Cognition and Instruction, in press

What Needs to Develop in the Development of Inquiry Skills?

Deanna Kuhn and Maria Pease

Teachers College, Columbia University

Contact:

Box 119

Teachers College Columbia University

New York NY 10027

dk100@columbia.edu

What Needs to Develop in the Development of Inquiry Skills?

Abstract

With the objective of identifying the challenges that students must meet to engage in effective self-directed inquiry, a class of diverse students were followed for three years, from the fourth through the sixth grades, as they engaged in a sequence of progressively more demanding inquiry activities. Students made substantial progress in understanding the objectives of inquiry, attending to evidence and identifying patterns, making controlled comparisons, interpreting increasingly complex data (that included interactive and probabilistic effects), supporting their claims, drawing justified conclusions, and inhibiting unjustified ones. Retaining awareness of inquiry objectives and integrating influences of multiple variables in predicting outcomes were two areas that remained challenging, despite progress. A comparison group of seventh graders who had not been involved in the program, by contrast, displayed strikingly different (and inferior) approaches to an inquiry task, indicating that the skills identified here are not ones that develop in the absence of appropriate kinds of educational experiences.

Inquiry skills occupy a prominent place in American science curriculum standards, beginning in the early grades and every year thereafter through grade 12 (National Research Council, 1996). Clearly, they are not regarded as skills to be taught and learned at one discrete time and thereafter expected to be in place. Rather, it is prescribed that students engage in inquiry activities repeatedly, during each year of elementary and secondary science instruction, presumably enriching, expanding, and consolidating their inquiry skills with each engagement.

The assumption that underlies and motivated the work described here is that it would be beneficial to be able to explicitly identify and describe these skills in as specific and detailed a manner as possible – an objective that could only enhance implementation of instructional goals related to inquiry. The present work is dedicated to this objective.

A further assumption underlying the work is that inquiry skills can be *defined* with some degree of generality (which is not to say that they should be taught as general principles – indeed, the assumption carries no instructional implications at all). Their repeated appearance in the curriculum standards across all the grade levels supports such an assumption. Skills that were entirely tied to a specific subject matter or specific instructional context would not be likely to warrant repeated inclusion in the curriculum standards across all grade levels.

A number of researchers, such as Klahr (2000), and indeed the NRC (1996) curriculum standards themselves, characterize a complete inquiry cycle as comprising identification of a question or questions(s), design of an investigation to address them, examination and analysis of empirical data, and drawing inferences and conclusions and justifying them. Klahr (2000) notes in this regard that only a minority of research studies investigate the entire cycle, rather than just portions of it. Despite widespread acceptance of this general definition of inquiry, beneath the surface of this apparent agreement there exists considerable debate as to exactly what inquiry does and does not entail (Duschl & Grady, 2005; Ford, 2005; Fortus, Hug, Krajcik et al., 2006; Krajcik, Blumenfeld, Marx, Bass, Fredricks & Soloway, 1998; Kuhn, 2005; Kuhn, Iordanou, Pease, & Wirkala, in press; Lehrer & Schauble, 2006: Metz, 2004; Reiser, Smith, Tabak, Steinmuller, Sandoval, & Leone, 2001; Sandoval, 2005). At one extreme, the inquiry process is regarded as narrowly as a control-of-variables strategy that can be taught to students in a single brief session (Klahr & Nigam, 2004); at the other it is an activity so complex and evolving with practice as to defy simple characterization (Lehrer & Schauble, 2006), **with many other**

conceptions intermediate between these two (for recent reviews see Lehrer & Schauble, 2006; National Research Council, 2007; Zimmerman, 2007).

It is thus particularly important that we begin by making clear the conceptual model of the inquiry process we bring to this work and the role of this model in the sequence of activities we employ here. We see inquiry as having both epistemological and strategic aspects, with developments on the two fronts reinforcing one another. To engage in authentic and productive inquiry, students must come to understand inquiry not as the accumulation of objective facts but as an enterprise that advances through the coordination of evidence with evolving theories constructed by human knowers. To achieve this epistemological understanding, we follow Sandoval (2005) in believing that such understanding is best fostered in a context of inquiry activities **that students themselves conduct. In other words, students learn best about the nature of science by engaging in it, albeit at a rudimentary level. Rather than demonstrations to students of outcomes that are known in advance, even if they participate in their production, authentic inquiry involves investigations characterized by clearly identified questions with answers that are not known to students in advance. Authentic inquiry is also motivated (Kuhn, 2002), in the sense that the process is driven by an explicit intention to find out.**

