Investigating Students' Collaborative Scientific Reasoning During a Natural Selection Investigation

Eleni A. Kyza
School of Education and Social Policy, 2120 Campus Drive, Northwestern University, Evanston, IL 60208
Tel: 847.491.3585, Fax: 208.247.1506
Email: ekyza@northwestern.edu

Abstract: Recent calls to reform science education propose changes in the content and structure of the learning of science. At the classroom level, among others, these calls emphasize investigations that parallel the nature of scientific work, and advocate the value of peer collaboration. Many existing studies of individual students' scientific reasoning claim that middle-school students do poorly, when asked to explain complex problems and provide evidence in support of their explanations. This study investigates how middle-school students, working in academically homogeneous dyads, investigate an extended, data-rich problem. Findings suggest that middle-school students can collaboratively construct evidence-based explanations, but that they still face important challenges that need further support. Both of these points are discussed in more detail in the paper.

Introduction

In recent years, science education reform efforts have advocated changes in the learning of science to reflect important elements from the practices of actual scientists (NRC, 1996). Among other things, such practices should help students understand the nature of science by providing them with extended problem solving opportunities. In these, students would have to take control of their own learning and would need to engage in practices that resemble the work of real-world scientists, such as making sense of multiple sources of data (Blumenfeld et al., 1991). Peer collaboration has also been advocated as a support to these processes, with several researchers arguing that collaboration can help foster collective and individual learning (Barron, 2000; Okada & Simon, 1997). Currently we do not have many rich accounts of how students reason while engaging in authentic inquiry. Understanding how students reason collaboratively in complex investigations can help us ground discussions and guide further design efforts. The goal of the present study was to characterize middle-school students' scientific reasoning, in particular the interaction of students' theories and interpretation of the data in guiding the problem solving activity and the formulation of evidence-based explanations. The coordination of theory and evidence is considered one of the most crucial aspects of doing science (Klahr & Dunbar, 1988; Kuhn, Amsel, & O'Loughlin, 1988) and is an interesting topic to investigate since there are conflicting reports regarding students' ability to do this. The results of this study can thus provide further insights into how students coordinate theory and evidence and inform our understanding of how the educational community could address potential challenges.

Theoretical Framework

Students' reasoning, and in particular scientific reasoning is a complex topic to study. This research looks at one aspect of it: that of coordinating theory with evidence. I define the coordination of theory and evidence as the ability to engage in scientific problem solving using one's theories to guide the discovery and interpretation of evidence, but without allowing a strong belief in the theory distort the examination of the data. The ability to maintain one's theory distinct from the interpretation of the data is proposed to be at the crux of theory-evidence coordination (Kuhn et al., 1988). Nonetheless, the literature has been divided on what parts of this process children and adolescents can do and by when. Klahr & Simon (1999) have called this "the paradox of children's thinking and its development". For instance, in a long line of laboratory-based research, Kuhn et al. argue that middle-school students cannot coordinate between theory and evidence. Several other researchers have argued that the fact that many students were not able to coordinate theory and evidence does not mean that they are not truly able to do so, providing evidence in support of the argument that individual children can reason scientifically from a young age (Koslowski, 1996; Leach, 1999; Samarapungavan, 1992). Finally, research in students' collaborative reasoning has shown that same-ability students working together perform better than if they were working alone; this benefit was measured at both the individual and the collective level (Barron, 2002).

The present study was undertook in the context of a larger research project, which examined the role of software scaffolding in supporting middle-school students' scientific reasoning. Our assumptions were that students have the ability to coordinate theory and evidence but that they face many challenges in trying to do so, and part of our work was to understand more about these challenges. The theoretical framework of this study is grounded in Simon and Lea's (1974) idea of a dual space model of scientific discovery. In this model, scientific discovery is seen as one example of problem solving, involving search and reasoning in two distinct spaces: the hypothesis and the experiment space. This model was later expanded by Klahr & Dunbar (1988) who described scientific discovery as the process of coordinating the two spaces. Klahr & Dunbar suggested that scientific discovery is difficult because the two problem spaces have different representations, which, in turn, require different search strategies. The reasoner can coordinate the two spaces through evidence evaluation (Klahr, 2000) using theories to constrain experimentation and the results of the search in the experiment space to generate new theories. In this study, I used this model of the dual space search and developed analytic categories to help specify what the evidence evaluation space would look like in the context of the students' inquiry. I have also replaced the terms "experiment space" with "inquiry space", and "evidence evaluation space" with "coordination space".

