

Persistence of Inquiry:
Evidence of Complex Reasoning Among Inner City Middle
School Students

Nancy Butler Songer
The University of Michigan
songer@Umich.edu

Paper presented at the American Educational Research Association (AERA)
Annual Meeting, San Diego, CA.

This research is funded by the Interagency Education Research Initiative (IERI), grant no. REC-0089283 and REC-0129331. The opinions, findings and conclusions or recommendations expressed in this paper are those of the author. I thank Michelle Astolfi, Amelia Gotwals, Pier Sun Ho, Anne Huber, Tricia Jones, Scott McDonald, Phil Myers and other members of the BioKIDS research group at the University of Michigan for their support.

Persistence of Inquiry: Evidence of Complex Reasoning Among Inner City Middle School Students

Introduction

Science education research suggests that the development of complex scientific reasoning takes longer than any one curricular intervention, yet surprisingly little research exists on fostering or evaluating complex inquiry thinking over multiple topics or years. While much research in the learning sciences supports the idea that the development of complex thinking takes time and that the development of concepts and thinking skills should be coordinated over multiple units and years, (e.g. Bransford, Brown and Cocking, 2000), few research projects develop and evaluate curricular programs with this idea in mind. National organizations such as the American Association of the Advancement of Science (AAAS) advocate standards-based curricular programs to foster the development of complex reasoning in science, including fostering both the ability to explain individual scientific concepts and the relationships and connections between concepts. Many educational reforms provide strong curricular programs and strong learning outcomes but these programs are nearly always associated with one science topic and one curricular unit.

Even with strong curricular programs, research diagnosing students' scientific inquiry skills reveals that students' understandings are incomplete in many respects (Jeong, Songer and Lee, submitted). While students are able to recognize certain features of inquiry thinking such as the objectivity of data, few students systematically demonstrated the difference between evidence and

explanations, or can generate explanations from data. This research suggests that more comprehensive, scaffolded inquiry-fostering programs are needed to systematically support students' development of complex reasoning skills over a given unit, several units, and several years.

While some curricular programs do not support young students' complex reasoning in science, current research on elementary students' inquiry development suggests that even students as young as first and second grades can develop complex reasoning about science phenomena, provided appropriate guidance and scaffolding of tasks is present (e.g. Metz, 2000). The National Research Council suggest that inquiry programs should "exploit the natural curiosity of children" (NRC, 2000; p. xiii) as children in K-4 are guided to "ask questions about objects, organisms and events in the environment, plan and conduct a simple investigation, use data to construct a reasonable explanation, and employ simple equipment and tools to gather data and extend the senses" (NRC 2000; p. 19). According to the NRC, programs in 5-8th grade should be organized to build on students' evolving questioning, investigation-building, and explanation-building talents first fostered in early elementary grades, yet again, a review of exemplary curricular programs suggests few to no programs systematically build inquiry-fostering. Clearly more research is needed to examine the nature of tasks and guides that can foster such complex reasoning in science with elementary-age children.

What do inquiry-fostering activities look like in classrooms? Contrary to a commonly-held view of classroom inquiry as unstructured open-ended activities, research and policy documents suggest that classroom-based inquiry-fostering activities can take many forms. As outlined by the NRC (2000),

“ Investigations can be highly structured by the teacher so that students proceed toward known outcomes, such as discovering regularities in the movement of pendulums. Or investigations can be free-ranging explorations of unexplained phenomena...The form that inquiry takes depends largely on the educational goals for students, and because these goals are diverse, highly structured and more open-ended inquiries both have their place in science classrooms” (NRC, 2000, p. 10-11).

How is complex reasoning evaluated? The National Research Council (2001) recommends robust assessment instruments that compliment standards-based curricular programs and that focus on the measurement of complex reasoning in science.

