“I’m trying to figure out how we started in science”: Elementary students’ developing scientific thinking

Adrienne Gelpi Lomangino
Susanna Hapgood
Shirley Magnusson
Annamarie Sullivan Palincsar
Nancy DeFrance

Educational Studies
University of Michigan

Paper prepared for the annual meeting of the American Educational Research Association

April 12 – 16, 2004

San Diego
Introduction

Being scientifically literate requires not only substantive understanding of concepts and theories, but syntactic understandings pertaining to the accepted forms and means of generating evidence and proof within the disciplinary community of science (Schwab, 1964). Millar, Lubben, Gott and Duggan (1994) describe these fundamental understandings about the “collection, validation, representation and interpretation of evidence,” as concepts of evidence (Gott & Duggan, 1996, p. 793). These ideas are important not only for students who seek to be scientists, but more broadly for participation in a range of fields. Engaging students in scientific investigation activity has been widely promoted as a means of developing these syntactic skills. Students are believed to learn scientific processes most effectively in an integrated way through whole investigation experiences. However, questions of what students should be and are actually learning through the experiences remain more elusive. Students do not simply shift from lacking these understandings to having them. These understandings develop over years and multiple experiences. What does progress toward deeper understanding look like? This study examines the knowledge and skills that elementary students are developing early on the path to scientific literacy (AAAS, 1993).

The goal of having elementary students develop scientific literacy through experiences conducting investigations might seem infeasible based on previous research findings. Psychological studies of elementary students’ scientific thinking reveal an array of ways in which students’ thinking and activity fall short of scientific experimentation (Bullock & Ziegler, 1999; Klahr, Fay, & Dunbar, 1993; Kuhn, 1989). For example, previous findings indicate that during investigations children in fifth and sixth grade did not record plans or conclusions, attend to the range of available problem space, design trials that tested a single variable, or systematically manipulate parameters (Schauble, 1990; Schauble, Klopfer, & Raghavan, 1991). Once they had generated data, children also had difficulty with interpretation (Schauble, Glaser, Duschl, Schulze, & John, 1995).

Looking past these difficulties, understanding the development of scientific thinking requires taking a broader view of the knowledge and practices that form the foundation for the complex enterprise of experimentation. Lehrer, Schauble, and Petrosino (2001) point out that by focusing attention on hypo-deductive reasoning, researchers have lost sight of aspects of scientific thinking that are fundamental to experimentation, including the central importance of argument, representation, and modeling. In developing an understanding of scientific argumentation, students need to develop facility with inscriptions, which include various means of concretely preserving a view of the world for examination, such as drawings, maps, text, recordings, and models. For example, students must develop the ability to record data in a fashion that enables interpretation (Lehrer & Schauble, 2002). In addition, use of inscriptions is essential to the process of building and refining models, which lies at the core of scientists’ work (Penner, 2001). One purpose of this study is to provide an enriched description of the range of ways in which upper elementary school children engage in scientific investigations.

The analysis of children’s activity in this study utilizes a framework for assessing children’s scientific thinking devised by Millar, Lubben, Gott and Duggan and their colleagues in the United Kingdom. They propose that students must develop both relevant skills and concepts of evidence in order to meaningfully engage in scientific inquiry. In response to concerns about standards in education and as part of an effort to examine student performance following the introduction of a national curriculum in the United Kingdom, they (1994, 1996) outlined a framework defining the knowledge base pertaining to conceptions of evidence. They argued that “practical” or laboratory work in science has “a distinct knowledge base which is connected directly and necessarily with the understanding of scientific evidence” (1996, p. 792).
Developing concepts of evidence

They intended for the construction of such a framework to both clarify the purposes for inquiry-based science instruction and facilitate examination of student progress toward important goals in science education. Although skills, such as how to create a graph or use a particular instrument, are important for effective experimentation, concepts of evidence require a higher level of understanding regarding the scientific argument one is trying to make, as well as broader understandings of validity and reliability of evidence. The framework divides the concepts of evidence into three broad categories, associated with design, measurement, and data handling. Concepts of evidence associated with design include variable identification, fair testing, sample size, and variable types. Those understandings associated with measurement pertain to understanding how relative scale, range of interval, choice of instrument, repeatability, and accuracy influence the validity and reliability of an investigation. With respect to data handling, concepts of evidence involve use of tables, graphing, and seeking patterns in the data. While we recognize that the authors of this framework made no claims to its comprehensiveness, a second purpose of this study is to examine the utility of this framework and the ways in which it might fruitfully be expanded for describing young children’s emerging competencies with scientific investigations.

The process of learning science is viewed within this study as involving a social process of enculturation into the practices and ideas of the scientific community. Knowledge and skills are the tools, the useful means for constructing and extending thought, that students gain expertise using. These practices and ideas include a body of theories, practices, and concepts that have accumulated over time and been accepted within the scientific community (Driver, Newton & Osborne, 2000; Metz, 1995; Rogoff, Baker-Sennett, Lacasa, & Goldsmith, 1995). Through multiple experiences engaging in activities of the discipline in conjunction with others, students develop the tools that will enable them to engage in complex inquiry (Brown, 1997; Brown & Campione, 1994; Hapgood, in press; Lomangino, 2000). Students should therefore not be expected to form this knowledge independently or easily.

Existing research has not captured students’ progression in developing these fundamental understandings. Studies that examine whether students have these understandings or lack them do not shed light on what the path to learning the conceptions involved in scientific investigation looks like. When teachers engage their students in investigation activity, they are not likely to find that students shift from novices to experts. This study examines students’ developing scientific thinking, as demonstrated through their activity during an investigation task.

This study examines students’ scientific thinking after relatively brief instructional experience. This study seeks to take a closer look at students who are just beginning to make sense of the scientific investigation process and seeks to address the following questions.

β What aspects of scientific thinking within the “concepts of evidence” framework do students reveal during an investigation task?
β In what ways does the “concepts of evidence” framework need to be expanded to describe how young inexperienced learners engage in scientific investigations?