Methods

Context

I tested the assumption that many middle-school students are able to coordinate theory and evidence but to various degrees, in the context of the enactment of a Center for Learning Technologies in Urban Schools (LeTUS) evolutionary biology curriculum, the Struggle for Survival (Carney et al., 2002). Most of the data were collected during the culminating task of this curriculum, the Galapagos Finches investigation [GF] (Tabak, Sandoval, Reiser, & Steinmuller, 2000), which is a computer-based natural selection investigation. In the GF investigation students investigate the reasons leading to the death of many finches on the Galapagos Island of Daphne Major during the late 1970's. Students can browse the GF database and examine finch records for a period of years, including measurements of the finches' physical characteristics, such as beak or wing sizes, and finch behaviours such as foraging and mating. Students can ask questions of the data, by selecting variables such as season, finch characteristic, and gender; as a result to these queries, the program generates graphs that are permanently stored in the Finch Data log section of the GF. With the help of the GF software, the students collect data to support their hypothesis on why many finches died and why some survived during the crisis years on Daphne Major. This is a valuable context to study because the data can support several different hypotheses. This presents the opportunity to study how students try to coordinate theory and evidence during a complex investigation.

Participants

I gathered data from six pairs of middle school students as they worked with the GF investigation. The pairs came from two seventh-grade classes taught by the same teacher. The school was located in a suburb of a large mid-western city and had a mixed student population. The activities that students engaged with, the time they had to complete their investigation, and opportunities to interact with the teacher were comparable across classes. Based on the teacher's assessment of his students' overall academic achievement in science, we selected homogenous pairs, representing different academic abilities to observe the potential range of inquiry activities and discussions.

Data Sources

The data for the current analysis come from the videotaped interactions of the six pairs of students as they worked on the computer and daily records of the groups' artifacts during the GF investigation. We closely followed three pairs of students in each class and collected data from all collaborative investigation sessions, gathering data from approximately eight 45-minute periods for each group. At the end of the unit we conducted exit interviews with each pair of students, and interviewed the teacher.

Data Analysis

Students' discourse and videotaped interactions were transcribed verbatim. The transcription process was contextual (Lemke, 1998), meaning that transcriptions included all of the groups' actions, relating to their investigation, on or off the computer, including gestures. The researcher and an undergraduate coded all of the data together. The primary goal of the data analysis was to understand the students' process and extent of theory-evidence coordination in the context of a classroom-based complex inquiry investigation. To operationalize this we looked at the students' investigation pattern in the hypothesis, inquiry, and coordination spaces, whether students kept their theories distinct from their interpretations of the data, and the sufficiency of the evidence the groups provided in support of their theories. We coded the data for episodes in which students were engaging in processes in each problem space. We also coded students' evolving theories for the four components of a natural selection explanation and the evidence they provided to support them using a rubric developed by Sandoval (2003). These four components are the concepts of environmental pressure, the selective pressure on the population, the existence of a differential trait, accounting for the differential survival in a population, and differential fitness, that is the causal mechanism in place.

Findings

The underlying assumption of this study was that if students were successfully coordinating theory and evidence they would exhibit different patterns of activity in the hypothesis, inquiry, and coordination spaces, as compared to the groups who were not. In particular, we would expect to see the groups who were successfully coordinating theory and evidence to frequently move between spaces, show sustained patterns of activity in each space until the last stages of the investigation, and use the activity in the coordination space to inform their activity in the hypothesis and inquiry spaces. The more successful groups would also keep their theories distinct from their interpretation of the data, and thus would not engage in biased reasoning. Finally, we would expect that these students' theories would show an evolving understanding of important epistemic domain-specific distinctions and that these theories would be supported by the data. Findings are presented in three sections: in the first section, I explore the students' activity patterns in the three spaces. The second section addresses the issue of whether the groups exhibited confirmation bias. The last section discusses the evidentiary support for the groups' theories.

