Published in G. Clarebout & J. Elen (Eds.), *Avoiding simplicity, confronting complexity: Advances in studying and designing (computer-based) powerful learning environments* (pp. 117-125). Rotterdam, Netherlands: Sense Publishers.

Patterns in Children's Online Behavior and Scientific Problem-Solving: A Large-N Microgenetic Study

David F. Feldon and Joanna Gilmore
University of South Carolina
127 Wardlaw College
Columbia, SC 29212, USA
feldon@gwm.sc.edu

Abstract. The purpose of this study is to investigate the viability of a computer-based system for collecting and analyzing data from large-N microgenetic studies of scientific reasoning. Most studies of children's scientific problem-solving strategies take place in classroom settings with established norms that homogenize student behavior. Small sample sizes and the limited range of settings limit the generalizability of the conclusions that can be drawn. The current study examined the problem-solving behaviors of 154 children using web-based science simulations. Of the participants who conducted genuine attempts using the simulations, 52 of the participants were members of a sixth grade science class that used the simulations as part of a curriculum unit on infectious disease and epidemiology while 42 of the participants were free-choice users of an informal science website. The data collected through the site and analyzed offline supported the hypothesis that previously reported patterns of scientific problem-solving do not represent the full range of scientific reasoning strategies that participants used when outside the classroom setting. Implications for the continued development of automated analyses of microgenetic studies and future research on children's scientific thinking are discussed.

The development of scientific reasoning in children has major implications for science education and, by extension, the public's understanding and production of scientific knowledge. In instructional contexts, reliable assessment of scientific reasoning skills is very difficult because children's strategy use is highly variable, both within and across individuals (Kuhn & Phelps, 1982; Schauble, 1990; Shavelson, Baxter, & Gao, 1993). Thus, it is particularly important to understand the processes by which scientific reasoning develops in order to design effective interventions (Chen & Klahr, 1999).

Past studies indicate that students approach scientific experimentation from different perspectives that govern their strategy selection, metacognition, and personal criteria for successful completion of presented tasks (Zimmerman, 2000). For example, Klahr and Dunbar (1988) identify two classes of problem-solvers—theorists and experimenters. Theorists develop a hypothetical model of a phenomenon and design tests to analyze the suitability of their model. In contrast, experimenters manipulate values for available variables and subsequently generate hypotheses that can account for their findings. In an alternative description, Schauble, Klopfer, and Raghavan (1991) characterize children's experimentation approaches as representing either engineering goals or scientific goals. Those students with engineering goals manipulate independent variables in order to generate a specific outcome (i.e. value for the dependent variable), such as designing the fastest possible vehicle in the context provided. Scientific goals, however, entail an attempt to understand the causal relationship between independent and dependent variables. Both dichotomies represent differences in fundamental approaches to scientific problem-solving that characterize students' understandings of what "doing science" means. Further, they both draw an important distinction between students who attempt to test a causal model and those whose first priority is to generate experimental outcomes.

Microgenetic approaches to the study of scientific strategy development and understanding typically provide the greatest insights into students' problem-solving processes (Kuhn, 1995). Microgenetic studies entail high frequency qualitative and quantitative observations of a change process in very short time intervals for relatively small numbers of participants (Siegler & Crowley, 1991). The power of the microgenetic method for the investigation of scientific reasoning lies in its ability to detect subtle and transient changes that may not be evident in process outcomes.

However, microgenetic methods have two weaknesses that limit the generalizability of their findings. First, sample sizes are quite small. Second, the collection of both qualitative and quantitative data usually requires the physical presence of the researcher in a setting that facilitates the systematic observation and recording of subjects during a pre-defined task. These constraints, in conjunction with other practical considerations, most often lead researchers to study children's scientific problem-solving within the context of the school classroom (but see Gleason & Schauble, 2000). Within this context, various empirical findings have been replicated (e.g., Kuhn & Phelps, 1982; Schauble, 1990). However, little is known about the respective impacts of the context and the constraints that the setting may place on the range of individual differences among participants on patterns of behavior (Klahr & Simon, 2001).