“Assessments that resonate with a standards-based reform agenda reflect the complexity of science as a discipline of interconnected ideas and as a way of thinking about the world.” (National Research Council, 2001; p. 12)

Despite the demand for assessment instruments that measure complex reasoning in science, few instruments exist that provide a systematic approach to the evaluation of complex reasoning in science (Mislevy et al, 2002). Many of the current high-stakes national and international science tests emphasize definitions of science concepts and/or fact-based knowledge over items measuring complex reasoning in science, no doubt because of the challenge of developing reliable instruments to systematically evaluate students' inquiry thinking such as the ability to develop explanations from scientific evidence. As high-stakes tests often attempt to match the learning goals of the standards-based reform programs but often fall short, schools must confront a difficult mismatch between the emphasis of the high-stakes tests and the emphasis of the reform-based programs. What is needed is a coordinated, systematic curricular and assessment program that collectively support the thinking and learning goals of

standards-based reform programs, and that has assessment tasks built on design principles intimately aligned with the reform programs.

Failing schools, such as those in many urban school districts, are particularly pressured to perform well on high-stakes tests. A systematic approach to inquiry-fostering curricular activities and assessment might be especially valuable for urban schools, both to provide tangible evidence of student learning trajectories, as well as evidence to evaluate effective reform programs from those that are less effective. The focus of this paper is the description and results of one systematic curricular and assessment program for the development and evaluation of learning among cohorts of 4-7th grade students in a high-poverty urban district.

Longitudinal Evaluation of Complex Science

Research on children's learning recognizes that the development of deep conceptual understandings in science requires the structuring of experiences, including catalysts to encourage curiosity and persistence, and mediation often in the form of scaffolds to guide children's attention to salient features amidst many complexities within natural world reasoning situations (Lee and Songer, in press; Bransford et al, 2000; Vygotsky, 1978). The development of complex inquiry thinking requires both the development of underlying science concepts as well as the development of reasoning skills in that context, such as building explanations from evidence (NRC, 2000).

How should these developing reasoning skills be evaluated? A model longitudinal inquiry assessment program would coordinate **student cognitive models** of how learning occurs and what should be measured, **observations** that

demonstrate learning associated with what should be measured, and **interpretation** to “reason from fallible observations” (Pellegrino, Chudowsky, and Glaser, 2001).

In BioKIDS, these dimensions of our assessment system are developed following guidelines of Evidence Centered Design and in association with Robert Mislevy and Geneva Haertel of the PADI project (see Gotwals and Songer, 2004 for details of this work and collaboration). In our evaluation, we wish to evaluate students’ development of inquiry reasoning skills (e.g. formulating explanations from evidence) and content (e.g. biodiversity), both as they are intertwined (e.g. complex reasoning in science) and when they are separate (e.g. the development of declarative knowledge in biodiversity that does not involve inquiry reasoning). In the past, we believe it has been difficult to assess students’ complex reasoning in a content area because the assessment systems were not designed for these multiple roles. We expect that the BioKIDS/PADI system will allow robust evidence of students’ content knowledge within a scientific discipline (e.g. ecology, weather) as well as robust evidence of students’ complex reasoning within a scientific discipline, e.g. the reasoning skills associated with interpreting species data or building explanations from atmospheric science evidence. At the conclusion of this project we expect to have the following:

- A coordinated set of assessment tasks that provide evidence of students’ development of particular inquiry reasoning skills in several content areas,
- A complimentary set of science activities within units that provide specific scaffolds for the development of complex reasoning in a range of science topics,

- Longitudinal trajectories of evidence on cohorts of inner city students demonstrating beginning, intermediate, and advanced levels of reasoning in science in several content areas.

We hope that this coordinated set of products will provide us with a much clearer view of both what students learn in standards-based reforms, and where the development of complex reasoning falls short of the ideals.

The Development of Scaffolded Inquiry Activities

In *BioKIDS: Kids' Inquiry of Diverse Species* (Songer, 2000), curricular units contain content-specific scaffolds to foster the development of inquiry reasoning skills. In these units, particular inquiry thinking skills such as the development of explanations from evidence are fostered through a carefully scaffolded activity sequence (Lee and Songer, 2004; Songer, 2000; Huber, Songer and Lee, 2003).