Strategic competence, as we elaborate below, entails both procedural and meta-level components. First of all, and most fundamentally, it requires recognizing that there is something to find out – that new evidence can be distinguished from and potentially bears on existing understanding – and identifying what this something is within the context of any specific inquiry activity. Strategic capability further entails designing investigations that will yield informative evidence, interpreting it appropriately, reaching justified conclusions, and revising one's theoretical understanding as warranted. We also include a strategic component not typically included in research on inquiry – generating predictions consistent with the understanding that has been achieved. Still another component of inquiry, one that has both epistemological and strategic aspects, is recognizing the importance of and having the skill to enter one's claims into social contexts in which they can be justified, debated, and potentially revised – in other words, the skills represented in the conception of science as argument (Kelly, Druker, & Chen, 1998; Kuhn, 1993, 2002, 2005; Lehrer, Schauble, & Petrosino, 2001; **Newton, Driver, & Osborne, 1999; Osborne, Erduran, & Simon, 2004; Patronis, Potari, & Spiliotopoulou, 1999).**

With respect to developmental process, our view is that these capabilities develop only gradually in the context of rich practice in activities that entail them (Dean & Kuhn, 2007). We report here a three-year longitudinal study in which we follow the progress of a group of students from their fourth-grade through their sixth-grade years, as they engage in a sequence of activities we designed to provide dense experience with a set of progressively more challenging inquiry activities. Our research goal was to better articulate what needs to develop in the realm of inquiry and to better understand the particular challenges this development poses. Beginning with their initial activities, students engage in the entire inquiry cycle – identifying a question or questions, accessing data of their choice to address the question, analyzing these data to identify patterns and make inferences, and, finally, drawing conclusions and making judgments based on them. Adult coaches provide support as needed but do not engage in any **explicit** instruction. The activities are socially situated, with an aim of helping to promote among students a sense of why there is a need for and value in the kinds of strategic practices that are developing.

The activity sequence we devised to support this development evolves in a number of respects over the three years. It begins in the first year with software-based activities that scaffold the inquiry process, supporting students in identifying a question, choosing relevant data to access, making and justifying interpretations of the evidence, and relating them to their existing understandings, including making predictions based on them. These activities gradually extend across varied kinds of content. Over time, this software support is reduced, becoming less structured, allowing students to conduct their investigations **more** independently **(within the general framework of investigating causal and noncausal influences of multiple variables on outcomes).** The kinds of human scaffolding that remain are comparable in form throughout, their purpose being to remind students to reflect, collaborate, ask themselves questions, and justify their conclusions. Initially such support is included within the software; later, when students begin to work independently of it, this support function is confined to human prompts, administered as needed when students show lapses in self-regulation of their activities.

An equally important evolution over time is in the complexity of the database students work with. This was gradually increased, posing new challenges, which students were provided support to meet as needed. **Initially, for example, the database students worked with involved only simple additive effects of individual variables. Over the course of their work with different databases involving different content, students gradually encountered more**

complex forms of data, including effects of variable sizes, probabilistic effects, and interactive effects. An examination of the challenges students need to meet as they address these increasingly complex forms of data, and the nature of the skill development they demonstrate in doing so, constitutes the empirically-grounded analysis we present here.

Should the present work be regarded as an empirical test of an a priori model of "what develops" or as a characterization of the inquiry process deriving from the present data? The answer to this important question is that neither of these extremes is the case. We undertook the study with definite conceptions of what inquiry at the middle-school level involves, informed by our own (as well as others') previous research on the topic (and later we connect the present findings to previous work). At the same time, the present work enhanced and clarified many of these conceptions, particularly with respect to the challenges that various aspects of the inquiry cycle pose to this age group.

Although the work we present here is thus empirically-grounded, it is well to make explicit the several assumptions that guided it. We have already noted the key assumption that inquiry skills develop best in the context of rich practice in activities that entail them. It is this dense practice that we undertake to provide over the three years of the present study, rather than any explicit instruction in inquiry procedures. A further assumption is that it is in the context of such practice – practice that in the present period of investigation becomes increasingly self-directed – that researchers can best observe and understand the developmental process.