The current study uses an informal, web-based environment to examine children's scientific problem-solving behavior through interactive simulations of infectious disease spread. A subset of the participants were students in a sixth grade science class that was studying infectious disease. The remainder of the participants varied widely in geographic location and age, and used the simulations without instructions to do so after school hours or on weekends as an independent activity.

The study addresses two research questions:

1. Can computer-based data capture and analysis overcome the limitations of the microgenetic approach to studying scientific reasoning while preserving the richness and complexity of trends in iterative problem-solving data?
2. Does the context in which students engage in scientific inquiry affect their problem-solving strategies?

It is expected that this large-scale microgenetic study will preserve the richness and complexity typical of smaller microgenetic studies. However, the use of algorithms to analyze patterns in performance data will allow researchers to inquire meaningfully into larger populations. Further, it is expected that this geographically distributed sample will demonstrate differences in problem-solving approaches when compared with the performance of participants in a classroom setting. Specifically, non-classroom participants are expected to adhere less to the models of reasoning described in previous research (e.g., Klahr & Dunbar, 1988; Schauble, et al., 1991).

Methodology

Participants. Participants in this study consisted of 154 children (mean age = 13.5; SD = 4.2) who were participating in an online community, Whyville.net. Approximately 31% of the participants were students in a sixth grade science class (mean age = 11.5 years; SD = 0.8) that was involved in other aspects of a larger project, of which the study reported here was one aspect. Their participation in the problem-solving task was a required aspect of their coursework. However, all other participants performed the task without any external direction to do so.

Online context of the Study. Whyville.net is a virtual community of 1.2 million users that can support up to 4000 simultaneous participants at any one time. There is no fee for membership or use, so the population represents a wide cross-section of internet users. The purpose of the site is to provide children who are interested in science an informal virtual environment in which to interact, learn, and play. Users are represented visually on the site through avatars of their own design that are able to "chat" in real time, communicate via internal email, and engage in games and science simulations. The current study coincided with the release of a virtual disease, Whypox, into the environment that altered the appearance of infected avatars (they had "spots") and interfered with normal chatting functions (typed text was replaced sporadically with the word "achoo"). Information about the spread of Whypox on the site was available at a pre-existing Center for Disease Control (CDC) within Whyville. On the CDC page were two simulations that modeled different aspects of disease transmission and epidemiology.

Each simulation also provided a link to a public threaded discussion (i.e., BBS) to post results and conclusions among interested members of the community.

Classroom context. As an aspect of their science class, students participated in Whyville.net activities throughout the academic year. The release of Whypox coincided with week 5 of a 10-week instructional unit on infectious diseases. The instruction emphasized principles of bacterial and viral biology that explained the processes and symptoms of infection. However, it did not include explicit instruction on scientific problem-solving or experimentation strategies. On two days, students were directed to use the online simulations during class time. The instructions provided were minimal. They specified only that over the course of many trials, they should attempt to determine the rules controlling the simulation and test their conclusions by comparing the similarity of their predictions to the outcomes they generated.

Data Collection. Problem-solving data was collected transparently by the server that hosts Whyville.net during participants' interactions with the simulation exercises. In addition to the predicted and actual values of the variables in the simulations (described below), the computer-generated record included user identification, a time/date stamp, and elapsed time between iterations of the simulation. Qualitative data in the dataset included users' optional explanations for their hypotheses and summative comments on their conclusions in the linked BBS.

Apparatus. Participants controlled the values of the independent variables in the simulations using sliders in preset ranges (see Figure 1).

