In Guzdial, M., Kolodner, J., & Bruckman, A. (Eds.) (1998). Proceedings of the Third International Conference of the Learning Sciences. Charlottesville, VA: Association for the Advancement of Computers in Education.

Implementation and Evaluation of the GenScope™ Learning Environment: Issues, Solutions, and Results

Paul Horwitz & Joyce Schwartz
The Concord Consortium. 37 Thoreau St., Concord, MA 01742 USA
paul@concord.org & joyce@concord.org

Ann C. H. Kindfield & Laura M. Yessis
Montclair State University, Dept. of Biology & Molecular Biology, Upper Montclair, NJ 07043 USA
ann.kindfield@montclair.edu

Daniel T. Hickey & Alex J. Heidenberg
Georgia State University, Dept. of Educational Psychology and Special Education, Atlanta, GA 30303 USA
dhickey@gsu.edu

Edward W. Wolfe
University of Florida, Dept. of Educational Foundations, Gainesville, FL 32611 USA
wolfe@nersp.nerdc.ufl.edu

Abstract: This special session describes a collaborative effort to implement and evaluate the GenScopeä software and associated curriculum in a broad range of secondary life science classrooms. This effort illustrates how technology-based environments can be used to meet existing content-area educational goals and highlights new approaches to conducting large-scale evaluations of such environments. The four presentations in this special session include (1) project overview and software design, (2) implementation design, (3) assessment design, and (4) learning outcomes. With the notable exception of four at-risk urban classrooms, students in the GenScope classrooms have demonstrated the same disappointingly modest reasoning gains as students using conventional genetics curriculum. Encouragingly, a forthcoming revision of the software will make it possible to build in the curricular features that contributed to the dramatically larger reasoning gains witnessed in the four at-risk classrooms that used GenScope.

1. Project Overview and Software Design (Horwitz)

GenScope is an open-ended exploratory software tool that students can use to investigate a variety of phenomena in genetics. The program is explicitly designed to encourage and support multi-level reasoning, from molecular biology to population genetics.[1] For the past five years, we have evaluated GenScope's educational effectiveness in both urban and suburban high school classrooms, in the greater Boston area and in Atlanta.[2] We videotaped the classes, interviewed teachers and selected students, and administered written assessments specially created to evaluate multi-level reasoning. We have conducted numerous workshops, ranging in length from two to thirty-five hours, to introduce teachers to the use of the software.

GenScope has been enthusiastically received by both students and teachers. Even in low SES inner-city schools with high drop-out rates and sub-average academic standards, students have become very engaged with the software and have shown themselves able to carry out quite complex investigations with minimal

[1]For more information, visit http://genscope.concord.org/

[2] This project is supported by the National Science Foundation, under grant number RED-955343

assistance. The effect on teachers has been equally remarkable, judging by their comments during workshops as well as subsequent self-reported changes in their teaching styles.

However, a central finding of the project (reported in the final section here) has been the general absence of any significant improvement on test scores that can be attributed directly to GenScope. We have administered identical tests to classes with arbitrarily assigned students taught by the same teacher. Some of the classes used GenScope while others followed the same curriculum, but without the computer. The results on the assessments of the two groups are statistically indistinguishable. This was unexpected.

This absence of larger gains in domain reasoning is now the focus of our continuing research. In our experience with GenScope we have found it to be too unstructured for most teachers and students to use effectively. Teachers have difficulty assessing what students are doing; students often do not perceive the connections between their activities on the computer and the scientific principles that underlie or explain them. Furthermore, the software does not familiarize students with the kinds of questions with which they will subsequently be confronted on the assessments. In the past we have taken the position that these important tasks are the responsibility of the teacher and the curriculum, and should take place "off line." However, we are now enhancing our technology in an effort to facilitate them. Specifically, we are implementing a successor to GenScope, called BioLogicaä, that will embed curriculum and assessment materials within the computer model.

BioLogica will run under the control of control programs called "scripts" that communicate with students by sending them messages and collecting information from them. Scripts will configure BioLogica to adapt it to particular activities. They can also be programmed to react to particular changes in the internal state of the model, which will enable them to respond appropriately to students' actions. The combination of these elementary functions can be quite powerful. For example, a script writer may choose to present students with a scientific puzzle, "look over their shoulder" as they work their way through it, ask them questions designed to induce reflection about their reasoning, and ultimately save a record of their work for later assessment.