A further key assumption is that the strategic skills that need to develop in the realm of inquiry encompass a meta-level as well as a procedural level (Kuhn, 2001a,b). The meta-level includes awareness of and reflection on both task goals and one's own repertory of procedures, as well as the monitoring and management that enable these to be coordinated (Kuhn, 2001a; Metz, 2004; Reiser, Smith, Tabak, Steinmuller, Sandoval, & Leone, 2001; White & Frederiksen, 2005). Arguably, the most fundamental of these meta-level competencies with respect to inquiry is recognizing its purpose. In the words of Reiser et al. (2001), "Learners need to ground their understanding and practice of inquiry processes in an understanding of the goals and products of inquiry" (p. 269).

Our previous empirical work with developing inquiry skills in middle and upper-elementary-aged students (see Kuhn, 2002, 2005 for summary) has led us to concur with Reiser et al. (2001) that the most significant challenge beginning students face is appreciating the goals of inquiry. Simply put, a great many young students (and even some older ones) engage in inquiry activities in science classes without understanding that the activity presents an opportunity to find out something. Unless students recognize that there is something to find out, they can go through all of the motions of inquiry activity and come away with little indeed. Supporting this perspective is the empirical finding that helping students to identify a question produces significant improvement in the remaining phases of inquiry (Kuhn & Dean, 2005).

The even greater danger we have observed is that the activity is experienced as an opportunity to illustrate what one already knows. "You see, this is what I've been telling you," summarized one sixth-grade student seeking to make sense of the outcomes he had observed (Kuhn, 2002). "It's exactly what I expected." When we asked him what exactly the data before him indicated, he responded only by elaborating his ideas as to why this was what one should expect. Never did it occur to him that the data might support conclusions at variance with what he came to the activity believing was the case.

The ability to envision this possibility, and to be able to interpret what a set of data implies independent of one's prior expectations, we claim, is fundamental to productive inquiry. This claim on our part can lead to misunderstanding, since on the other hand it is often said that learning is most satisfactory when learners bring their existing knowledge base to bear on new information, attempting to connect the new to what they already know. What Stanovich (2004) calls "decoupling" or "decontextualization," (also referred to as "bracketing"), however, is essential to inquiry and indeed all scientific thinking (Kuhn, 2002). It is also, Stanovich claims, essential to good everyday thinking, supporting his claim with the example of needing to be confident that my already positive disposition toward the product or salesman will not influence my assimilation of the information he provides about the car I am contemplating buying.

In the case of inquiry activities, students need to be able to interpret what the present data indicate, distinct from what they already believe to be the case. This **"data-reading"** capability carries no implications for whether we ultimately weigh more heavily our prior knowledge or the new evidence in reaching a decision, nor does it indicate that one is more important or should be weighed more heavily than the other. Instead, this ability is meta-level in nature and involves maintaining awareness of the sources of one's own knowledge, specifically in the inquiry case being able to keep track of whether the source of a claim is one's prior understanding or the new evidence being contemplated.

Once students begin to examine new evidence in its own right and to coordinate it with their prior beliefs, meta-level functioning continues to be critical. Students must monitor and manage the strategies they bring to the task. To support this meta-level, students are asked frequently to reflect on their thinking and to **explicitly** justify their conclusions **("What told you so?").** In addition, the social context of students working together helps to externalize these meta-level processes and make them more explicit, in so doing aiding us as researchers by making them more visible.

In the elementary inquiry activities we have designed, we have focused on causal claims as the most common and generic kind of scientific (as well as everyday) claim for students to address through inquiry. Causal claims have the advantage of being cognitively accessible to students due to their familiarity in everyday experience. And causal claims are the building blocks of scientific knowledge, as well as integral to the experimental method of controlled comparison fundamental to science. Again in the words of Reiser et al., "We see that a general goal of scientific argumentation is to articulate a causal mechanism that explains patterns of data. The need to generate causal mechanisms suggests the use of controlled comparisons, because they enable us to isolate and identify causal factors" (2001, p. 271). **As Reiser recognizes, although we of course want students to learn to reason rigorously about the mechanisms that link cause and effect, development of the "data-reading" capability referred to above constitutes an essential foundation for doing so.**

It should be noted finally that in the activities we engage students in, we do not have as one of our goals to teach students particular science content or concepts. Our main goal, rather, is to help them learn how to think about and engage in science. From a pedagogical point of view, we see it as feasible to engage students in working toward this goal, in a context of scientific content, certainly, but without taking on the additional goal of their mastering specific scientific knowledge. To the extent this undertaking is fruitful, the influence of the competencies developed should be felt in other science learning contexts in which the goal of acquiring scientific knowledge is primary. This, however, remains a hypothesis that has yet to be fully investigated.

