meiosis—What Do Meiosis and Fertilization Have to Do With Making Offspring?

Meiosis is also subdivided into three subactivities. Introduction to Meiosis, Meiosis Shuffles and Deals! and Designer Dragons. *Introduction to Meiosis* focuses on learning to use the interface and linking the representations with the concepts of gametes, meiosis, and fertilization. It provides an introductory view into the process of gamete creation and the random distribution of the alleles and allows the student to inspect the alleles in each gamete and to choose combinations of gametes to fertilize. *Meiosis Shuffles and Deals!* links the representations with the names of the phases of meiosis found in textbooks. *Designer Dragons* offers students a series of challenges in the form of creating specific offspring by examining the chromosomes in the gametes from each parent and selecting those that will produce the desired phenotype in the offspring.

Key Concepts

- Meiosis
- Fertilization
- Chromosome segregation
- Gamete selection

(4) Inheritance—What Determines What the Offspring Look Like?

Unlike the Meiosis activity where students select gametes to determine combinations of alleles that will be “passed down” to the offspring, the Inheritance activity uses the pedigree view, which emphasizes the randomness of this process in nature. It thus serves as an introduction to the role of probability in genetics. The Monohybrid activity is a useful follow-on to this one.

Key Concepts

- Monohybrid inheritance
- Proportions

(5) Monohybrid—Do Traits Really Skip Generations?

Monohybrid introduces and uses Punnett squares to help students understand how the combinatorics of meiosis and fertilization (an exhaustive count of all possible combination of parental genes) enables one to derive the probability that an offspring will possess a certain trait, and leads eventually to a prediction of the likely fraction of offspring with the trait in a sample of size n .

Key Concepts

- Monohybrid inheritance
- Punnett squares
- Combinatorics → probability → statistics

(6) Mutations—What Happens When You Change the DNA?

This activity introduces and uses the BioLogica DNA View, which provides an expanded view of the red lines on chromosomes that represent genes. The activity introduces students to the DNA and base pairs that form particular alleles and challenges students to modify the DNA of a dragon and to observe the consequences.

Levels/Views: Organism View, Chromosome View, DNA View

(7) Mutation Inheritance—How Are Mutations Inherited?

This activity uses mutations to explore alternative modes of inheritance of traits that are controlled by genes with more than two alleles.

Key Concepts

- Modes of inheritance, non-Mendelian genetics

(8) Horns Dilemma—Can Two Horned Parents Have a Hornless Baby?

This activity focuses students' attention on the connection between parental genes and those of the

offspring, in the context of posing a challenge that requires the student to alter a parental gene, making it heterozygous so that a homozygous recessive offspring can result.

Key Concepts

- Inheritance of “hidden” traits

(9) Dihybrid

This activity explores what happens when you study the inheritance patterns for two traits at a time. It provides experience with dragons with traits on the same chromosome and with peas with traits on different chromosomes.

Key Concepts

- Effect of independent assortment of chromosomes on the inheritance of two traits

(10) Sex Linkage—What Difference Does It Make if a Gene Is on the X Chromosome?

This activity begins with a review of how the X and Y chromosomes interact to produce male and female (introduced in the Rules activity). Once it is clear to the students that female dragons are XX and males are XY, the activity uses the firebreathing (recessive, X-linked) trait to help them learn how sex-linked traits are inherited.

Key Concepts

- Sex-linked inheritance

(11) Scales—What Causes That Scaly Skin?

This is an advanced activity that guides the student through the process of investigating a trait—scaly skin—the gene for which has been hidden. Four questions are asked: Are scales genetically inherited? Are they recessive or dominant? Are they sex-linked or autosomal? And which chromosome are they on?

Key Concepts

- Inference of mode of inheritance from statistics

(12) Plates

Plates is similar to scales, but with incompletely dominant traits.