One characteristic of the curricular sequence leading to the scaffolding of inquiry and content development is the repeated presence of guided-learning approaches. For example, a central science concept fostered in *BioKIDS* is an understanding of the concept of biodiversity, a definition of which involves several factors on which scientists often disagree. In the *BioKIDS* program, sixth grade students are asked to collect animal species data on a particular area or the schoolyard in preparation for the development of a claim and evidence addressing the question, "Which zone in the schoolyard has the greatest biodiversity?" Scientists might evaluate which zone is most diverse using Simpson's index, $D = 1 - E (n/N)^2$, a formula that represents species evenness taking into account both the total number of animals (abundance) and the

number of different species (richness). While our sixth graders are not taught to use Simpson’s index, our program does encourage students to develop a qualitative understanding of biodiversity that takes into account species abundance and richness. In order to gain this understanding, students work with the concepts of abundance and richness in complimentary ways throughout several activities, and the repeated presence of approaches makes this challenging concept understandable to students. Similarly, a central inquiry concept emphasized is “building explanations from evidence”. As with the biodiversity concept, sixth graders are provided with repeated opportunities to make claims, determine what evidence is salient, and build explanations from data towards a deep understanding of inquiry thinking with biodiversity concepts. Figure 1 presents the scaffolding format used in ten different inquiry activities in the curricular program to guide students in formulating explanations from evidence. Notice the presence of sentence starters, e.g. “I think.....because...”, and direct content prompts, e.g. “How many animals and different kinds of animals were found...”, to guide students’ in selecting relevant evidence for their explanation and in composing a claim.

Question: Which schoolyard zone has the highest biodiversity?

<p><i>Claim</i></p> <p>SENTENCE STARTER→ <i>I think zone _____ has the highest biodiversity because.....</i></p>
<p><i>Data or Evidence</i></p> <ul style="list-style-type: none"> • <i>How many animals and different kinds of animals were found in this zone compared to other zones?</i> • <i>Where were animals found in this zone?</i> • <i>How does this zone support both high abundance and high richness of animals?</i>
<p>CONTENT PROMPTS ^</p>

Figure 1: Curricular Scaffolding For Formulating Explanations from Evidence

This paper presents control and experimental student learning data evident after “the persistence of inquiry”, e.g. the coordination of curricular scaffolds and assessment systems designed to systematically foster and measure inner city students’ inquiry reasoning within three coordinated units in the sixth grade.

Methods

Subjects

The sample consists of 1924 sixth grade students in fifteen high-poverty urban schools containing 94% underrepresented minorities. The students were taught by 21 teachers. All teachers were encouraged to attend professional development workshops supporting the same eight-week, inquiry-fostering curricular program in biodiversity called BioKIDS: Kids’ Inquiry of Diverse Species (Songer, Huber and Lee, 2003). Classes of students were divided into control (N=595) and experimental (N=1329) populations. Control classrooms performed from 0-30% of the curricular activities. Experimental classrooms performed nearly all curricular activities (the average was 95%). All students took identical pre and posttests at the same time points (what would have been the beginning and the end of the curricular intervention).

Instruments

Sixth grade control and experimental students experienced three eight week inquiry-fostering units in: biodiversity, weather, and simple machines. Pre and post assessments were developed in association with each curricular unit. Biodiversity pre and posttests consisted of 19 items worth a total of 37 points.

All tests were developed using components of the PADI assessment system, particularly design patterns and templates (see Gotwals and Songer, 2004). In all three content areas, assessment tasks were developed to evaluate student performance on three areas of inquiry reasoning within the given content area: “formulating scientific explanations from evidence”, “analyzing data”, and “developing hypotheses and predictions”. At the time of this paper, a complete analysis is available for the biodiversity unit pre and posttest data; the data for the weather and simple machines units are still being analyzed.

Results

Student Learning Outcomes

Table 1 and Figure 2 demonstrate control and experimental students' performance on the pre and posttest tasks for the biodiversity unit. Note that control and experimental students are statistically identical on the pretest, and experimental population scores significantly higher on all posttest measures.