In what follows, after a description of the participants and context, we divide our report of methods and results into three major sections that describe three phases of the work (and correspond to the three years of the project). We chose this chronological form of presentation since at each phase of the work, the particular challenges that we observed students to have and the degree of progress that they made shaped the methods we employed in the next phase. We then present the results of an explicit comparison we undertook, comparing students' achievement at the end of the three years to that demonstrated by a comparison group of slightly older students who had not participated in the inquiry program. We conclude with a general discussion of what we believe we have learned.

Participants and Instructional Context

The 30 initial participants consisted of the entire fourth grade at a university-affiliated independent school in a large urban setting. The school has a unique population in that 50% of the school's slots are reserved for children of university faculty and high-level administrators, while the remaining 50% of slots are filled by children from the surrounding low-to-middle income community and chosen by lottery. The school was in its first year of operation when our work began. Hence, all students had attended a diverse range of other schools the preceding year. Students were equally divided by gender and were of diverse ethnic, racial, and socioeconomic backgrounds.

The curriculum featured at the school is an "integrated curriculum" in which common themes are identified that connect different subjects the students are studying. Teachers in all content areas indicated to us that they valued and employed inquiry as a teaching method, but review of the curriculum and classroom observations we conducted confirmed that none of the specific kinds of skills and strategies that were the focus of our activities were an explicit part of students' other coursework.

All inquiry activity except the initial individual assessment (see below) took place in the students' classroom, as part of a class that was introduced as a class in inquiry, defined for students as "ways for finding out about things." During years 1 and 2, the class met regularly, one or two times per week. By year 3, when it was reintroduced for only part of the year, on a twice-weekly basis, students were very familiar with and accepting of inquiry as a subject of study in their curriculum.

Except for occasional individual assessments, students worked in pairs or in teams of 3-4. This collaborative context we (and many others) have come to regard as productive in fostering cognitive engagement and progress, although studying the nature and role of collaborative (versus solitary) cognition is not the purpose of our work here. Whatever its benefit to those who engage in it, its additional benefit to us as researchers is one of making more visible the thinking processes we are undertaking to study.

Inquiry activities were introduced and implemented by members of our research team, who were identified to students as coaches rather than teachers. A regular teacher was often present in the room but did not become involved in the activity and sometimes left for brief periods to attend to other tasks. The number of members of our research team who were present in the classroom varied according to the size of the class and the nature of the activity, but ranged from two to four. Typically, one of us played a lead role in introducing a new activity and giving guidelines for a particular day's work. All members of the team then functioned as coaches, circulating and providing assistance as needed.

Year 1

1. Initial assessment

Method

All students participated in an initial individual assessment in the fall of year 1, using the computer simulation *Earthquake Forecaster*, one of several parallel programs we designed to both assess and provide students practice in inquiry skills. They were supervised individually by an adult who provided any guidance needed with the software.

In *Earthquake Forecaster*, students play the role of junior earthquake forecaster and are asked to assess the causal status of five dichotomous variables in contributing to level of earthquake risk (table 1). The introduction to the program explains the importance of developing means to predict earthquakes in order to protect others and maintain safety. To accomplish this, students must learn which features do and do not make a difference to risk. **The very simple causal structure of three independently acting dichotomous variables having identically**

sized effects on an outcome reflected in table 1 was chosen based on earlier work with students of this age, in which we observed the difficulty they had with this basic structure; complexities we had included in earlier work (Kuhn, Schauble, & Garcia-Mila, 1992; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995), such as multi-level variables and interactive effects, were therefore eliminated.

As seen in the text at the top of each screen (figures 1-3), students are first asked to choose what they want to find out about in their first selection of an instance (or case) to examine (see figure 1). Students identify whether they are or are not finding out about a feature by clicking the feature picture(s) corresponding to their choice(s). Then, students construct an instance of their own choosing (figure 2), by selecting the level (table 1) of each feature. When completed, these choices yield an outcome displayed in the form of a gauge representing the earthquake risk level. Students are then asked to make any inferences they believe to be justified regarding the causal or noncausal status of any of the features **and to justify them explicitly** (figure 3), or they are permitted to defer judgment. The final screen prompts the student to enter any notes they wish to (figure 4). As each screen is displayed, a voiceover presents the identical text orally, thus eliminating any challenge that reading the text may have posed for any of the students while at the same time accommodating those who prefer this mode.