Activity Pattern in the Hypothesis, Inquiry, and Coordination Spaces

All groups attempted to coordinate theory and evidence but to different degrees. I use investigation process maps (IPMs) from two of the groups to illustrate this range of activity: one from a group who were coordinating their theories and evidence well and one from a group who were not. The IPM of Group 1 (Figure 1) shows a sustained pattern of activity in each space throughout the investigation and frequent interaction between the spaces. This pattern is consistent with the pattern we would expect from a group who were successfully coordinating theory and evidence. Looking at the activity within each space we find that Group 1 frequently revisited the hypothesis space, generating and revising hypotheses. The search in the hypothesis space was informed by students' work in the inquiry space. The group was systematic in querying the data and was careful to triangulate their interpretations by looking at a variety of data sources. The students' activities were primarily theory-driven but students appropriately considered data that went against their current theories, resulting in a reconsideration of existing theories and the development of new ones.

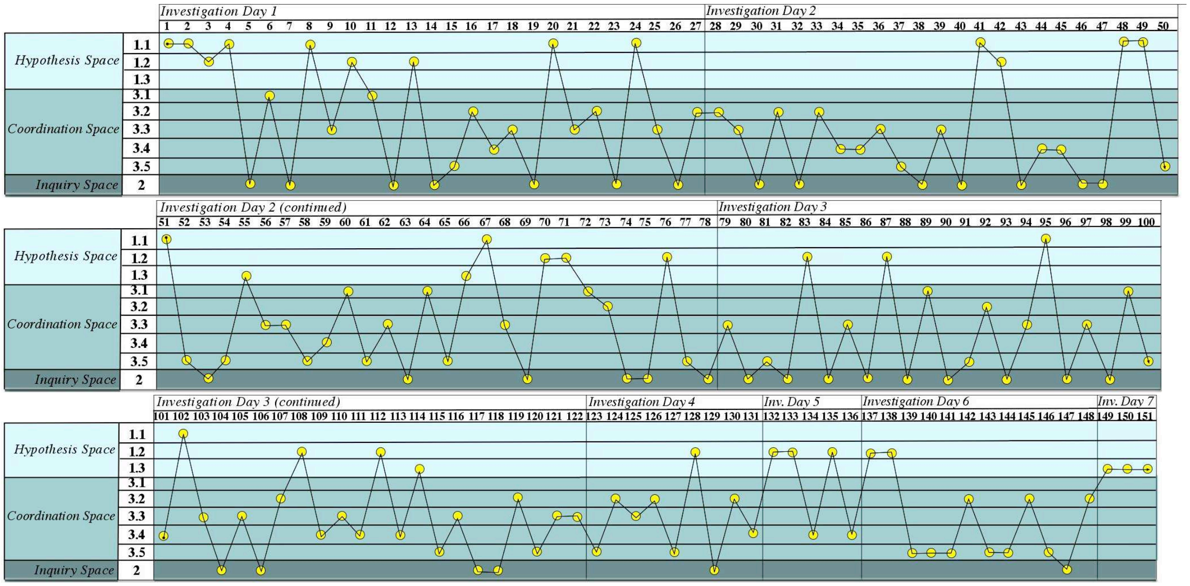

Figure 1. Group 1 Investigation Process Map (IPM1)

In contrast, Group 2 pursued the testing of their theories much less than Group 1, as suggested by the activities coded as attempts to coordinate theory and evidence. From their IPM (Figure 2) we see that this group generated all of their new hypotheses at the very beginning of their investigation, did not generate any other new hypotheses, and revised their hypothesis on only two more occasions. Even so, our examination of the revisions indicated that students extended their hypothesis a little but not to the extent that it significantly expanded their theory or impacted their subsequent investigation. In their activity in the inquiry space, the students' discourse showed that they were aware of their theory as they were interpreting their data but that they did not use the theory to constrain their searches. Rather, students ignored a significant portion of the data since they did not query the database to compare the finches' physical characteristics. We saw several instances of biased interpretations of the data with this group. Students rarely expressed doubts about the plausibility of their theory and almost always concluded that they had enough evidence to support their hypothesis.