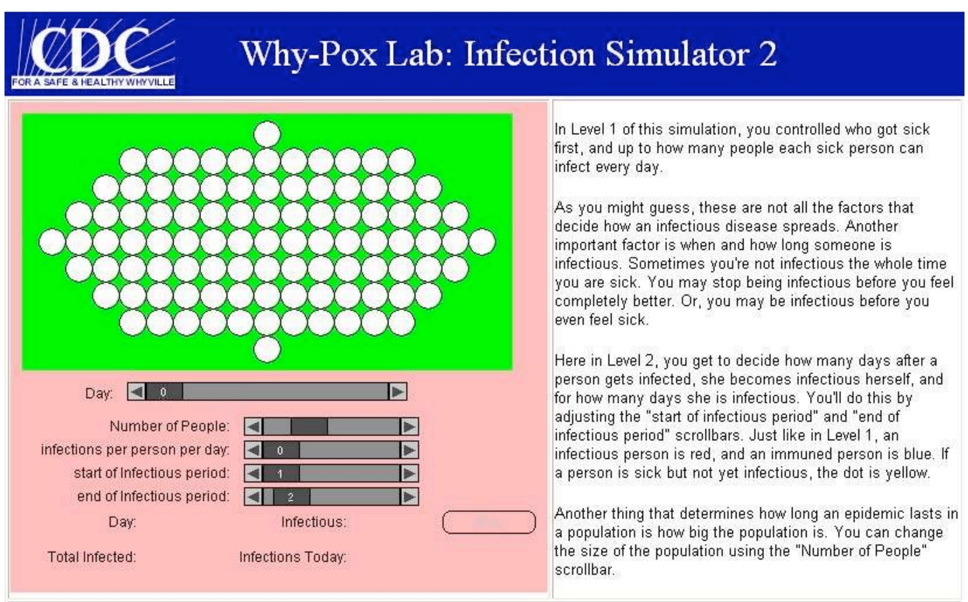

Figure 1. *Screenshots from a typical simulation interface.*

Once participants were satisfied with their settings, they clicked a "run" button. At that point, they were prompted to enter their predicted (i.e. hypothesized) values via slider for each of four dependent variables. They were also prompted to enter text explaining their reasoning (see Figure 2). After entering their responses, participants clicked the "submit" button to observe the outcomes emerging over time and compare the results to their predictions (see Figure 3). The simulations did not constrain the number of trials that a participant could run.

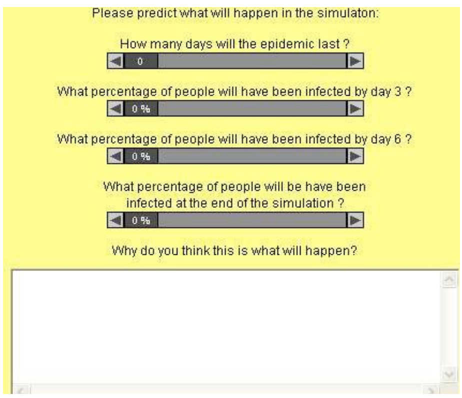

Figure 2.

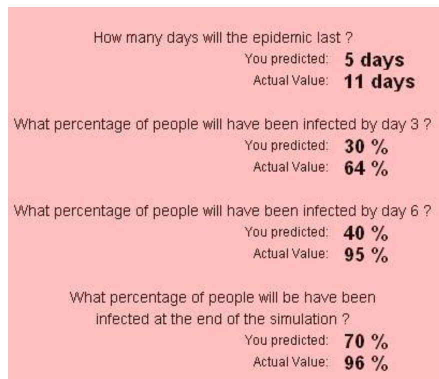

Figure 3.

Data Analysis. The general approach of the current study is intended to reduce the amount of labor-intensive qualitative analysis typically associated with microgenetic studies in order to facilitate investigations of larger, more widely geographically distributed samples. However, the quantitative analyses must converge with available qualitative data to support the validity of this strategy (Creswell, 2003). Thus, after quantitative analysis, available qualitative data was matched to quantitative outcomes to determine congruence.

Quantitative analysis. To analyze such a large volume of data, several algorithms were developed to identify relevant trends in performance. These routines represented three priorities for quantitative analysis of the data. The first priority was that the data represent only valid attempts at using the simulation. Preliminary review of the data suggested that a number of participants tried only to match the predictions and outcomes as precisely as possible without engaging in scientific reasoning. They held the values of the independent variables constant and entered the same values for their predictions that were displayed as outcomes in the previous trial. Others simply "clicked through" the simulation without manipulating independent variables or entering any predicted values. Two algorithms identified problematic data of this nature. They are presented in Table 1. Iterations displaying either of the behaviors described above were removed from the larger data set. In some cases, it eliminated all data from an individual participant.