2. Implementation Design (Schwartz)

Students and teachers have been involved in the design of GenScope since the beginning. As each level became operational, students were either brought into the office or we took the program to them and their teachers in the classroom. It was from the students and their problems in understanding various mechanisms of genetics that changes in design and curriculum came about. Very early on we were in an inner city school in Boston, trying out the meiosis part of the program. We had about 8 students in the class—all 10th graders. One young girl, Cynthia, didn't want to participate, hated computers, in fact, didn't know how to use a mouse. She refused to get involved. We asked the students to "make a baby by running meiosis and choosing the gametes with the proper alleles" to match the problem. She tried it, and kept making new babies long after we had moved on to another problem. Because of her we wrote a group of increasingly harder problems based on "making babies". These are still being used. Meiosis is particularly difficult for beginning students to understand. The graphics display of meiosis was modified several times until we were sure that it was clear enough for students to understand the process.

Teacher training workshops have been conducted over the past two years. In these workshops teachers have been taught how to use the software and have written puzzles of their own to use in their classrooms. The range of computer literacy among the teachers varied from those who barely know how to turn the computer on to those who can search the Internet using any one of several search engines.

Carolyn, an inner city school teacher, came to the workshop with the knowledge and ability to scan the Internet for information. She had developed a nice project about sickle cell anemia using the Net. As a result of the workshop, she expanded the project by using GenScope to study the genetics of the disease. Special Education teachers have found that GenScope is a real help in teaching this particular population of students. Ginny, a teacher of the deaf modified existing worksheets to take advantage of the graphic display of meiosis, crossover and alignment to help her students better understand these abstract processes. GenScope is now being used in urban and suburban classrooms in various parts of the country and even overseas. Although the workshops have been very successful in training teachers, the total number has been small. New methods of training are being investigated now with the hope of implementing them in the near future.

3. Assessment Design (Kindfield, Hickey, & Yessis)

For the purpose of evaluating student learning as a result of interacting with the GenScope Learning Environment, we developed an assessment instrument called the *NewWorm* assessment. The *NewWorm* uses a fabricated species, NewWorms, to systematically explore student understanding of introductory genetics concepts and reasoning. Constraints on the design of the NewWorm Assessment were the need to (a) use a paper-and-pencil format, (b) satisfy both (proximal) research and (ultimate) dissemination goals, (c) assess multi-level reasoning, (d) compare GenScope and non-GenScope users, and (e) assess a broad range of student populations. The NewWorm Assessment addresses these constraints by using a species whose genetics mimics that of GenScope dragons, but is novel and understandable to both GenScope and non-GenScope students, and by using questions that progressively shift from simple to complex forms of reasoning all in a paper-and-pencil format.

All NewWorm items can be classified along two primary dimensions: (1) Domain-general Reasoning Type (cause–to-effect, effect-to-cause, and process reasoning) and (2) Domain-specific Reasoning Type (within-generations and between-generations). In general, reasoning within generations is easier than reasoning across generations and reasoning from causes to effects (from genotypes to phenotypes[3]) is easier than reasoning from effects to causes (from phenotypes to genotypes) (Stewart, 1988; Stewart & Hafner, 1994). Reasoning about processes can be further divided into reasoning about process inputs and outputs versus reasoning about process events with the former generally being easier than the latter (Kindfield, 1994). Typical introductory genetics instruction focuses on within- and across-generation, cause-to-effect reasoning which often can be accomplished through pattern matching and application of algorithms with little to no understanding of the relationship between Mendelian inheritance and the underlying process of meiosis. To the extent that process reasoning is dealt with in typical introductory instruction, it is almost exclusively confined to reasoning about inputs and outputs (Kindfield, 1994). GenScope was designed to support the development of reasoning in all of these categories and thus all categories were represented in the NewWorm Assessment.[4] Table 1 displays each item reasoning type category along with brief descriptions of example problems from the NewWorm Assessment. In addition to these primary dimensions, some items can also be distinguished according to the particular genetics involved (e.g., autosomal vs. X-linked inheritance), the explicitness of provided information, and/or type of information used/sought (i.e., categorical, probabilistic, diagrammatic, short answer).

		Domain-General Dimension of Reasoning (Novice ⟷ Expert)		
		Cause-to-effect	Effect-to-cause	Process Reasoning
Domain-Specific Dimension of Reasoning (simple ⟷ complex)	Between-generations	**Monohybrid inheritance I**: given genotypes of two parents, predict genotypes and phenotypes of offspring	**Monohybrid Inheritance II**: given phenotypes of a population of offspring, determine the underlying genetics of a novel characteristic	**Punnett Squares** (input/output reasoning): describe Punnett Squares in terms of ploidy; **Meiosis-The Process** (event reasoning): given genetic make-up of an organism and the products of a single meiosis, describe the meiotic events that resulted in this set of products
	Within-generations	**Genotype to Phenotype Mapping**: given genotypes and info about NewWorm genetics, predict phenotypes	**Phenotype to Genotype Mapping**: given phenotypes and info about NewWorm genetics, predict genotypes	none (see footnote 3)

Table 1: Two dimensions of reasoning represented by items in the *NewWorm* assessment.