Key Concepts

- Investigating how new traits are inherited from statistics

(13) Invisible Dragons

In this activity we test the student's ability to solve a real genetics puzzle. They are presented with two

invisible dragons and their task is to figure out the genetic makeup of this couple. They may make crosses, look at the chromosomes and even make a backcross, but all of these costs money. The players start out with \$20000 in the bank. Each procedure costs money as does each wrong answer. Players make money by answering questions correctly.

GenScope

This is an open-ended interface to all the BioLogica functionality. It has no "story line" or monitoring functions. GenScope is essentially a BioLogica implementation of GenScope that runs on the Mac and the PC.

APPENDIX B: SUMMARY TABLE FOR SITE "L" (STUDY 1)

Summary Table of Univariate Analyses of Variance Using Pretest as a Covariate

Question subgroup	Dependent variable	Mean		F statistic	Significance (p value)	Number of items	Total number of points
		Control	Experimental				
All questions	Posttest	30.000	34.431	5.355	0.007 ^a	52	106
Concept	Inheritance	5.778	6.472	1.870	0.162	10	15
	Dominance	3.875	4.722	2.524	0.088 ^b	8	13
	Sex-linkage	1.000	1.139	1.143	0.325	6	10
	Genotype/phenotype	8.208	9.028	1.527	0.225	11	11
	Monohybrid	7.444	9.167	5.798	0.005 ^a	8	19
	Dihybrid	1.250	1.694	4.261	0.018 ^a	4	6
	Pedigree	0.778	0.750	1.098	0.339	3	6
Type of reasoning	cew	4.472	5.083	5.734	0.005 ^{a,c}	6	6
	ecw	2.569	2.750	2.630	0.079 ^b	4	4
	ceb	5.958	7.056	5.951	0.004 ^a	11	18
	ecb	5.792	5.694	3.572	0.033 ^{a,c}	15	22
	Punnett square	5.250	5.653	4.280	0.018 ^a	4	16
	Structure	2.917	4.139	3.985	0.023 ^a	7	10
	Process	0.667	1.194	7.420	0.001 ^a	3	6

Notes. Ninth grade classes: total number of students = 72 (control group = 36 students; experimental group = 36 students).

^aSignificant at $p < 0.05\alpha$.

^bSignificant at $p < 0.10\alpha$.

^cLevene's Test for Equality of Error Variances is significant at $p < 0.05\alpha$.

APPENDIX C: LEARNING PATH OF GH

Based on responses to open-ended questions on the pretest, GH began the study with a model of inheritance as something you receive from parents or grandparents and that it is a random process. "I guess it's when genes are passed down to you." The chromosome is "some sort of gene" and traits as "some quality of you." No responses were offered for fertilization, pedigree, or sex-linked. This is a potentially good basis for beginning to use BioLogica in that GH recognizes

genotypic entities (genes and chromosomes) as well as phenotypic entities(traits).

According to available logs, GH did not complete the Scales activity. GH seemed to have little difficulty completing most of the challenges in the activities—until encountering the pedigree view in Monohybrid. GH answered questions, could define dominant and recessive in terms of representations by upper case and lower case letters, respectively, but sometimes had difficulty applying them to multiple choice questions embedded in the Rules activity. GH was able to reason

with and manipulate the meiosis representation and make the connection between the representations of chromosomes and alleles in chromosome and meiosis views. GH could reason with Punnett squares but seemed to lack a strategy for making crosses and reasoning about the results in pedigree view. [Monohybrid seems to be a place where scaffolding is needed to help learners figure out and reason with this new representation.] Dihybrid and Plates logs provide further evidence of difficulties with pedigree use and reasoning. In Plates, for example, GH made just one cross, looked at chromosomes and concluded (incorrectly) that plates are recessive because “the quality didn’t seem to travel through the genes often.” The evidence for dominance seems to be frequency.