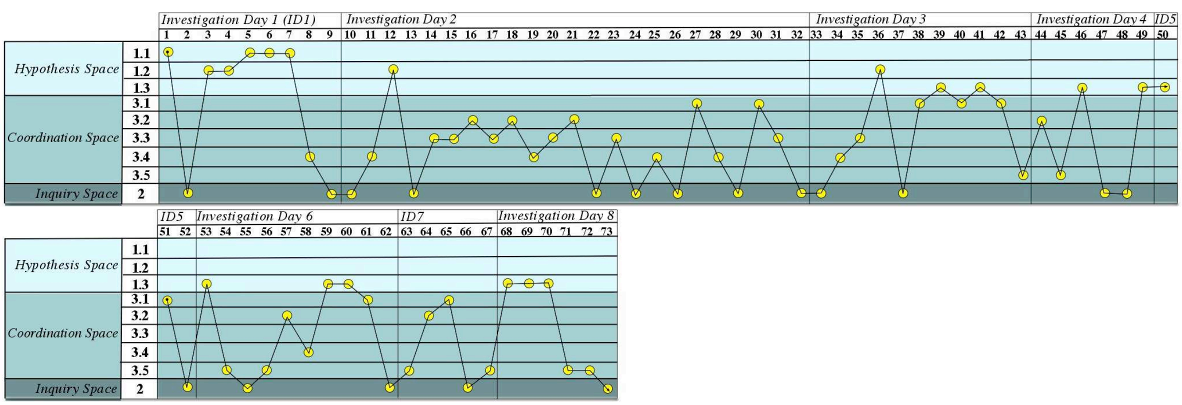

Figure 2. Group 2 Investigation Process Map

On the other hand, we observed other groups who did not use their hypotheses to guide their search in the experiment space. Even though these groups did engage in activities in the coordination space, these activities did not help them as they were not able to use them in deciding which data would be most informative to examine. These students came up with several hypotheses but their pattern of activity suggests that they were not metacognitively aware as to how they could use their theories to guide their experimentation: for instance, the students might have articulated one hypothesis but queried the data for another, or abandoned a hypothesis, without really having explored it.

Challenges in Students' Coordination of Theory and Evidence

Our analysis suggests that some middle-school students are able to engage in data interpretation without exhibiting confirmation bias. The following is such an excerpt from one group's discourse. In this, Group 1 is querying the Finch database for the finches' weight in the wet 77 season, when they find evidence disconfirming their current hypothesis that the lighter finches survived.

1 Lydia: [*Specifying the finch population comparison.*] Okay, we want live to dead?
2 Lilly: Try dead. Try the weight of the people, of the finches who died.
3 *Looking at the results of the comparison in the GF data log:*
4 Lilly: **Only two died?**
5 Lydia: **And they were light.** [*They are referring to the finches' weight*]
6 Lilly: **That's weird.** How many died in [*season*] 77? Yeah, see 77 is when they all died and they were the heavier ones.
7 Lydia: But look at the wet. [*season*]
8 Lilly: The lighter ones died. **That's odd. It's like the normal [*average]* for the living ones was higher so**…who knows…all right wet 77.
9 Lydia: Okay, look all died in wet, not dry.
10 Lilly: Yeah, I know.

The results of the students' query (lines 4-5) contradict their current hypothesis since they show that the finches which died in wet 77 were lighter than the finches which had survived. If students were interpreting these data with a confirmation bias, they could have chosen to ignore or reject these results, since they were only valid for two finches. Nonetheless, as the steps they next took show, this group not only acknowledged the discrepancy but they questioned whether they had queried the data for the wrong season. Following this, students repeated the query only to get the same results. These findings led them to the formulation and testing of several new hypotheses (e.g. the finches which died were lighter because they were fledglings), by constructing queries of the data that could help them explain the discrepant phenomena.