Methods

Participants

This study examines activity of 12 pairs of fourth grade students during a clinical interview (described below). Students were drawn from two classrooms in two different schools within a district on the urban fringe of a large city. The student population in the district was approximately 50% African American and 50% European American, and 58% of the students qualified for free or reduced lunch. Pairs of students were selected through teacher nomination based on their perceived abilities to work together and their participation during a brief science unit that preceded the clinical interviews analyzed in this paper.
Related prior experiences. The students had had minimal prior experience conducting scientific investigations. However, during the two weeks preceding the conduct of the interviews, the children had engaged in two investigations of motion, one on an incline and one on a horizontal plane. This instruction was brief, challenging and novel for the students. Half of the students had used materials similar to those used in the clinical interview during their investigation of horizontal motion. The other half had read a text describing a similar investigation and only used materials themselves during their study of motion on an incline.

The Clinical Interview

This study focuses on students’ responses to a modified performance-based assessment, which we are calling a clinical interview, that was conducted as a means of attaining information about students’ abilities to conduct investigation. The clinical interviews engaged students in manipulating materials and collecting data (Magnusson, Templin, & Boyle, 1997).

We considered administering a performance-based assessment (PBA), however, due to the students’ limited experience with 1st-hand investigation and lack of experience with performance-based assessment, we decided to conduct interviews that were a cross between a performance-based assessment and a clinical interview. They were like PBAs in that students worked with the materials to complete an investigation (which has not been common in clinical interviews in science education). They were like clinical interviews in that students were sometimes asked to provide information about their thinking, and they were sometimes supported to conduct the investigation.

Figure 1. Diagram of investigative set-up.

Protocol. The problem presented to students in the clinical interview involved motion on a flat surface. The materials provided to students to investigate the problem included those used for previous classroom investigations of motion on a flat surface, as well as a few additional items used for investigating motion on an incline.

The interview involved presenting children with a design problem that also required modeling of an everyday situation. The rationale for using a design problem was in large part to promote students’ motivation by having a clear goal. A reason for including modeling in the interview task was to reflect an important aspect of scientific practice. The interview involved six phases, which are described below.
The assessment began by presenting students with a problem scenario, shown below. The scenario ended with a question for them to consider.

### Problem Situation
Dante needs to go grocery shopping. He uses his wagon to carry the groceries home. Between his home and the store he has to cross a busy street. When he walks with an empty wagon, he just makes it across the street before the light turns red.

How many groceries can he carry in his wagon and still safely walk across the street?

**Determining the problem.** The next phase of the interview focused on students having to determine the problem. After reading the scenario along with the pair, the interviewer asked students what ideas they needed to think about to answer this question. The interviewer noted whether students referred to mass, force, or speed, as well as other factors.

**Modeling the Problem.** The students then selected materials to use in order to model the problem situation. The provided materials included: a string with a paper clip attached, 5 washers, a calculator, a spring, a cart, a tape measure, and 4 rubber weights. The box of available materials included several items that were used in the students’ prior investigation experiences, but were not necessary for examining the current problem. Students had the option of modeling the situation together or individually, depending on whether or not they had congruent ideas. The interviewer sketched the models.

**Investigation Planning.** After setting up a model, the students were asked about planning to investigate with the selected materials. They were asked what they would be trying to find out and how they would do so (including follow-up questions such as, “What would you do first?”). The interviewer coded for mention of the following elements within students’ plans: appropriateness given the question, trials, systematic change in the variables, control of variables in set-up, changing only one variable at a time, consistent number of trials, measurement of variables, and recording data.

Students were then given blank paper for recording information regarding their investigation and were asked to talk about what they wanted to record about their investigation. The observation record allowed the interviewer to note whether the students’ plan for recording included reference to: a table, list, drawing, graph, narrative, or averaging.

**Data Collection.** The students then had time to collect data using the materials. Each pair had 45 minutes to complete the investigation, so the amount of time for data collection varied, depending on how long the students had spent preparing to investigate. While the students collected data, the interviewer took note of whether their activity demonstrated the following: recording data, measurement of variables, collecting data that fits the question, completion of trials, systematic change in the variables, control of variables in set up, changing only one variable at a time, and conducting a consistent number of trials.

**Interpretation.** After completing data collection, students were asked to examine their records and think about what they found out from their investigation. Students were also asked how they reached their conclusion (e.g. “How did you figure that out?”). Any references to collected data, observations of motion, or other experiences were noted on the interviewer’s observation record. The assessment ended with questions about what the students would do next.

**Data sources**

The first two authors conducted interviews with 12 pairs of students. Students were paired for the assessment to promote conversation that could reveal the nature of their reasoning on the assessment. Each assessment lasted approximately 45 minutes. The students’ activity was videotaped and the interviewer also coded student responses during the interview.
Developing concepts of evidence

The data sources for this analysis include the codes recorded during the interviews, videotape recordings of the interviews, and student records. Descriptive codes of students’ behavior recorded by the interviewers were compiled for examination across pairs. All videotapes were transcribed and entered into Nvivo for coding and analysis.

Coding of transcripts

Interview transcripts were coded to address the following areas: problem identification, modeling, planning, procedure, changing variables, identification of variables, attention to outcomes, data organization, interpretation, and conceptual understanding. A coding scheme was developed through a recursive deductive and inductive process of examining transcripts, informed by previously described literature on conceptions of evidence, modeling, and control of variables. Segments of transcript were assigned descriptive codes. Larger categories of codes were then created pertaining to the assessment phases and concepts of evidence framework (e.g. Variable identification, Modeling). After an initial coding scheme was established, a second rater coded one of the transcripts and inter-rater agreement was examined. Any differences in coding were discussed to joint resolution, leading to minor modifications to the coding scheme. All of the transcripts were then coded using NVivo. Table 1 below shows examples from the coding scheme with related transcript excerpts.