All of the screens shown here are depicted as they would appear during the course of the second of the four instances the student chooses for investigation. Note that the screen includes not only the outcome for the current instance the student is investigating but also shows the outcome for the instance chosen immediately preceding this one. After answering questions regarding the outcome of the fourth instance, the student is prompted to make any additional notes desired on the final Notebook screen. The program then thanks the student for participating and shuts down.

Results

Each student's performance was categorized in one of the levels in table 2, which is drawn from earlier work in categorizing performance on this kind of task (Kuhn, Schauble, & Garcia-Mila, 1992; Kuhn, Garcia-Mila, Zohar, & Andersen, 1995; Kuhn, Black, Keselman, & Kaplan, 2000; Kuhn & Dean, 2004, 2005; Schauble, 1990, 1996), **in particular on their conclusions and the justifications they provide for them (in response to the software probe, "What told you so?").** **These codes are based on the mostly close-ended responses to the software prompts and require very little inference on the coder's part; a sample of approximately one-fourth of the entire corpus of student responses confirmed an inter-coder agreement of 100%.**

As seen in table 2, pretest performance is confined to the first three of five levels, and almost half of the students are at Level 1. At this level, students fail to appreciate that there is information to be accessed that they may want to take into account in making their judgments, **i.e., the recognition that there is something new to find out that we described earlier as a basic foundation for productive inquiry.** In the absence of this understanding, they make their claims as to which factors make a difference and which don't based exclusively on their prior intuitions.

Once students begin to attend to the evidence (Levels 2 and above), the coding scheme attempts to distinguish the forms of evidence they choose to access and the kinds of interpretations they make of this evidence. It is worth emphasizing again that the evidence-evaluation skills involved at these levels require the decontextualization, or bracketing, identified earlier, allowing the student to encode and represent the new information as an object of cognition distinguished from the student's own prior beliefs. They do not require the student to weigh the new evidence more or less heavily than prior beliefs in reaching conclusions or possibly modifying beliefs.

As reflected in table 2, the majority of students who do attend to the evidence do not get very far in drawing valid inferences based on it since they limit themselves to interpreting single, isolated cases (Level 2), thus not affording themselves the opportunity for comparison that is essential to valid analysis and interpretation. Finally, among the minority of students who do make comparisons, none incorporates the control of variables essential to valid inference (Levels 4 and 5). **In order to make the valid evidence-based justification required for this level, students will have had to generate and interpret a controlled comparison and drawn the**

appropriate conclusion of causality or noncausality for the variable examined. This sequence of levels thus provides an overall indicator of students' ability to investigate a multivariable database and draw justifiable conclusions.

2. Scaffolded practice in inquiry

Method

The initial assessment having confirmed the anticipated weakness in students' inquiry skills, we began by providing them concentrated engagement and practice with a version of this software, a method that had met with success in previous work (Dean & Kuhn, 2007; Kuhn, 2002, 2005; Kuhn & Dean, 2005) and in particular offered them scaffolding in identifying **an effective** question to serve as the objective of the inquiry**, i.e., in identifying a single variable whose role will be the focus of investigation. The scaffold embedded in the software (fig. 1) prompts students to identify a question. If after several sessions (see below), students continued to indicate the intent to investigate the effect of multiple variables at the same time (fig. 1), a coach made the additional suggestion, following Kuhn and Dean (2005), "Why don't you try finding out about just one feature?"**

Approximately one month following the pretest assessment, students began their work in pairs with the parallel *Ocean Voyage* program. This program is identical in all respects to *Earthquake Forecaster* except for content, which involves the variables that affect the success of an ancient ocean voyage across the sea. The five variables are captain's age (young or old), crew size (large or small), navigation (compass or stars), sail type (latteen or square), and ship hull shape (round or V). This content was chosen to relate to a unit on the sea that was part of the fourth-grade curriculum.

Because of student absences and other scheduling issues across the two fourth-grade class sections, the number of times a student worked on the *Ocean Voyage* program varied somewhat, with a range from 5 to 9 and a mean of 7.25. Students worked in pairs, and only occasionally alone (when an uneven number of students was present). They worked with a new partner at each session. Sessions lasted from 30 to 45 min. Sessions took place once or occasionally twice per week, depending on the class schedule, over a period of nine weeks, interrupted by a two-week school vacation midway through.