Biased Interpretation of the Data

The espoused hypotheses were at times detrimental to some groups' interpretation of the data. We saw several data misinterpretations with these groups, while at other times they failed to notice data that could have been helpful in solving the problem. The following excerpt is one of these cases. This group's (Group 2) current hypothesis is that some of the finches survived because they were smart and flew away, a hypothesis not supported by the data in the GF database.

1 Researcher: So what is your explanation for why some finches survived?
2 Tom: Cause some finches had the beaks to dig into the ground and some were smart to fly away.
3 Researcher: Do you have evidence for that?
4 Tom: Yeah, Because in our graphs it showed that...
5 Researcher: Where does it show that they were smart and flew away? Show me the evidence for that.
6 Tom: See here, it's 60 (pointing to a data page with a pie graph).
7 Miles: But there, it's only 32.
8 Tom: And then for the dry season, there were…
9 Miles: 32...38
10 Tom: So we know that some flew away.

In line 2, Tom states that some finches were smart and flew away to avoid death. When prompted to support this with evidence, he interprets the difference in the number of finches from one season to the next as evidence that the finches must have flown away, even though this response can not be supported using the data in the GF database.

Considering Competing and Alternative Explanations of the Data

During the early stages of scientific inquiry one may contemplate several competing hypotheses, while towards the later stages one needs to look back at the interpreted data to examine if any other possible explanation exists. If students were successfully coordinating theory and evidence we would expect to see evidence for both these things. All groups considered multiple hypotheses during their investigation of the data but to varied degrees. The groups who more successfully coordinated theory and evidence treated these hypotheses as competing ones, whereas other groups examined them one at a time, moving to another one when having trouble to support the first one. We also saw a case of one group whose belief in their hypothesis was so great that they were not willing to examine the data for the plausibility of any other explanation.

The groups who considered competing explanations within the natural selection framework acknowledged the possibility of a link between the finches' physical characteristics and the environmental pressure. This led these groups to question which one of the available four characteristics (weight, leg length, wing length, or beak length) was the one that gave the surviving finches the advantage. We found differences with the two groups who fully considered these. One of the two groups followed a process of elimination, exploring each one of these characteristics until they concluded that the trait was not responsible for the differential survival. The other group found supporting evidence for their initial hypothesis early on in their investigation and, because of this, briefly addressed the ones for which they did not hold a strong belief about. Even though this group queried the data for competing explanations, their belief about the plausibility of their primary hypothesis prohibited them from pursuing them to the point where they would have adequate evidence to refute them. Late in their investigation, when they are confident that the key characteristic is the beak size, students continue to query the data for alternative explanations. The following excerpt is from their fifth investigation day; in this excerpt students are querying the database comparing measurements of the finches' leg size in wet 78, the season after the drought that killed many of the finch population:

1 Alicia: Now let's look at the live ones.
2 *They do another query and move to the data log, presumably looking at the live finches for wet 78.*
3 Alicia: They tend to have longer legs but it doesn't really matter. Let's look at wing length, oh wait let's look at weight actually.

In this excerpt, the group concludes that one trait (e.g. the leg size) is not important for the finches' survival. In this case, and several others, they choose not to explain why the trait does not matter or look for evidence supporting their claim. We found a similar pattern of not fully considering other plausible hypotheses, or not willing to entertain the idea of looking into possible alternative explanations in most other groups.

Students' Theories and Sufficiency of Evidence

We expected that groups who were coordinating theory and evidence would show changes in their understanding of the natural selection phenomenon. Examining the evolution of the groups' theories over the course of their investigation, we see significant changes in their conceptual understanding of the domain. All of the groups proposed simple but plausible theories before they delved into the data, most influenced by the activities which led to the GF investigation. Nonetheless, with one exception, these theories were later abandoned in favor of theories which evolved from an interaction with the data. In all cases, the final theories had more explanatory power than the initial ones, and included important domain-specific distinctions, such as references to an environmental pressure, the relation of that to specific finch traits, and explanations of the finches' differential fitness. Another place where one could look to see if students were coordinating theory and evidence appropriately is the evidentiary basis of their claims. In the best-case scenario this support would be explicitly cited and, at a minimum, it should be warranted by the data students interpreted during their investigation. We found that all groups presented adequate support for the existence of an environmental pressure and provided some evidence describing how this pressure had an effect on the finches' survival. Four of the groups provided sufficient evidence for claiming the existence of a differential trait, and two of the groups provided some evidence. Most groups did not provide good evidence to support the claim for differential fitness. This would require them to show evidence that related to the differential trait, for both the dead and the surviving finch population. Most groups provided evidence only for the surviving finches. Thus, in brief, groups provided evidence in support of their theories to varying degrees and even though they did manage to support their theories with evidence they did not achieve full coordination of theory & evidence.