Responses to the questions on the posttest suggest that there is some confusion about the physical relationship among chromosomes, genes, and alleles. The chromosome is now “the group of genes or alleles that give traits” and a *gene* is “a group of DNA.” But *allele* is “a group of genes/DNA,” which suggests that allele is not understood. Heterozygous, homozygous, and autosomal alleles are understood, but dominant, recessive, and sex-linked are not. Dominance seems to be associated solely with the frequency with which it turns up in the offspring. For example, dominance is “the overpowering type of alleles that determine more in the offspring outcome” while recessive is “the ‘weak’ type of alleles that determine less in offspring.” This is echoed in a later explanation “If there is complete dominance, the offspring should all have the trait.” Sex linked is “when an allele is different in each of the sexes (on either the X or the Y chromosome).”

The mechanisms of inheritance (meiosis and fertilization) were defined as “the process in which DNA is taken from the parents and made into an offspring” and “when an egg from the female and sperm from the male is brought together to form an offspring,” respectively. It is not clear whether this is part of a mental model of meiosis, or a text-based definition.

In the NewWorm portion of the posttest, the learner was able to map from genotype to phenotype and vice versa, determine the sex of the NewWorm, and create and use Punnett squares effectively to reason about monohybrid inheritance. The facts that (1) only one of the three dihybrid questions was answered correctly without explanation and that (2) GH was able to identify the gametes and offspring in Punnett square but (3) did not associate them with the outcomes of meiosis and fertilization, suggests to me the

lack of a mental model. If GH had a mental model of the chromosomes and the process of meiosis and fertilization, it would be possible to reason about dihybrid inheritance by visualizing the movement of alleles on the same or different chromosomes.

GH’s ability to reason with the pedigree representation is unclear. GH seems to have forgotten or did not apply BioLogica activities in this explanation, “if having small nostrils was recessive, then the baby couldn’t have them, because the parents don’t. It then has to be dominant.” This completely ignores the possibility that the parents could be heterozygous dominant in the trait, the focus of more than one activity. Answers to the multiple-choice questions were little changed from the pretest, just one point increase.

ACKNOWLEDGMENTS

This paper is based on research conducted as part of the Fostering Transfer Project supported by the National Science Foundation under Grant No. REC-0087579. The Modeling Across the Curriculum Project is supported by the National Science Foundation and the U.S. Department of Education under the Inter Agency Education Research Initiative (IERI #011569). Any opinions, findings, and conclusions expressed are those of the authors and do not necessarily reflect the views of our funding agencies.

REFERENCES

- Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of Educational Psychology* 84: 261–271.
- Bereiter, C., and Scardamalia, M. (1989). Intentional learning as a goal of instruction. In Resnick, L. B. (Ed.), *Knowing, Learning and Instruction: Essays in Honor of Robert Glaser*, Erlbaum, Hillsdale, NJ, pp. 361–392.
- Bransford, J., Sherwood, R., Vye, N., and Rieser, J. (1986). Teaching thinking and problem solving. *American Psychologist* 41: 1078–1089.
- Brewer, W. F. (1987). Schemas versus mental models in human memory. In Morris, P. (Ed.), *Modelling Cognition*, Wiley, Chichester, UK, pp. 187–197.
- Buckley, B. C. (1992). *Multimedia, Misconceptions and Working Models of Biological Phenomena: Learning About the Circulatory System*, Unpublished Doctoral Dissertation, Stanford University.
- Buckley, B. C. (2000). Interactive multimedia and model-based learning in biology. *International Journal of Science Education* 22: 895–935.
- Buckley, B. C., and Boulter, C. J. (1999). Analysis of representations in model-based teaching and learning in science. In Paton, R., and Neilson, I. (Eds.), *Visual Representations and Interpretation*, Springer, Liverpool, England, pp. 289–294.
- Buckley, B. C., and Boulter, C. J. (2000). Investigating the role of representations and expressed models in building mental models. In Gilbert, J. K., and Boulter, C. J. (Eds.),