Table 1. Examples of codes

<table>
<thead>
<tr>
<th>Code</th>
<th>Description</th>
<th>Example from transcripts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Problem identification</strong></td>
<td><em>Sections of interview in which students are asked what they will be investigating</em></td>
<td></td>
</tr>
<tr>
<td>Students’ description of the problem presented in the scenario</td>
<td>Everyday action solution</td>
<td>Students suggest ways for Dante to solve his problem involving everyday actions.</td>
</tr>
<tr>
<td></td>
<td>Question-narrative focused</td>
<td>Student defines problem using language from the scenario, rather than variables.</td>
</tr>
<tr>
<td></td>
<td>Define problem using key independent variables</td>
<td>Student identifies Dante’s problem as involving mass and/or force.</td>
</tr>
<tr>
<td><strong>Modeling</strong></td>
<td><em>Section of the interview in which children are asked how they will model Dante’s situation.</em></td>
<td></td>
</tr>
<tr>
<td>Students’ efforts to use the materials provided to create a model of the problem posed for investigation</td>
<td>Confusion variables (mass/force/weight) in model</td>
<td>Student states s/he is adding mass when actually adding force, or vice versa. Or, student confuses weight with the other variables.</td>
</tr>
<tr>
<td></td>
<td>Correctly models</td>
<td>Students model scenario correctly, using cart, string, washers, and “grocery bags”</td>
</tr>
<tr>
<td></td>
<td>Pull/push to create force</td>
<td>Student pulls or pushes on the cart to make it move during investigation.</td>
</tr>
<tr>
<td><strong>Data Collection: Procedure</strong></td>
<td><em>Section of the interview in which children use materials to collect data</em></td>
<td></td>
</tr>
<tr>
<td>Students’ efforts to follow a data collection procedure during investigation activity.</td>
<td>Determine mistrial</td>
<td>Students determine that a particular trial/outcome should not be counted or recorded.</td>
</tr>
<tr>
<td></td>
<td>Monitors trials</td>
<td>Student explicitly attends to progression of trials (number, changes in mass/force)</td>
</tr>
</tbody>
</table>
The analysis drew upon interviewer codes of student behavior recorded during the investigation activity, student records, and codes associated with segments of transcripts. Patterns were determined by examining codes and associated transcript segments both within and across categories. The frequency and range of categories of behavior were examined using Nvivo and compared to coded data recorded during the interviews. Due to the fact that students were interviewed in pairs in order to promote verbal interaction, individual contributions were typically inextricable from the interaction of the pair. For this reason, the “pair” was used as the unit of analysis.

Findings
The findings section will largely be organized around the framework of concepts of evidence defined by Millar, Lubben, Gott and Duggan (1994) and Gott and Duggan (1996). Appendix A includes a description of the particular aspects of this framework. The children’s behavior during the investigation interviews provided a number insights into their developing conceptions of evidence. Children’s activity is reported by pairs only because individual contributions were typically inextricable.

Concepts of Evidence Associated with Design
Variable Identification
Problem Identification. When initially asked about the problem that they were trying to figure out, after being presented with a scenario and before collecting data, all of the pairs responded with a narrative-focused question. For example, one child stated the question, “How can he get the groceries across the street after he got them from the store in his wagon?” They did not readily abstract variables from the scenario to formulate a question. However, some of their responses provided potential inroads to key variables (See Table 2). Eight of the pairs stated a question pertaining to how many groceries the character could carry across the street, suggesting a focus on mass as key variable. Five of these pairs, plus one additional pair, also made reference to how long the character would take to cross the street, pointing toward consideration of the dependent variable of time.

Pairs responded in more than one way when probed to define the problem. Another common response, among 8 of the 12 pairs, was to suggest an everyday action that the character in the scenario could do to solve his problem, such as, “Maybe he can have a friend and his friend can carry some.” Similar to the narrative-focused questions, these responses did not involve the abstraction of variables. To an even greater extent than the narrative-focused questions, these responses did not typically situate the problem as one for investigation.

The same number of pairs however, also referred to at least one of the key variables, with five pairs referring to both independent variables, and 3 referring to one of them.

Identifying variables during investigation. While collecting data, students made more explicit reference to the independent and dependent variables in abstract terms than they had prior to data collection. Three-quarters of the pairs made reference to both variables, and half referred to the dependent. (More students collected time data, but did not explicitly refer to time as an outcome.)

Table 2. Identification of independent and dependent variables

<table>
<thead>
<tr>
<th></th>
<th>Prior to data collection</th>
<th>During data collection</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identify one indep. variable</td>
<td>8</td>
<td>-</td>
</tr>
<tr>
<td>Identify both indep. variables</td>
<td>5</td>
<td>9</td>
</tr>
<tr>
<td>Identify dependent variable</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>
Confusion over variables. Many pairs showed some difficulty keeping the variables straight in their minds during the investigation, with 8 pairs demonstrating some confusion. Confusions arose between mass and force, force and gravity, and mass and weight. In some instances, the students did not necessarily confuse the mass and force as concepts, but referred to a part of the modeled scenario by the wrong variable name. For example, when trying to decide whether the “groceries” were force or mass a student deliberated, “The mass. No the force. Hold on. (Puts his head down on arm. Looks up.) The mass (nodding), the mass.” In other instances, students more explicitly expressed an intertwining of the concepts, for example, “2 masses… isn’t mass like weight? What’s force like? How heavy something is?”

Modeling

After stating the problem for investigation, and before collecting data, all of the students were asked to model the character’s situation with materials. All of the pairs had seen a similar scenario modeled using materials at some point during their previous instruction. Half of the pairs had collected data with a similar set-up. The other half had read about a scientist modeling a similar scenario with materials and then seen the materials set-up after completing the text, but had not collected data with the materials.

All of the pairs who had read a text\(^1\) in which a similar situation was modeled, had greater difficulty using the materials to model this situation than those who had used materials to investigate a similar situation. The primary difficulty they encountered was not including a means of manipulating force in their model. When asked about how the cart would move, they resorted to physically pushing or pulling the cart. All of the six pairs that had used materials to investigate a similar situation, set up a similar model for this investigation, typically with minimal hesitation (See Figure 2).

With varying degrees of support from the interviewer, five of the six pairs that had read a text describing a similar study rather than investigating with materials successfully modeled the situation. In several instances, drawing students’ attention to the string seemed to trigger a shift in their view of the situation, as observed in the example below.