Discussion

One of the goals of this study was to characterize students' scientific reasoning during an inquiry investigation. Our data show evidence that middle-school students, working collaboratively, *can* coordinate their theories with evidence, even though the extent to which they do varies. The question then becomes what accounts for the differences between the picture depicted in our data and that painted by Kuhn et al.'s research. One answer is that the task and the support provided to the students, as well as the assumptions of the study were different. In contrast to the Kuhn et al's study setup, this study's task setup had students working in groups and allowed room for students' theories and independent exploration of the data in pursuing these theories. The GF investigation environment provided domain-specific support to the students by structuring the task so that their search in the hypothesis and inquiry spaces would focus on domain important aspects, even though students could choose to ignore it. In our analysis of the data we also considered the ecological validity of students' thinking, looking at their investigation as an unfolding story rather than considering their unsuccessful choices as evidence of an inability to coordinate their theories with the evidence. These results support prior research findings that some middle-school students have the ability to develop coherent theories, evaluate data against their theories and provide evidence to support their ideas (Koslowski, 1996; Leach, 1999; Samarapungavan, 1992). At the same time we saw that some groups faced some of the problems that Kuhn et al. documented in their studies, such as confirmation bias and not considering alternative explanations of their data. Other observed problems had to do with understanding how to coordinate search the hypothesis and inquiry space: for instance, using one's hypothesis to guide their selection of queries to run. Even though this study was not designed to address issues of causality, there is good reason to believe that if some of our subjects were able to successfully engage in coordinating theory and evidence, then the problem of not doing so might be caused by reasons other than developmental ability. In fact, these kinds of problems might be caused by several other things, such as students' repertoire of inquiry strategies or gaps in their domain-specific understanding. These problems can be potentially addressed through the appropriate scaffolding.

Implications for Design

The findings suggest that the students we followed needed further support in two categories: a) addressing inquiry challenges, such as biased interpretations of the data and understanding the value of searching for alternative explanations of the data; and, b) how to use their theories to constrain their search in the inquiry space and vice versa. Both issues are quite complicated, especially since in an ill-structured curriculum one cannot anticipate which challenges are going to be most problematic for whom. Addressing these most probably requires more time than the duration of the enactment of any single curriculum, since they touch on students' understanding of the nature of science. For instance, it was extremely difficult to help students understand that they were engaging in a biased interpretation of their data. Confronting students with the fact, or simply asking them probing questions did not seem to work well. Encouraging students to consider alternative explanations did help them re-examine part of their data, but students were reluctant to do so, and could not see the need for it. Following Collin & Ferguson's (1993) discussion of epistemic forms and games, these students seemed not to fully understand both the form of scientific explanations and the game they were supposed to play. It is, thus, my impression that the problem transcends the specific investigation and it is something that cannot be instilled to students; rather they have to come to understand it more deeply through extended experiences. Other researchers have also suggested the value of repeated exposures to complex data coupled by appropriate scaffolding (Chinn & Brewer, 2001). Secondly, some groups were not able to use their theories productively in their search of the inquiry space and vice versa. This might be a product of two things: a) the theories were not at a level that could guide their inquiry, in which case students might need support in formulating better theories -one such kind of support could be exposure to more domain-specific core ideas; b) students were not explicitly aware of their theories and past interpretations at all times, especially when they interpreted new data. One possible way to assist with these processes could be through metacognitive facilitation, providing students with the means to keep theories and data interpretations in the working memory and encouraging them to engage in processes that help the coordination of the two.