- Developing Models in Science Education*, Kluwer, Dordrecht, Holland, pp. 105–122.
- Chinn, C. A., and Brewer, W. F. (1993). The role of anomalous data in knowledge acquisition: A theoretical framework and implications for science instruction. *Review of Educational Research* 63: 1–49.
- Christie, M. (1999). “We understood it more ‘cause we were doin’ it ourself:” Students self-described connections between participation and learning. *Paper Presented at the American Educational Research Association*, Montreal, Canada, April 20, 1999.
- Christie, M. (2001). Portraits of academic beliefs. *Paper Presented at the Annual Meeting of the New England Educational Research Organization (NEERO)*, Portsmouth, NH, April 26, 2001.
- Christie, M. (2002). The role of classroom culture and learning contexts in achievement beliefs. *Paper Presented at the Annual Meeting of the American Educational Research Association (AERA)*, New Orleans, LA, April 1–5, 2002.
- Clement, J. (1989). Learning via model construction and criticism: Protocol evidence on sources of creativity in science. In Glover, J. A., Ronning, R. R., and Reynolds, C. R. (Eds.), *Handbook of Creativity: Assessment, Theory and Research*, Plenum, New York, pp. 341–381.
- Deci, E., and Ryan, R. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*, Academic Press, New York.
- deKleer, J., and Brown, J. S. (1983). Assumptions and ambiguities in mechanistic mental models. In Stevens, A. L., and Gentner, D. (Eds.), *Mental Models*, Erlbaum, Hillsdale, NJ, pp. 155–190.
- Dweck, C. S. (1986). Motivational processes affecting learning. *American Psychologist* 41: 1040–1048.
- Gentner, D., and Stevens, A. L. (Eds.). (1983). *Mental Models*, Erlbaum, Hillsdale, NJ.
- Gobert, J. (2000). A typology of models for plate tectonics: Inferential power and barriers to understanding. *International Journal of Science Education* 22: 937–977.
- Gobert, J., and Buckley, B. C. (2003). *Scaffolding Model-Based Reasoning: Representations and Cognitive Affordances*, The Concord Consortium, Concord, MA.
- Gobert, J., and Discenna, J. (1997). The relationship between students’ epistemologies and model-based reasoning. *Paper Presented at the American Educational Research Association*, Chicago.
- Gobert, J., Snyder, J., and Houghton, C. (2002). The influence of students’ understanding of models on model-based reasoning. *Paper Presented at the Annual Meeting of the American Educational Research Association*, New Orleans, LA, April 1–5, 2002.
- Gobert, J. D., and Buckley, B. C. (2000). Introduction to model-based teaching and learning in science education. *International Journal of Science Education* 22: 891–894.
- Gobert, J. D., and Clement, J. J. (1999). Effects of student-generated diagrams versus student-generated summaries on conceptual understanding of causal and dynamic knowledge in plate tectonics. *Journal of Research in Science Teaching* 36: 39–53.
- Hickey, D. T., Kindfield, A. C. H., Horwitz, P., and Christie, M. A. (1999). Advancing educational theory by enhancing practice in a technology-supported genetics learning environment. *Journal of Education* 181: 25–55.
- Hickey, D. T., Kindfield, A. C. H., Horwitz, P., and Christie, M. A. (2003). Assessment-oriented scaffolding of student and teacher performance in a technology-supported genetics environment. *American Educational Research Journal* 40(2): 495–538.
- Hickey, D. T., and Kindfield, A. C. H. D. (1999). Assessment-oriented scaffolding of student and teacher performance in a technology-supported genetics environment. *Paper Presented at the Annual Meeting of the American Educational Research Association*, Montreal, Canada.
- Hickey, D. T., Wolfe, E. W., and Kindfield, A. C. H. (1998a). Assessing learning in a technology-supported genetics environment: Evidential and consequential validity issues. *Paper Presented at the Annual Meeting of the American Educational Research Association*, San Diego, April 1998.