Extensive data supporting the evidential validity of our assessment practice has been gathered throughout the project (e.g., Hickey, Wolfe, & Kindfield, 1998). One key piece of validity evidence is derived by using multifaceted Rasch scaling of students scores . This technique locates each assessment item and each students score on a single linear scale. Figure 1 shows the scaled results in terms of the relative difficulties of the various item clusters for students in the 27 classrooms that completed the NewWorm. These results validate

[3] An organism's genotype for a particular characteristic is the organism's genetic make-up for that characteristic and its phenotype is its observable appearance for the characteristic.

[4] In the NewWorm Assessment, the processes of interest were meiosis and fertilization, both of which typically contribute to generational change and thus fall into between-generation domain-specific reasoning. Within-generation processes like transcription and translation were not dealt with in the GenScope curriculum or the NewWorm assessment.

our assumptions about the two dimensions of reasoning described above. Specifically reasoning within-generations is easier than between-generations, and cause-to-effect reasoning is easier than effect-to-cause, which in turn is easier than process reasoning. Other results not presented here validated our assumptions about more specific dimensions of reasoning (e.g., categorical versus probabilistic reasoning about inheritance) and various aspects of the domain (e.g., autosomal versus sex-linked traits).

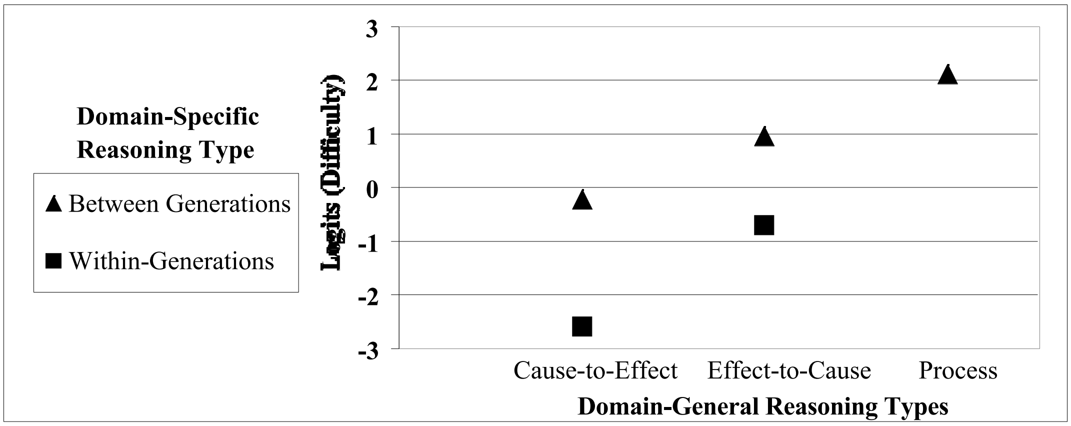

Figure 1: Relative difficulty of clusters of different items, by reasoning type.

4. Learning Outcomes. (Hickey, Kindfield, Wolfe, & Heidenberg)

In the 96-97 school year, students in 14 secondary life science classrooms, including three classrooms where GenScope was implemented, completed the *NewWorm* assessment (or the *NewFly* variant) before and after genetics instruction. Multifaceted Rasch analyses of these data revealed that students in all classrooms made similarly modest gains in domain reasoning, and very few developed the kind of higher-level reasoning targeted both by GenScope and by current educational reform standards. In response, the software developers made additional changes to the GenScope software and the assessment team developed a set of curricular activities that use the familiar GenScope dragons to scaffold development of the domain reasoning represented by the NewWorm. These activities were called *Dragon Investigations* and were made available to implementation teachers as supplementary activities.

In the 97-98 school year, students in 24 GenScope classrooms and two comparison classrooms completed the NewWorm before and after genetics instruction. Figure 2 displays the increases in mean proficiency of various groups of students. With one obvious exception, students in both the GenScope classrooms and the comparison classrooms again demonstrated the same modest reasoning gains documented in the 96-97 school year. These gains ranged from .60 to .83 of the 4.6 logit range of the scale displayed in Figure 1. Consider that such gains left the suburban general track students *after* instruction below the level of their college prep schoolmates *before* instruction.

The exception to these unimpressive findings was in the four Urban At-Risk classrooms that used GenScope in 97-98. Students in these classrooms (represented by the gray squares) gained 2.1 logits overall, significantly larger than the .38 gain in the three Urban At-Risk comparison classrooms (represented by the white squares) [$F(1,93) = 14.6$, $p < .001$]. While the number of class periods devoted to genetics was roughly the same in both groups (25-30), teacher surveys and interviews revealed that the two teachers who taught the four Urban At-Risk GenScope classes relied very heavily on the curricular activities that the assessment team developed to scaffold student performance on the NewFly assessment.