Table 1: Comparison of Experimental and Control Schools

	Experimental (N=1329)	Control (N=595)	Significance
Pretest: total	39.4%	39.0%	p=.721
Posttest: total	54.7%	45.8%	p<.001
Pretest: explanations	46.0%	44.6%	p=.248
Posttest: explanations	63.9%	53.5%	p<.001
Pretest: interpreting data	59.7%	56.9%	p=.061
Posttest: interpreting data	74.5%	67.1%	p<.001
Pretest: Hypotheses and Predictions	44.7%	43.6%	p=.405
Posttest: Hypotheses and Predictions	55.6%	47.1%	p<.001

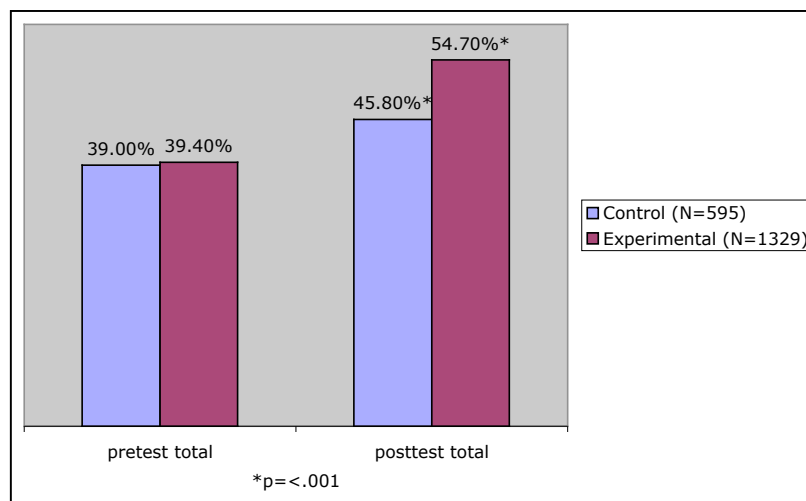


Figure 2: Control and Experimental Performance on Pre and Posttest

Looking within particular areas of inquiry reasoning, our results are similar to the overall trends. Figure 3 demonstrate control and experimental

students' performance on the pre and posttest tasks for the tasks designed to evaluate students' performance on the inquiry reasoning skill, "formulating scientific explanations from evidence". Note once again that control and experimental groups are statistically identical on these pretest measures, but experimental populations demonstrate significantly better on posttest measures.

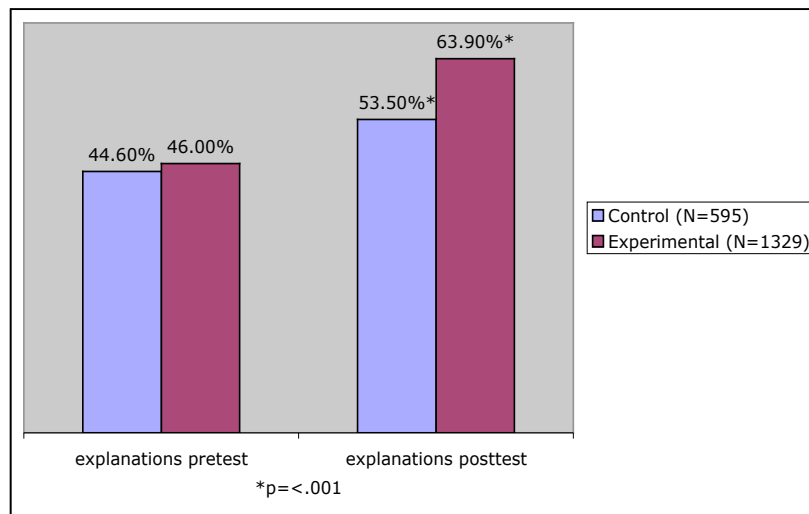


Figure 3: Control and Experimental Performance on Explanations Items, Pre and Posttest

Student Learning Outcomes By Complexity Level

One advantage of assessment tasks coordinated around PADI templates (Riconscente, Mislevy, Hamel and the PADI Research Group, 2004) and content-inquiry matrices (Songer, 2003; Gotwals and Songer, 2004) is the ability to provide evidence of student reasoning at step 1, step 2 and step 3 levels of content and inquiry complexity. Table 2 and Figure 4 present control and experimental students' performance on all types of inquiry reasoning tasks as organized by complexity level (e.g. step 1 simple tasks, step 2 moderate tasks, and step 3 complex tasks).