Conclusion

The findings of this study argue that, as expected, there is diversity in students' processes and understanding of scientific reasoning. This study was exploratory and only looked at the investigation of six groups of students. Even though further studies will provide more insights into this topic, one important implication is that some students are able to collaboratively coordinate theory and evidence but that they need further support in addressing the documented challenges. This support might be the result of repeated exposure in engaging in solving complex scientific problems. Furthermore, it might be possible to scaffold learning environments and provide cognitive facilitation tools to help students consider the different aspects of scientific reasoning. For instance, such scaffolding might introduce the idea of the scientific enterprise as a search for the falsification of theories rather than their confirmation, and support students' understanding of the game of constructing scientific explanations.

References

AAAS. (1993). *Benchmarks for Science Literacy*. New York: Oxford University Press.

Barron, B. (2000). Problem solving in video-based microworlds: Collaborative and individual outcomes of high achieving sixth grade students. *Journal of Educational Psychology, 92*, 391-398.

Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational Psychologist, 26*, 369-398.

Carney, K., Golan, R., Holum, A., Laczina, E., Mundt-Leimberer, J., Reiser, B. J., et al. (2002). *The Struggle for Survival: an ecosystems and natural selection project*. Evanston, IL: NSF Center for Learning Technologies in Urban Schools (LeTUS).

Chinn, C. A., & Brewer, W. F. (2001). Models of Data: a theory of how people evaluate data. *Cognition and Instruction, 19*(3), 323-393.

Collins, A., & Ferguson, W. (1993). Epistemic forms and epistemic games: Structures and Strategies to guide inquiry. *Educational Psychologist, 28*(1), 25-42.

Klahr, D. (2000). *Exploring Science: The cognition and development of discovery process*. Cambridge, MA: MIT Press.

Klahr, D., & Dunbar, K. (1988). Dual Space search during scientific reasoning. *Cognitive Science, 12*, 1-48.

Klahr, D., & Simon, H. A. (1999). Studies of Scientific Discovery: Complementary approaches and convergent findings. *Psychological Bulletin, 125*(5), 524-543.

Koslowski, B. (1996). *Theory and evidence: the development of scientific reasoning*. Cambridge, Mass.: MIT Press.

Kuhn, D., Amsel, E., & O'Loughlin, M. (1988). *The development of scientific thinking skills*. San Diego: Academic Press.

Leach, J. (1999). Students' understanding of the co-ordination of theory and evidence in science. *International Journal of Science Education, 21*(8), 789-806.

Lemke, J. L. (1998). Analysing Verbal Data: Principles, Methods and Problems. In B. J. Fraser & K. Tobin (Eds.), *International Handbook of Science Education*. Dordrecht:Kluwer.

NRC. (1996). *National science education standards*. Washington, D. C.: National Academy Press.

Okada, T., & Simon, H. A. (1997). Collaborative Discovery in a scientific domain. *Cognitive Science, 21*(2), 109-146.

Samarapungavan. (1992). Children's judgements in theory choice tasks: scientific rationality in childhood. *Cognition, 45*, 1-32.

Sandoval, W. A. (2003). Conceptual and epistemic aspects of students' scientific explanations. *Journal of the Learning Sciences, 12*(1), 5-51.

Simon, H. A., & Lea, G. (1974). Problem solving and rule induction: A unified view. In L. W. Gregg (Ed.), *Knowledge and cognition* (pp. 105-128). Hillsdale, NJ: Erlbaum.

Tabak, I., Sandoval, W. A., Reiser, B. J., & Steinmuller, F. (2000). *The Galapagos Finches*. San Diego, CA: Academic Press.

Acknowledgments

This research was funded by the U.S. National Science Foundation under grants #REC-9980055 to the KDI/ASSESS project, and #REC-9720383 to the Center for Learning Technologies in Urban Schools. The opinions expressed herein are those of the authors and not necessarily of these foundations. I would like to thank Danny Edelson for his invaluable feedback on this work, and Helen Wong for helping with the data analysis.