The validity inquiry reported in Hickey, Wolfe, and Kindfield (1998) concluded that the assessment-oriented curriculum activities would not fundamentally compromise the evidential validity of the NewWorm assessment. Specifically, we argued that such activities can appropriately scaffold the development of student reasoning while not reducing the complex problems on the NewWorm to simple algorithms. Implementation studies now underway will more rigorously examine the impact of efforts to enhance our assessment system's positive consequences for learning (as in the notion of *systemic validity* advanced by Collins and Frederiksen, 1989) on evidential validity. Given that some of the scripted activities being planned for the new BioLogica software will be modeled after these activities, such inquiry will be vital for subsequent evaluations of learning in the GenScope environment.

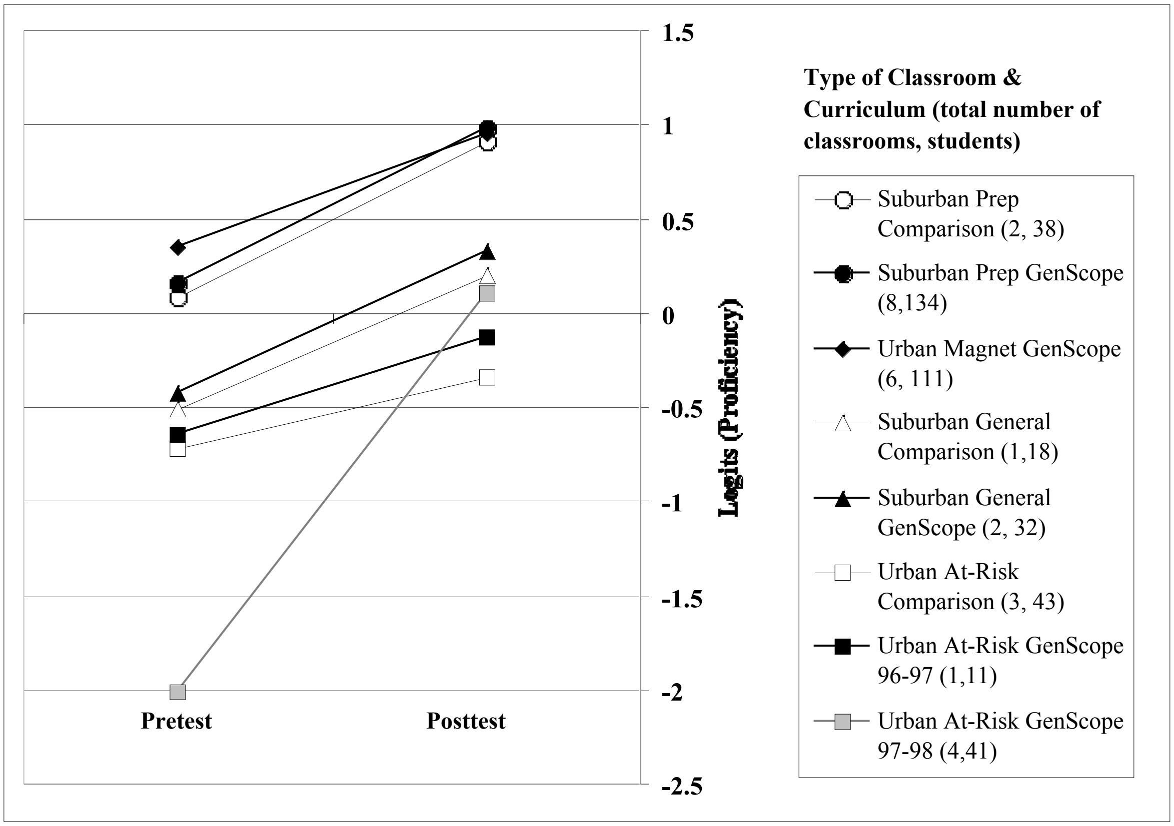

Figure 2: Reasoning gains in GenScope and comparison classrooms, by classroom type.

5. References

Frederiksen, J. R., & Collins, A. (1989). A systems approach to educational testing. *Educational Researcher, 18* (9), 27-32.

Hickey, D. T., Wolfe, E. W., & Kindfield, A. C. H. (1998, April). *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.

Kindfield, A. C. H. (1994). Understanding a basic biological process: Expert and novice models of meiosis. *Science Education*. *78*, 255-283.

Stewart, J. (1988). Potential learning outcomes from solving genetics problems: A typology of problems. *Science Education*. *72*, 237-254.

Stewart, J., & Hafner, R. (1994). Research on problem solving: Genetics. In D. Gabel (Ed.) *Handbook of research on science teaching and learning* (pp. 284-300). New York: Macmillan.