Table 2: Comparison of Experimental and Control Schools by Complexity Type

	Experimental (N=1329)	Control (N=595)	Significance
Pretest: Step 1, Simple	65.8%	63.6%	p=.101
Posttest: Step 1 Simple	77.7%	73%	p<.001
Pretest: Step 2, Moderate	41.5%	40.4%	p=.361
Posttest: Step 2, Moderate	58.8%	49.1%	p<.001
Pretest: Step 3, Complex	23.2%	24.6%	p=.239
Posttest: Step 3, Complex	43.5%	32.2%	p<.001

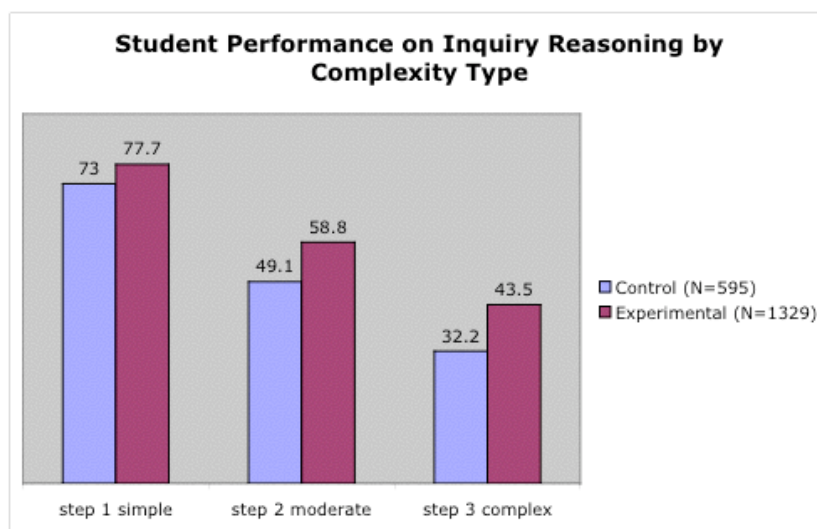


Figure 4: Control and Experimental Performance on Posttest by Complexity Levels

Several important trends are evident in these data. First, while control and experimental groups are statistically identical on pretest measures by complexity level (not shown in this table), experimental groups demonstrate significant improvement on the posttest relative to control students at all three complexity levels. In addition, the percentage gain between control and experimental students is not the same at each complexity level. As the complexity level increases, the difference in gains between the control and the experimental populations also increases. In other words, while control and experimental students can perform more similarly on items measuring declarative of simple levels of inquiry reasoning and scientific knowledge in biodiversity, the experimental students demonstrate a much larger advantage over the control students on the items of highest complexity. (The control/experimental differences on the posttest are 4.7% (step 1 simple), 9.9%, (step 2 moderate) and 11% (step 3 complex).

The trends seen by complexity level overall are also mirrored within each area of inquiry reasoning. Figure 5 presents student performance by complexity level on interpreting data tasks only. Note once again that the differences between control and experimental students are smallest in the step 1 simple items, and larger in the step 2 and step 3 levels.

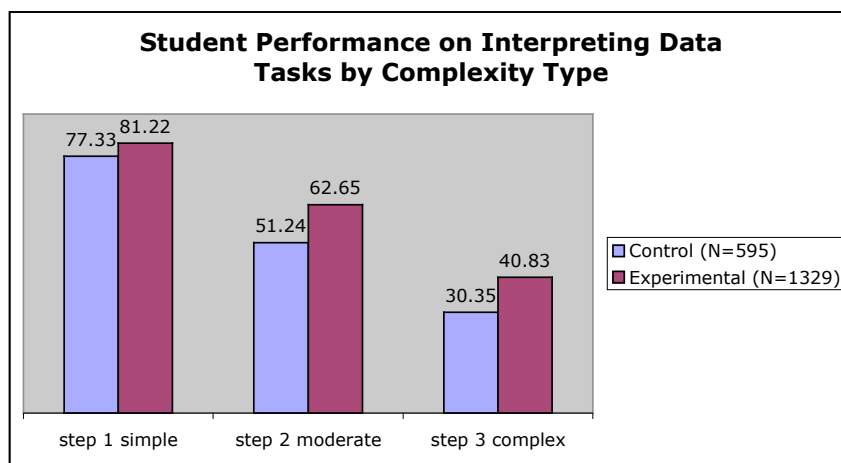


Figure 5: Control and Experimental Performance on Interpreting Data Tasks by Complexity Levels