Monica: Oh yeah. Connect that to there (the cart) and use the washers.
Tisha: Where the washers going to go? Right here (puts washer on cart).
Monica: No, remember, we stuck them on the paper clip.
Monica attaches washers to the end of the clip.

Thus, similar numbers of pairs realized an accurate model at some point during the investigation. However due to the time taken to reach this model by pairs in the second-hand condition, they had less time left for collecting and interpreting data.

---

1 The texts the students read had been written especially for this investigative context. They were designed to be like a scientists’ “notebook” and chronicled the investigations of a fictional scientist. Design features of these texts included the use of figures, data tables, modeling of real-world situation, and knowledge claims based on represented data (Palincsar and Magnusson, 2001)
Whether students conduct fair tests lies at the heart of many studies of scientific thinking (Chen & Klahr, 1999; Schauble, et al., 1995; Toth, Klahr, & Chen, 2000). The children exhibited various behaviors related to conducting a fair test including: changing variables, control of variables, determining mis-trials, and monitoring data collection activity. These many efforts toward the goal of conducting a fair test were sometime accompanied by subtle deviations from the procedure that undermined a fair test.

Change of variables. Students varied in their approach to changing variables (See Table 3). Four pairs engaged in control of variables, as commonly defined, by only changing one independent variable at a time and keeping the other independent variable constant. Two of these pairs changed only the force and two changed only the mass. Another pair took an alternative approach to varying one variable at a time by changing the mass and holding the force constant, and subsequently changing the force and holding the mass constant. One of the students specifically identified this plan for changing the mass and force, “Four four (referring to the mass then the force), three four, three three, two three, two two, whatever.”

Table 3. Student approaches to changing variables during data collection

<table>
<thead>
<tr>
<th>Change of variables</th>
<th>Number of pairs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Change force only</td>
<td>2</td>
</tr>
<tr>
<td>Change mass only</td>
<td>2</td>
</tr>
<tr>
<td>Change mass or force</td>
<td>1</td>
</tr>
<tr>
<td>Change mass &amp; force at the same time</td>
<td>7</td>
</tr>
<tr>
<td>Change mass; mass &amp; force</td>
<td>3</td>
</tr>
<tr>
<td>Change mass; force; mass &amp; force</td>
<td>3</td>
</tr>
<tr>
<td>Change mass &amp; force</td>
<td>1</td>
</tr>
</tbody>
</table>

Over half of the pairs changed both the mass and the force simultaneously at some point during their data collection. In several instances, students would sometimes change only one variable, and at other times change both. In one instance the pair consistently changed both the mass and force, reducing both by one unit for each set of trials.

Control of Variables/Systematicity.
Developing concepts of evidence

As previously mentioned, 4 pairs engaged in consistent control of variables as it is commonly described. Another pair followed an atypical approach to control of variables in which they alternated varying mass and force. While this approach was systematic and did not lead to confounded results, it led to data that were organized in a manner that was not conducive to noticing patterns. In their data table, recorded times related to different amounts of mass were separated by recorded times associated with different amounts of force. They therefore had appropriated some practices for conducting trials, but did not have a big picture view of how to design their investigation in order to facilitate interpretation of the evidence to address a specific question. The pair that consistently changed the force and mass by one unit also followed an explicit system and had appropriated the practice of incrementally changing the variable, but did not understand the importance of only changing one variable at a time. Thus, in a couple of instances, students revealed progress toward systematic approaches to investigation, but had not yet developed the conception of variables in a manner that facilitates interpretation of evidence.

Mistrials. Another behavior that provided insights into children’s developing conceptions of how to conduct a fair test was their determinations that certain outcomes should not be recorded, that in effect the trial should be re-done. In a couple of instances, students determined a trial was a mis-trial due to inconsistencies in the set-up of materials or conduct of the trials— for example, the cart falling off the side of the ramp before reaching the end because it was crooked. They thus, perhaps only implicitly, recognized that the situation was not comparable to their other trials.

Monitoring. Through various monitoring efforts, the students revealed attention to how they were manipulating the variables and reducing the potential impact of extraneous variables. All but one pair of students, or all of the pairs that conducted trials, showed some effort to monitor their activity, as seen in Figure 3.

![Figure 3](image-url)

**Figure 3.** Observed types of monitoring behaviors.

Although 11 pairs exhibited both monitoring of trials and procedure, the type of monitoring that occurred with the most frequency was monitoring trials. Monitoring trials included overt efforts to track the progression of trials and attend to the amount of force or mass for a given trial. For example, before conducting a trial, a student commented, “Okay, now we’re going to do with a mass of 2, force of 1, and trial 2, and groceries are going to be 2.” Monitoring trials was observed about twice as frequently as the other types of monitoring, occurring 53 times as compared to 23, 28, and 28 for monitoring of set-up, procedure, and recording respectively.
This behavior was actually observed most frequently by a pair of students who never explicitly referred to the variables by name throughout the investigation, but made comments such as, “we didn’t do it with four groceries yet.” Thus, even though these students were not explicitly identifying the independent variables, they were keeping track of these variables in practice. Efforts to monitor trials revealed that students were making efforts to keep track of their actions and the sequence of actions, even though most were not at the point of effectively controlling variables.

The same 11 pairs also made some effort to monitor the procedure, which involved attending to the consistency of enacting the trial. These behaviors were observed among the same groups of students, but less frequently. As an example, after a student dropped the washers more forcefully during one trial, her partner responded, “You can’t drop it. You can’t drop the thing.”

Another effort to maintain the consistency of the trials involved monitoring the set-up of materials. Although this occurred almost as frequently as monitoring procedure, it was only observed among six, or half, of the pairs. Monitoring the set-up of materials focused on either lining up the materials or noting the need to change the materials being used. For example, one partner commented to the other, “Put that straight. Put the tire at the end.” Again, although few of the students adequately controlled variables, such attention to procedure shows progress toward removing extraneous variables and appropriation of the norms of consistency in use of materials.