- Hickey, D. T., Wolfe, E. W., and Kindfield, A. C. H. (1998b). Assessing learning in a technology-supported genetics environment: Evidential and systemic validity issues. *Paper Presented at the Annual Meeting of the American Educational Research Association*, San Diego, April 1998.
- Horwitz, P. (1995). Linking models to data: Hypermodels for science education. *The High School Journal* 79(2): 148–156.
- Horwitz, P., and Christie, M. A. (1999). Hypermodels: Embedding curriculum and assessment in computer-based manipulatives. *Journal of Education* 81(2): 1–23.
- Horwitz, P., Schwartz, J., Kindfield, A. C. H., Yessis, L. M., Hickey, D. T., Heidenberg, A. J., and Wolfe, E. W. (1998). Implementation and evaluation of GenScope™ learning environment: Issues, solutions, and results. *Paper Presented at the Third International Conference of the Learning Sciences*, Charlottesville, VA.
- Johnson-Laird, P. N. (1983). *Mental Models*, Harvard University Press, Cambridge, MA.
- Kindfield, A. C. H. (1993/1994). Biology diagrams: Tools to think with. *Journal of the Learning Sciences* 3: 1–36.
- Kindfield, A. C. H., and Hickey, D. T. (1999). Tools for scaffolding inquiry in the domain of introductory genetics. *Paper Presented at the Annual Meeting of the American Educational Research Association*, Montreal, Canada.
- Kozma, R., Jones, T., Wykoff, J., and Russell, J. (1992). Multimedia, multiple representations, and mental models in chemistry. *Paper Presented at the Annual Meeting of the American Educational Research Association*, San Francisco.
- Larkin, J. H. (1989). Display-based problem solving. In Klahr, D., and Kotovsky, K. (Eds.), *Complex Information Processing: The Impact of Herbert A. Simon*, Erlbaum, Hillsdale, NJ, pp. 319–341.
- Larkin, J. H., and Simon, H. A. (1987). Why a diagram is (sometimes) worth ten thousand words. *Cognitive Science* 11: 65–99.
- Lepper, M., and Chabay, R. W. (1985). Intrinsic motivation and instruction: Conflicting views of the role of motivational processes in computer-based education. *Educational Psychologist* 20: 217–230.
- Lepper, M., and Malone, T. (1987). Intrinsic motivation and instructional effectiveness in computer-based education. In Snow, R. E., and Farr, M. J. (Eds.), *Aptitude, Learning and Instruction III: Conative and Affective Process Analyses*, LEA, Hillsdale, NJ, pp. 255–286.
- Lowe, R. (1993). *Successful Instructional Diagrams*, Kogan Page, London.
- Monaghan, J., and Clement, J. (1994). Factors affecting the efficacy of computer simulation for facilitating relative motion concept acquisition and visualization. *Paper Presented at the Annual Meeting of the American Educational Research Association*, New Orleans, LA.
- National Research Council. (2002). *Technology and Assessment: Thinking Ahead: Proceedings of a Workshop*, National Academy Press, Washington, DC.
- Nicholls, J. G. (1989). *The Competitive Ethos and Democratic Education*, Harvard University Press, Cambridge, MA.
- Norman, D. A. (1983). Some observations on mental models. In Stevens, A. L., and Gentner, D. (Eds.), *Mental Models*, Erlbaum, Hillsdale, NJ, pp. 7–14.
- 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.

- Schauble, L., Glaser, R., Raghavan, K., and Reiner, M. (1991). Causal models and experimentation strategies in scientific reasoning. *Journal of the Learning Sciences* 1: 210–238.
- Schommer, M. (1993). Epistemological development and academic performance among secondary students. *Journal of Educational Psychology* 85: 406–411.
- Songer, N. B., and Linn, M. C. (1991). How do students' views of science influence knowledge integration? *Journal of Research in Science Teaching* 28: 761–784.
- Stewart, J., and Hafner, B. (1991). Extending the conception of problem solving. *Science Education* 75: 105–120.
- Vosniadou, S., and Brewer, W. F. (1992). Mental models of the earth: A study of conceptual change in childhood. *Cognitive Psychology* 24: 535–585.
- White, B. Y., and Frederiksen, J. R. (1998). Inquiry, modeling, and metacognition: Making science accessible to all students. *Cognition and Instruction* 16: 3–118.