Conclusions

Analysis of control and experimental student performance relative to complex reasoning in biodiversity associated with curricular units designed to scaffold complex reasoning suggests a correlation between curricular programs that specifically scaffold inquiry reasoning and evidence of complex reasoning in science. While improvements are seen in experimental students' outcomes at all three complexity levels, greatest improvements are seen in the highest complexity levels. This result suggests that curricular programs and assessment systems specifically tailored to foster and measure complex reasoning in science can provide evidence of not just improvements in reasoning, but a great deal of

information on the areas where complex reasoning is not progressing as effectively as others. This kind of data will allow more careful alignment of curricular programs to the goals of complex reasoning in science, as well as detailed information on students' development of complex reasoning over time and topic. Analysis is continuing to evaluate students' performance on these same inquiry reasoning skills associated with curricular programs in weather and simple machines. This research has implications for both the careful design of curricular scaffolds for fostering conceptual development and inquiry thinking, and the potential of systematic assessment systems to provide detailed information on students' development of complex thinking over time.

References

- Bransford, J. D., Brown, A. L., Cocking, R. R. (2000). *How People Learn: Brain, Mind, Experience, and School*. Washington D.C., National Academy Press.
- Huber, A., Songer, N.B. and Lee, S.Y. (2003) BioKIDS: A Curricular Approach to Teaching Biodiversity through Inquiry in Technology-Rich Environments. Paper presented at the annual meeting of the National Association of Research in Science Teaching (NARST), Philadelphia, April.
- Lee, H-S. and Songer, N.B. (in press) Making Authentic Science Accessible to Students. *The International Journal of Science Education*.
- Jeong, H., Songer, N.B., Lee, S.Y. (submitted) Diagnosing Urban Sixth Graders' Inquiry Skills and Understanding. *Science Education*.
- Linn, M.C., Tsuchida, I., Lewis, C. and Songer, N.B. (2000) Beyond Fourth Grade Science: Why do US and Japanese students diverge? *Educational Researcher* 29(3), 4-14.
- Mislevy, B., and the Principled Assessment Design for Inquiry Project (in alphabetical order: Chudowsky, N., Draney, K., Fried, R., Haertel, G., Hamel, L., Kennedy, C., Long, K., Morrison, M., Pena, P., Rosenquist, A., Songer, N., Wenk, A.) (2002) Design Patterns for Assessing Science Inquiry. Technical Report, SRI International.
- Mislevy, R., Steinberg, and Almond, (in press) On the structure of educational assessments. To appear in *Measurement*.

National Research Council (2001) Classroom assessment and the National Science Education Standards. Committee on Classroom Assessment and the National Science Education Standards. Atkin, J.M., Black, P.I., and Coffey, J. (Eds.) Center for Education, Division of Behavioral and Social Sciences and Education. Washington, D.C.: National Academy Press.

National Research Council (2000) Inquiry and the National Science Education Standards. Committee on Development of an Addendum to the National Science Education Standards on Scientific Inquiry. Center for Education, Division of Behavioral and Social Sciences and Education. Washington, D.C.: National Academy Press.

Songer, N.B. and Wenk, A. (2003) Measuring the Development of Complex Reasoning in Science. Paper presented at the annual meeting of the American Educational Research Association, April.

Songer, N.B., Lee, H., and Kam, R. (2002) Technology-rich inquiry science in urban classrooms: what are the barriers to inquiry pedagogy? *Journal of Research in Science Teaching* (39) 2: 128-150.

Songer, N.B., Lee, H. and McDonald, S. (in press) Research towards an expanded understanding of inquiry science beyond one idealized standard. *Science Education*.

Songer, N.B. (2000) BioKIDS: Kids' Inquiry of Diverse Species. Proposal funded by the Interagency Education Research Initiative (IERI), \$5.5 million.

Vygotsky, L. (1978) *Mind in Society: The development of higher psychological processes*. Cambridge, MA: Harvard University Press.