**Deviations from procedure.** For all of their efforts to conduct consistent trials, a number of pairs showed occasional lapses in their attentiveness to following a standardized procedure. Of the 11 pairs that followed a procedure, 8 were noted to deviate from it at some point in varying ways, including elevating the washers before dropping them, altering the plan for changing mass and force, tugging the string to make the cart move, and placing the cart off the edge of the ramp.

**Concepts of Evidence Associated with Scale Measurement**

Within the context of the provided task, students revealed less about their conceptions associated with scale measurement than they had about their conceptions associated with design. However, their behavior did provide some insights related to their understanding of choice of instruments, repeatability, and accuracy.

**Choice of instrument**

The students’ choice of instrument was constrained by the materials provided. Other than the stopwatch, the only other measuring instrument available to select was a tape measure. The materials also included a spring, which had been used during investigation on incline planes to examine force. All of the pairs did use the stopwatch at some point during their investigation activity and used it to time the motion of the cart. Half of the pairs also used the tape measure to determine the length of the board.

Two of the pairs used the measuring tape to measure the stretch of the spring. Similar to the investigations they had engaged in during instruction on inclined planes, they attached the spring to the back of the cart, which had objects inside it and was being pulled forward by a string with washers attached. A student held the spring at the back of the board and they examined the length of the spring. One student explicitly described this approach as a means of measuring gravity.

**Confusions about how to measure.** A few pairs experienced confusion related to measurement. Questions arose about how to measure force and time that suggested the lack of ideas about appropriate instruments. For example, a student suggested using the measuring tape to measure time, and another suggested using it to measure force. Other confusions were related
more to students’ need to develop greater skill at using these tools, rather than their conceptions of what instrument was most appropriate to use. For example, one pair revealed confusion about how to interpret the measurement from the stopwatch, reading the seconds as minutes. Another pair expressed confusion about how to use the measuring tape. However, only four of the pairs demonstrated such confusions.

**Repeatability**

Even when students closely attended to the set-up and timing of the cart, the measured times across trials varied enough to make multiple trials necessary to collect reliable data. Seven pairs conducted multiple trials. However, their behavior did not necessarily reflect an understanding of inherent variability. They did not show any indication of deciding to conduct more trials based on the variability observed in previous trials.

**Accuracy**

Students’ developing attention to the accuracy of outcomes was most clearly evident in their determination of mistrials due to problems with timing. The most common reason, observed among 5 of the 12 pairs, that students decided not to count an outcome was due to inaccuracies in timing—they either started or stopped timing early or late. In making such a determination, these five pairs made an effort to remove measurement error from their trials. The students all timed trials to hundredths of a second, consistent with the practice during their investigation experiences. Thus, while they collected data with a degree of accuracy that would allow for interpretation, it was not clear that the students made determinations about what degree of accuracy was appropriate, but rather that they followed previous experiences.

**Concepts of Evidence associated with Data Handling**

**Recording trials.**

Only 1 of the 12 pairs did not record data. Thus, almost all of the students recognized the need to record outcomes from their activity. Seven students within four pairs initially only recorded dependent variable data. That is, they listed the time outcomes. Five of those seven students subsequently recorded both dependent and independent data when probed. Thus, students in 9 of the 12 pairs recorded independent and dependent variable data at some point during their investigation activity. And another 2 pairs recorded only the outcomes.

Almost half of the pairs, 5 of 12, did not record outcomes when they should have at some point during their data collection period. Even a pair that explicitly planned to record initially forgot to do so until probed.

**Creating a table.**

Half of the pairs recorded their data in a table. This practice had been introduced to the students during the instruction. The teachers had supported the formation and use of tables during the instruction, and thus they had little experience constructing tables independently. The other students either recorded in the form of lists or sentences.

**Concepts of Evidence associated with Patterns in Data**

Students had greater difficulty finding patterns in the data or making claims based on data than they did collecting the data. Once they had a data record, they seemed uncertain what to do with it. Even when they did generate claims, these were not necessarily grounded in the data. None of the students made a clear claim based on examination of the full body of data that they collected. Use of data, when it was used, was more selective.
Claims based on data

Student statements were coded as claims if they made a generalized statement pertaining to the key concepts involved in the investigation, such as, “The less mass, the faster the cart goes.” Claims were further differentiated by whether they were conceptually framed in terms of mass or force, or contextually-framed in terms of the materials used in the task (washers, groceries).

Across pairs, students made a similar number of conceptually- and contextually-based claims (see Table 4). Approximately 33% (4 of 12) of the pairs made accurate conceptual claims and 42% (5 of 12) made accurate contextual claims. For both types of claims, only two pairs supported claims with evidence from their collected data. Students were considered to support their claims with evidence if they referred to collected data relevant to their claim. Thus, students could provide evidence for their claim without effectively drawing conclusions from across the full range of their data. Using a stricter definition of providing support from the data, involving reference to patterns across the set of collected data, none of the pairs made evidence-supported claims.

Table 4. Frequency of Accurate Conceptual and Contextual Claims

<table>
<thead>
<tr>
<th></th>
<th>Conceptual</th>
<th>Contextual</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of claim statements (with evidence)</td>
<td>8 (1)</td>
<td>8 (2)</td>
<td>16 (2)</td>
</tr>
<tr>
<td># pairs making claim statements (with evidence)</td>
<td>4 (1)</td>
<td>5 (2)</td>
<td>7 (2)</td>
</tr>
</tbody>
</table>

One-third of the pairs made inaccurate claims. One inaccurate claim was perhaps due to confusion over mass and force in the model. For example, after noting that if the boy in the scenario had more force, he would go faster, one child concluded from her data, “when he had more force, he went slower.” Actually, she and her partner had controlled variables by only changing the mass.

In another example of an inaccurate claim, a student referred to the data in stating a claim, but asserted that the outcome was atypical in order to defend her claim. In the face of contrary evidence, and a contrary claim from her partner, she did not disregard the evidence or warp it to fit her claim, but distinguished what their data showed from what she considered to be the more typical outcome. Within her interpretation she stated, “Usually more mass makes stuff go faster. But down here has three washers and two groceries on it—that’s mass— and it says lower…I’ve been saying up here (points high on paper) it’s a good time. But down here (points lower on paper) usually more mass makes it go faster.”