Table 1

Algorithm and Description of Data Analysis Priorities

Priority	Procedure	Algorithm	Decision
1. Eliminate invalid attempts at using the simulation	1. Identify data in which student's copied the outcomes of the previous simulation trial. 2. Identify data in which students did not change any independent variables from the default of zero.	1. $x_{t+1} - x_t = 0$ 2. $\Sigma y = 0$	1. eliminate trial from further analysis 2. eliminate trial from further analysis
2. Assess changes in the accuracy of participants' predictions over time	Obtain the slope of correlations between predicted and actual values. The slope of the correlations indicates the direction and magnitude of the changes in the accuracy of participant's predictions.	$m=(y-b)/x$ x = trial #; y = accuracy correlation	If the slope of the correlations between predicted and actual values is: - positive, then the participant is classified as having displayed increasing accuracy in prediction. - less than or equal to zero, then the participant is not classified as making increasingly accurate predictions.
3. Assess the robustness of participants' mental models	The correlations computed between predicted and actual outcomes in priority 2 are transformed using Fisher's Z' to ensure a normal distribution. Values of the independent variables are transformed into z-scores. The slope of the regression line between the transformed correlations and the transformed values of each independent variable is computed. This value provides evidence of the robustness of the participants' mental models over time.	$p(t\text{-value}) < 0.05$	If the slope of the regression line between the transformed correlations and the transformed values of the independent variable: - is not significantly different from zero, then the participant is classified as having a robust mental model with respect to the independent variable in question. - is significantly different from zero, then the participant is classified as having a non-robust mental model with

			respect to the independent variable.

The second priority in data analysis was to assess the change in the accuracy of participants' predictions over time. Schunn and Anderson (1999) introduced the notion of correlating numerical hypotheses with the values of the experimental outcomes as a method for determining successful predictions in simulations of scientific problem-solving. They considered Pearson correlations greater than or equal to 0.90 to indicate validation of a participant's theory. As shown in Figure 3, participants in the current study were provided only with the raw values, because it was not expected that participants in this age range would be able to make sense of a correlation as feedback. However, changes in correlations were used as the primary basis of offline data analysis. Increasing correlations across multiple iterations of the simulation for a participant indicated an increase in prediction accuracy. The algorithm developed to identify trends in the accuracy of predictions is provided in Table 1.

Because accuracy correlations are computed without regard for the values of the independent variables, it must be verified that accuracy was not dependent on a set of static conditions. Children's scientific mental models often contain misconceptions that are difficult to detect when activities elicit only a narrow range of responses (Vosniadou, 1994). A fully robust mental model would yield a high level of accuracy across a range of values of the independent variable. Thus, the third priority in the data analysis was to assess the robustness of participants' mental models. The transformations and algorithms used to estimate the robustness of the participant's mental models are described in Table 1.

For the purposes of classification in the current study, participants' data were interpreted according to Schauble, et al.'s (1991) scientific-engineering dichotomy. Participants whose accuracy and robustness both increased were considered to employ a successful scientific model of inquiry, because they sought to attain accuracy across a range of conditions. In contrast, participants whose accuracy increased without a concomitant increase in robustness were considered to employ a successful engineering model. These participants pursued accurate predictions without replicating them across multiple conditions. Because the simulations were self-directed and open-ended, it is assumed that participants who met the criteria described above engaged in these activities until they were satisfied with their outcomes.

Qualitative analysis. After quantitative analysis, subjects who generated numeric data that were strongly indicative of either scientific (i.e., robustness code of 3) or engineering models (i.e., robustness code of 0) of inquiry were matched against the text data they generated in the simulation (see Figure 2) or in subsequent threaded discussion. Statements describing their goals, reasoning, and conclusions were coded independently as representing the scientific model, representing the engineering model, or being uninterpretable. The statements identified as indicative of either model were then compared to the quantitative data from those participants and assessed for congruence.

Results

Of the 94 participants whose performances remained after invalid simulation runs were excluded, changes in accuracy were interpretable for all participants. However, changes in robustness were interpretable for only 83. As expected, of those participants who were classified as having successful science or engineering models, 64.5% were members of the elementary school classes participating in the study (see Table 2). Data also indicated that overall, more school-based participants had greater improvements in accuracy (67.3% vs. 50.0%) and robustness (21.1% vs. 14.3%).