T**he software structure required students** to justify explicitly each of their **claims about the effect of a feature.** The pair of students working together **thus** needed to reach agreement on many choices and decisions, **which required that they talk about and justify** these decisions to one another **whenever there was disagreement. Although we were not able to record the dialog within pairs, it was frequent and often vigorous.**

The computer program provided students a high degree of scaffolding of the inquiry process, with the range of choices at each point circumscribed. While students had to identify a question, for example, a set of potential questions were provided for them to choose among (i.e., the five possibilities in fig. 1). Furthermore, the sequence of phases of the inquiry cycle was provided by the program structure – students did not need to generate it and keep track of where they were. We note the strong scaffolding that the program provided for purposes of comparison with later activities when it is greatly reduced. However, the two adult "coaches" who supervised each section of 15 students as they worked on *Ocean Voyage* provided minimal additional scaffolding, infrequently intervening or commenting on students' work and largely limiting their role to addressing an occasional technical issue with the software, as the students expressed considerable confidence in proceeding on their own. This changed to a degree during the latter half of the nine-week period and targeted a minority of students who appeared not to have made progress beyond a certain level for several weeks.

Three kinds of probes were introduced to students **who required further support beyond the software itself – about half of the group.** Among students whose work remained entirely theory-based, one of the coaches would occasionally pose the question, "What do the findings that you see now tell you?" **As noted above,** among students who consistently indicated the intention to find out about multiple variables at once, a coach would occasionally suggest "Why don't you try to find out about just one feature at a time." This scaffold was introduced as earlier work (Kuhn & Dean, 2005) had shown it to be highly effective in focusing students' inquiry and promoting effective investigation and inference strategies. Finally, among students who consistently designed uncontrolled comparisons, a coach would sometimes ask, at the point of data interpretation, "Is there anything else that could be making the outcomes different?"

At the end of the nine weeks, in order to ascertain that any progress students made was not limited to the particular content in which the skills developed, the *Earthquake Forecaster* pretest was readministered **individually** to all students.

Results

Progress on *Earthquake Forecaster* from pretest to posttest was substantial, as seen in table 3. As detailed in table 3, a majority of students progressed to a higher level ($p < .001$, McNemar change test). Of the 16 students who had shown some processing of the evidence at the pretest (Levels 2 or 3), a majority progressed to Level 5. Of the 14 students who had performed at Level 1 at the pretest, 11 showed progress at the posttest. Thus, at the posttest half of the students (15 of the 30) appear to have consolidated the relevant skills (Level 5) and 80% (24 of 30) have begun to attend to and evaluate the evidence by making comparisons across instances (Levels 3 and above).

How should we account for the failure to progress or even to regress on the part of the five students categorized at Level 1 at the posttest? Examination of the records showed no tendency for these five students to have participated in fewer sessions than others. Our anecdotal observations showed that these students were ones who warranted more coaching during the sessions, indicating that their posttest performance was consistent with their performance during the sessions. Thus, these were not cases of transfer failure, i.e., cases in which students performed well during the intervention but were unable to transfer their skill to new content. Nor was there any obvious association with generally low academic performance. Two students from our sample who at the end of the year were not promoted to fifth grade were not among these five low-performing students in inquiry. We thus can offer no specific explanation for these five students' failure to progress in contrast to their classmates and the possibility remains open that other methods or additional support would have led to success.

Despite the indication of significant transfer of skill development to new content on the part of the majority of students, the real test, we recognized, would come several months later, when school resumed the next autumn, and we would assess the extent to which students had maintained their newly developed skills, as we report under year 2. First, however, we report on another aspect of students' work during year 1.

3. Practice in making and justifying predictions

Method

After students had worked with the *Ocean Voyage* program for one month, the prediction/attribution module of *Ocean Voyage* was introduced as a separate "assess your skill" module. The number of times a student worked on the prediction module varied, with a range from 2 to 6 and a mean of 4.5. It required about 5 min. and was done after the basic *Ocean Voyage* program whenever the student had time available. Students had access to their notebooks if they wished to consult them.