Seeking Patterns. Five pairs, four of the pairs that made accurate claims and one additional pair, expressed some effort to find a pattern in the data (See Figure 3). These efforts did not lead to clear interpretations of the data, but show progress toward trying to make sense of the data. For example, after conducting several trials, a student remarked to her partner, “It went highest to lowest. First it was 2.56, then 2.31, and then 1.87.”

Drawing Conclusions. Rather than making claims, students in over half the pairs (7 of 12) responded to probes about what they found out from their investigation by drawing conclusions related to the scenario outcome, such as, “He could take four across the street” or “he safely made it across.” These statements were not typically made with reference to the data. While such responses reveal an emerging understanding of doing an investigation in order to answer a question, they do not indicate a search for patterns and relationships in the data.

Attending to outcome. As perhaps even an earlier step in the process toward interpretation, students described and evaluated outcomes while conducting trials. Nine of the 12 pairs (75%) described outcomes as “faster” and “slower,” or made evaluations such as “impressive” and “it’s getting bad” (See Figure 5). Such remarks do not involve establishing
relationships between variables, but perhaps indicate a step toward examining outcomes in relation to each other, as part of a larger set of outcomes. They also demonstrate active engagement in examining outcomes, rather than rote running of trials.

Figure 5. Student pair responses pertaining to patterns in the data.

Discussion

Our discussion of these analyses and finding have two foci. First we examine our findings in relation to the concepts of evidence framework and suggest several ways in which this framework could be elaborated and expanded. We then discuss several major points about our specific results.

Reflections on the concepts of evidence framework

We found the concepts of evidence framework to be very useful for describing most of the fourth-graders’ investigation activity. However, some of the students’ responses were not readily accounted for within the framework. Below we examine each of the three categories of concepts of evidence in turn to discuss the ways in which the framework matched or did not match what we observed.

Concepts of evidence associated with design. The UK framework includes four concepts of evidence associated with investigation design: variable identification, fair testing, sample size, and variable types. The task was not conducive to considering issues of sample size and variable type, but within the task presented, children revealed some of their developing understandings pertaining to identifying variables and fair testing.

At the beginning of the clinical interview children had difficulty with identifying the problem for investigation, which is not even an aspect of investigative work considered within the UK framework. This is probably an artifact of the differences in the tasks. In the Procedural and Conceptual Knowledge in Science (PACKS) Project (1994) from which the concepts of evidence framework was derived, the tasks presented to students had specific questions to investigate (e.g. “Find out how the type of surface affects the amount of pull needed to drag a brick along”). However, in this study, children had to translate a real-world situation into an investigative context.

Capturing children’s activity required expanding the idea of fair testing in order to broaden the notion of systematicity and acknowledge the role of monitoring. In a few instances, children attempted to be systematic in the way they varied key variables but did not adhere to canonical control of variables strategy typically associated with fair testing. However, these attempts were more than purely random manipulation of the materials. In addition, analysis of children’s behaviors potentially related to fair testing highlighted the prevalence of several forms of monitoring, such as efforts to be consistent with their procedure and recording of the
Developing concepts of evidence investigation. Children also showed progress toward implicit control of variables through distinguishing between mistrials and fair trials.

Finally, within the investigation task used in this study, modeling was a central issue related to defining variables and conceptualizing the problem. This critical role of modeling in theory construction and testing (Lehrer & Schauble, 2000) is not included within the concepts of evidence framework. As Lehrer and Schauble have described it, modeling extends beyond simply having objects “stand” for other objects and involves selecting “representations for the purpose of highlighting objects and relations that are theoretically important” (2000, p. 40). Within this study, students who had investigated a similar situation with materials more readily constructed a model than those who had only read about how a scientist had modeled a similar situation. However, with prompting, almost all of the children who had read text were also able model the scenario.

**Concepts of evidence associated with measurement.** The concepts of evidence associated with measurement include: relative scale, range of interval, choice of instrument, repeatability, and accuracy. The children’s activity provided limited insights into their understandings of these issues due to the task constraints, particularly with respect to relative scale and range of interval.

Children’s choice of instrument was highly constrained by the available materials, however the difficulties experienced by some pairs did draw attention to the fact that a critical precursor to selecting an appropriate instruments knowing how to measure the variable under consideration. Knowing how to measure force proved problematic for some pairs. Observations of students’ activity revealed insights into their awareness of the need to conduct multiple trials (repeatability) and attention to accuracy. They often followed a rote approach to repeating trials and recording outcomes that suggested they did not have a deep understanding of why to run multiple trials or what degree of accuracy was optimal for subsequent interpretation. Pairs that conducted multiple trials typically completed three or four trials for each variable configuration, reflecting the procedure followed during their previous investigation experiences. Students’ awareness of the need for accuracy was most evident in their determination of mistrials.

**Concepts of evidence associated with data handling.** The concepts of evidence associated with measurement include: tables, graph types, and patterns. Observations of these children’s activity revealed that they were early in their progress toward developing conceptions of evidence associated with data handling. Use of graphs was not included in the assessment or prior instructional experiences. Half of the pairs did record their data in a table, however their use of tables did not indicate that they had reached the point of using them to organize the design the analysis of their investigation. The students demonstrated a broader repertoire of ways of recording data, of which using a table would be one, but not the only way. Efforts to capture the investigation of young learners with minimal investigation experience will thus require broader consideration of children’s recording attempts.