Table 2
Number of Participants by Accuracy Change Estimate, Robustness Change Estimate, and Educational Context

	Robustness						
Accuracy	Model Not Robust	Model Robust for 1 I.V.	Model Robust for 2 I.V.'s	Model Robust for 3 I.V.'s	Too little change in accuracy to determine**	Too little change in I.V. to determine**	Too little data to determine**
Not a Student							
Does not increase over time	15	1	0	3	0	1	1
Increases over time	16*	1	1*	0*	0	1	2
Student							
Does not increase over time	8	4	1	0	1	3	0
Increases over time	27*	2	4*	0*	2	0	0

*Subjects in these cells were identified for qualitative analysis, because they represent successful engineering (i.e., no increase for any I.V.) and scientific (i.e., increases for all I.V.'s) models of reasoning.

**Correlations incalculable due to insufficient variation or insufficient data.

Analysis of the qualitative data was limited due to a paucity of usable responses. Of the successful scientific and engineering model participants, only one in each category provided meaningful text. Both participants were not members of the classroom group, but each set of comments was clearly representative of the model indicated by the quantitative data. The participant classified as using a scientific model referred repeatedly to "surprising" results relative to her theory and asserted the need to "test my theory" with additional trials. The participant classified as using an engineering model articulated post hoc theories to explain the data that was produced (e.g., "It took alot longer to start off, then skyrocketed. This is because the first people weren't infecting people for 3 days. This infections was very evry efficient, as it infected 100% of people" [sic]).

Discussion

The results of the current study show initial promise for a computer-based, large-n microgenetic method. It is possible that providing incentives to enhance persistence in the simulation tasks and increase the number and specificity of comments would improve the overall performance of the system by increasing the quality of the data. The differences observed in participants' performance patterns suggest that the setting may play a role in the strategies used for scientific problem-solving.

Conclusion

The findings from the current study suggest that the internet is a viable means to collect and analyze rich data from problem-solving processes. Further, the differences in strategy that were attained indicate that further research is necessary to understand the scientific reasoning strategies that children use in a variety of settings.

Acknowledgments

The work reported in this paper is supported in part by a grant of the National Science Foundation (NSF-0411814) to Yasmin Kafai at the University of California at Los Angeles. The views expressed are those of the author and do not necessarily represent the views of the supporting funding agency or the University of South Carolina.

References

Chen, Z., & Klahr, D. (1999). All other things being equal: Acquisition and transfer of the control of variables strategy. *Child Development, 70(5),* 1098-1120.

Creswell, J. W. (2003). *Research design: Qualitative, quantitative, and mixed methods approaches* (2nd ed.). Thousand Oaks, CA: SAGE Publications.

Gleason, M. E., & Schauble, L. (2000). Parents' assistance of their children's scientific reasoning. *Cognition and Instruction, 17(4),* 343-378.

Klahr, D., & Dunbar, K. (1988). Dual space search during scientific reasoning. *Cognitive Science, 12,* 1-55.

Klahr, D., & Simon, H. A. (2001). What have psychologists (and others) discovered about the process of scientific discovery? *Current Directions in Psychological Science, 10(3),* 75-79.

Kuhn, D. (1995). Microgenetic study of change: What has it told us? *Psychological Science, 6*, 133–139.

Kuhn, D., & Phelps, E. (1982). The development of problem-solving strategies. In H. Reese (Ed.), *Advances in child development and behavior* (pp. 2-44). New York: Academic.

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology, 32,* 102-119.

Schauble, L., Klopfer, L. E., & Raghavan, K. (1991). Students' transition from an engineering model to a science model of experimentation. *Journal of Research in Science Teaching*, *28*, 859-882.

Schunn, C. D., & Anderson, J. R. (1999). The generality/specificity of expertise in scientific reasoning. *Cognitive Science, 23(3),* 337-370.

Shavelson, R. J., Baxter, G. P., & Gao, X. (1993). Sampling variability of performance assessments. *Journal of Educational Measurement*, *30*(3), 215-232.

Siegler, R. S., & Crowley, K. (1991). The microgenetic method: A direct means for studying cognitive development. *American Psychologist, 46(6),* 606-620.

Vosniadou, S. (1994). Capturing and modeling the process of conceptual change. *Learning and Instruction, 4,* 45-69.

Zimmerman, C. (2000). The development of scientific reasoning skills. *Developmental Review, 20,* 99-149.