**Broader reflections on the findings**

Overall, the students had difficulty, as will be discussed further below, with finding patterns in the data. However, they exhibited potential precursors to the notion of finding “patterns” as described by Gott & Duggan (1996). We found a range in the extent to which pairs even recognized that their collected data could be examined for patterns. Some pairs seemed to have difficulty moving beyond rote data collection to trying to make sense of the numbers they had recorded. In addition, before they can seek patterns in the manner assumed in the concepts of evidence framework, students must develop the expectation that assertions about phenomena will be based on the range of collected data, and explicitly connected to that data. We saw few examples of students making knowledge assertions based on data, and none in which they considered the full complement of their data.
Having considered the usefulness and potential elaborations upon the conceptions of evidence framework, we now address three major points suggested by the specific results of our analyses. The first point relates to the affordances and constraints of presenting an everyday situation as the impetus for a formal scientific investigation. On one hand, using everyday scenarios is intended to make the activity more engaging, meaningful and authentic. However, as the students in this study demonstrated, students may have difficulty shifting from a context-specific, narrative view of the scenario to a more abstracted view of the situation. Students showed a strong narrative focus in their identification of the problem for investigation. It is not that narrative contexts cannot be conducive to generating complex thinking (Bloom, 1992). Rather, as Hogan and Maglienti (2001) have also suggested, students bring their own frames of reference to classroom science experiences and our challenge becomes one of expanding students’ repertoires to include the ability to abstract critical variables from real world situations in order to formulate scientifically testable situations.

It was striking that although most pairs did not initially frame the problem in terms of abstract variables, most of them (9 of 12) did identify the independent variables during data collection. Duggan, Johnson, and Scott (1996) hypothesize that students can more readily identify and operationalize variables that are more “concrete.” In the present study, the variable mass has a more tangible, concrete quality than the relatively abstract variable of force. Students’ capacity to identify this abstract variable, albeit following instruction, is noteworthy. Use of washers to create force perhaps supported students to identify it as a variable. However, use of washers, which have mass, to create force also introduced a potential source of confusion. Unless students remained focused on the washers as a way to create the pull of force, changing the number of washers used to create force could have seemed like a change in mass.

Still, despite these potential confusions and the fact that most of the pairs did not initially identify a complete set of key variables, all but one of the pairs did collect some pertinent data over the course of the clinical interview. This brings us to our second major point that, even after limited classroom experiences, the students had begun to appropriate some of the norms and conventions of scientific inquiry. Yet, there is “decalage” or uneven-ness in students’ appropriation scientific practices.

Nine of the 12 pairs recorded both independent and dependent variable information at some point. This finding is not surprising given that in the previous instructional experiences all children had documented outcomes and examined data tables. The fact that a number of pairs did not initially record both dependent and independent variables demonstrates that they are at the beginning of their enculturation into the norms and expectations of scientific investigation.

We didn’t see very many instances of children demonstrating awareness, nor certainly command, of the control of variables strategy. Students’ appropriation of the control of variables strategy is an example of an important norm and convention of science that only a minority of the children had begun to appreciate. Perhaps the fact that more pairs did not use the control of variables strategy is not surprising given that the instruction did not focus primarily on this strategy and was relatively brief in duration. Even after explicit training in the control of variables strategy, Chen and Klahr (1999) found that change in students’ design of investigations occurs gradually. Even after learning the control of variables strategy, students used ineffective strategies on some trials. We did, however, find it encouraging that several pairs did vary only one variable at a time, and even specifically described their intentions to do so, thereby providing further evidence that elementary students can make sense of this strategy (Chen and Klahr, 1999; Toth, Klahr, & Chen, 2000).

The complexity of the problem may also have contributed to students’ difficulty with control of variables. Duggan, Scott, and Johnson (1996) found when the number of independent variables in a problem shifted from one to two, students’ performance in terms of identifying the
Developing concepts of evidence

independent variable and controlling other variables declined. Faced with problems with multiple independent variables, some students seemed to experience “cognitive overload” and demonstrated less sophisticated conceptions of evidence.

Our third major point is to raise to the foreground of our attention the complexity of data interpretation. In descriptions of the importance of specific practices, such as the control of variables strategy, it seems assumed that if students manage to collect unconfounded data, interpretation of these will follow as a matter of course (Chen & Klahr, 1999). The responses of the children in this study point out the potential disconnect between data collection and interpretation. Although they could record outcomes, once they were completed with data collection, students seemed at a loss as to what these numbers could tell them. In many respects this finding is not surprising given that most of the students had not collected or recorded data in a manner conducive to identifying patterns. Even when the data was organized to facilitate interpretation, the variability in the outcomes (in part due to the difficulties students had with being precise in their conduct of the investigation) may have posed difficulties for identifying patterns. In addition, finding patterns requires substantial higher-level thinking, and students may not be accustomed to engaging in such cognitive work. Perhaps students’ tendency to frame the investigation in a narrative or “everyday” frame may also have contributed to their difficulties relating their data collection activity to the initial problem scenario.

The continuing prominence of the narrative scenario for the students is suggested by the fact that over half of the students drew conclusions about the scenario outcome when asked to describe what they had found. This response may have been due to the phrasing of the problem in the scenario. The scenario presented at the beginning of the interview ended with a specific question, “How many groceries can he [Dante] carry in his wagon and still safely walk across the street?” During their investigation activity and the interpretation phase of the interview, students were prompted to focus on their own question, rather than the scenario. However, including this question may have primed students to reach a specific, concrete conclusion, rather than focusing on the broader conceptual relationships. In addition, use of a “real life” scenario may more generally have promoted efforts to complete the story, in a sense, through formulation of a conclusion to Dante’s situation.

When students in this study did make claims about the relationships between mass, force and speed, they rarely supported them with evidence from their data. Perhaps they did not recognize using data as a basis for one’s conclusion as a norm for their activity. In other words, they were not aware of the “expectation” to explicitly refer to their data as part of the asserting a conclusion. A more common behavior relating to data interpretation was attending to the outcomes of trials. This behavior presents a potentially valuable inroad to more sophisticated levels of interpretation and may serve as a precursor to more advanced analysis and synthesis of outcomes.

In considering students’ challenges with interpretation, Duggan, Johnson, and Scott (1996) point out the difference between measuring a variable and understanding it. As a case in point, the students in this study were able to measure time; they had the necessary skills to use the stopwatch. However the fact that they can readily measure time does not mean that they had a firm understanding of time in relation to speed, of speed as a continuous variable, or of the relation between speed and the independent variables. Their facility with timing in contrast to interpretation perhaps illustrates the distinction in demands for competency between skills and concepts of evidence.

It may well be the case that students also need a great deal of time to “muck about” and be immersed in their data. For example, in contrast to our findings, Metz (in press) observed that 4th and 5th grade students “manifested a keen attention to their data” and successfully reasoned about their investigation activity as a means to address a question. The difference between observed outcomes in the Metz study and this one perhaps accentuate the importance of
experience. The students in the Metz study were interviewed following an 8-week science unit, as compared to the 2 weeks of instruction in this study. In order to reach the point of “keen attention to their data” students will need repeated, supported experiences.

It is also important to recognize that the activity of interpreting data is one that is extremely challenging and deserves discussion with students in its own right. As Elfin, Glennan & Reisch (1999) have suggested, "educators should discuss the idea of empiricism more generally. This should include discussion of the many and different ways that experience and the use of experiments informs scientific beliefs, and also the important fact that scientists often argue about how to understand and interpret the results of measurements and experiments" (p. 114). In a study examining the activity of 6th graders over the course of 10 weeks, examining and reexamining phenomena related to the density of liquids, Vellom and Anderson (1999) found that with sustained investigative activity, the students needed to use progressively more scientific persuasive strategies, especially those focusing attention on experimental technique and replicability (rather than the merely rhetorical strategies they used initially) to reach consensus. This suggests that a highly integrated approach in which the norms and conventions of data collection and handling are discussed in relation to the interpretations of the data collected can have an impact on the ways students use come to support their knowledge claims.

To summarize, across the various conceptions of evidence, students’ behavior suggested difficulty engaging in higher order thinking. They had difficulty analyzing the scenario and abstracting variables, and a number of pairs did not show a meta-level awareness of how to collect data in a manner that would be readily interpretable, either in their design or recording of trials. The students’ general difficulty with interpretation also suggests weaknesses in their competence at extracting pertinent information and synthesizing information from across the body of data.

At a general level, it’s fair to say that the students’ typically had a proximal rather than global view of their activity. They successfully monitored their procedure for individual trials, discussed how to change the variables for the next trial, recorded outcomes for each trial, and noted outcomes for specific trials – thus there are ways in which their activity is very productive and will provide a basis for their subsequent science inquiry learning. Yet, they did not manage to integrate these activities and think about the big picture of their work.

The findings of this study are constrained by several limitations. In particular, not having had pre-instruction interviews limits our ability to attribute the children’s performance to learning from the instruction. As mentioned earlier, the wording of the question in the Dante scenario may have unduly constrained the children’s identification of the problem and interpretation of their data. Finally, although conducting the interview with the students in pairs yielded valuable discussion, this approach makes it difficult to determine individual children’s conceptions.

Rather than determining the presence or absence of very complex conceptions regarding how to conduct scientific investigations, this study highlights the value of examining children's progress in a range of competencies related to investigations. It provides examples of children’s behavior that suggest potential precursors to the complex conceptions of evidence as described by Millar, Lubben, Gott and Duggan (1994). Ultimately, issues related to experimentation and observation are not considered the cornerstones of scientific activity, “rather they are the handmaidens to the rational activity of constituting knowledge claims through argument. It is on the apparent strength of arguments that scientists judge competing knowledge claims and work out whether to accept or reject them.” (Newton, Driver & Osborne, 1999, p. 555) Yet, we suggest that having knowledge of the intricacies of data collection and data handling may give students an inroad to acquiring the means with which to critically examine their own and others’ knowledge claims – certainly the hallmarks of being scientifically literate.
References


<table>
<thead>
<tr>
<th>Associated with Design</th>
<th>variable identification</th>
<th>Understanding the idea of a variable and identifying the relevant variable to change (the independent variable) and to measure, or assess if qualitative (the dependent variable)</th>
</tr>
</thead>
<tbody>
<tr>
<td>fair testing</td>
<td></td>
<td>Understanding the structure of the fair test in terms of controlling the necessary variables and the importance that the control of variables has in relation to the validity of any resulting evidence.</td>
</tr>
<tr>
<td>sample size</td>
<td></td>
<td>Understanding the significance of an appropriate sample size to allow for instance for random or biological variation.</td>
</tr>
<tr>
<td>variable types</td>
<td></td>
<td>Understanding the distinction between categoric, discrete, continuous and derived variables and how they link to different graph types.</td>
</tr>
<tr>
<td>Associated with</td>
<td>relative scale</td>
<td>Understanding the need to choose sensible values for quantities so that resulting measurements will be meaningful: for instance, a large quantity of chemical in a small quantity of water, causing saturation, will lead to difficulty in differentiating the dissolving times of different chemicals.</td>
</tr>
<tr>
<td>measurement</td>
<td>range and interval</td>
<td>Understanding the need to select a sensible range of values of the variables within the task so that the resulting line graph consists of values which are spread sufficiently widely and reasonably spaced out, so the ‘whole’ pattern can be seen; a suitable number of readings is therefore also subsumed in this concept.</td>
</tr>
<tr>
<td>choice of instrument</td>
<td></td>
<td>Understanding the relationship between the choice of instrument and the required scale, range of readings required, and their interval (spread) and accuracy.</td>
</tr>
<tr>
<td>repeatability</td>
<td></td>
<td>Understands that the inherent variability in any physical measurement requires a consideration of the need for repeats.</td>
</tr>
<tr>
<td>accuracy</td>
<td></td>
<td>Understanding the appropriate degree of accuracy that is required to provide reliable data which will allow a meaningful interpretation.</td>
</tr>
<tr>
<td>Associated with</td>
<td>tables</td>
<td>Understanding that tapes are more than ways of presenting data after it has been collected; they can be used as ways of organizing the design and subsequent data.</td>
</tr>
<tr>
<td>data handling</td>
<td>graph types</td>
<td>Understanding that there is a close link between graphical representation and the type of variable they are to represent: for example, a categoric independent variable such as type of surface, cannot be displayed sensibly in a line graph; the behavior of a continuous variable on the other hand is best shown in a line graph.</td>
</tr>
<tr>
<td></td>
<td>patterns</td>
<td>Understanding that patterns represent the behavior of variables and that they can be seen in patterns in tables and graphs.</td>
</tr>
</tbody>
</table>