Our intent was to engage students in the strategic component noted earlier that is not typically included in examination of inquiry – generating predictions consistent with the understanding that has been achieved. This is a task that in previous work we have found students have considerable difficulty with in its own right, **even when they have identified the causal variables influencing an outcome** (Kuhn & Dean, 2004; Kuhn, Iordanou, Pease, & Wirkala, in press). **As shown in figure 5, the task as we presented it here** requires the student to predict an outcome based on the levels of the five dichotomous variables and to indicate which variables influenced their predictions. No feedback on correctness was provided. **Students worked individually on this module in order to provide a clearer assessment of their own consistency (in the implicit causal attribution reflected in predictions, as described below) across prediction cases.**

In each iteration of the prediction module, a series of three instances is presented consecutively on the screen, each instance consisting of a particular constellation of variable levels, without any outcome depicted (figure 5). The student is asked to predict the outcome for that case (of four levels ranging from least to greatest) and then in the list of variables that follows, indicate those that affected the prediction. ("Why is this the outcome? Which feature or features made a difference in your prediction?"). Each of the variables is listed and the instruction indicates “Choose one or more.” The second and third predictions involved new instances but the format remains identical. At subsequent sessions, students were given the general feedback that their previous predictions had not been entirely correct and that they should keep working on their prediction skills.

Results

Our interest in assessing students’ multivariable prediction skills was on the reasoning processes that produced these predictions, rather than the correctness of predictions. In particular, previous work with this type of prediction task (Kuhn & Dean, 2004; Keselman, 2003) led us to focus on two characteristics. First, how many of the variables student identified as causal in the main task entered into their predictions? Second, within a given prediction session, was the student consistent in the variables to which causality was implicitly attributed (in response to the question, “Which feature or features made a difference in your prediction?”)?

We consider here the 19 students who achieved levels 4 or 5 on the main *Ocean Voyage* task. All had successfully identified the three causal variables as causal and the two noncausal variables as noncausal in their work on *Ocean Voyage*. To what extent could they use this knowledge to make predictions based on the effects of multiple variables?. (The remaining 11 students, who did not have this knowledge in place, might be expected not to do as well in predicting outcomes due to lack of knowledge of the causal effects and therefore must be considered separately.)

We focus on the last three sessions of students’ work on the prediction module, so as to allow the first one or two sessions with the prediction task (of the earlier-reported average of 4.5 session that students spent with this task) for the student to become familiar with it, thereby eliminating task unfamiliarity as a possible source of inconsistency in judgment. Inconsistency can be shown for any number from 0-5 of the five variables presented in the task. For causal variables the mean of the individual means (of number of variables for which inconsistency is shown) is 1.42 (of a possible 3). For noncausal variables the mean of individual means is .61 (of a possible 2). Thus, inconsistency tends to be more frequent when the variable is causal (almost half of the time, versus less than a third). Nonetheless, inconsistency in causal attribution, we see, remains a significant limitation with respect to both causal and noncausal variables for a majority of these students, despite the progress they have made in developing the inquiry skills that allow them to identify causal and noncausal effects in multivariable data. Comparable levels of performance were observed on the part of the 11 students who did not meet the criteria indicated and were not classified in the successful category. Hence, students who successfully identified the causal and noncausal effects present in the data available to them did not show any greater consistency in causal attribution in the prediction task, as a result of their achievement with respect to inquiry skills, than did students who did not show this achievement.

One other possibility to note is the possibility that inconsistency in causal attribution in the prediction task is attributable to a concern on the student's part with interaction effects. Considerable evidence is now available, however, that students of this age do not conceptualize interaction effects among variables (Kuhn et al., 1995; Kuhn, 2002; Zimmerman, 2000). A concern about the possibility of such effects was never voiced during students' *Ocean Voyage* investigations, and it is hence unlikely that a concern about interactions among variables had a detrimental effect on their predictions. In any case, it would not account for inconsistency from one prediction to the next in their implicit causal attributions (i.e., the variables they implicated as influencing the prediction).

The other major question we wished to ask of the prediction-task data is the extent to which students' judgments successfully incorporate the roles of all three causal variables. We look here at their performance only at the final prediction session, when their knowledge and skill should have been at its maximum. Again, despite the fact that the 19 students considered here had by this point all successfully identified the three variables that had causal effects on outcomes and the two that did not, underattribution of causality remained a significant constraint on their causal reasoning. Among these 19, only seven consistently implicate all three causal variables as having contributed to the outcome. Of the remaining 12, four implicate a median of two as causal (over the three prediction judgments that constitute the session), and the remaining eight implicate a median of one as causal. (Results were similar for the 11 students not in the successful category.) When the two characteristics, inconsistency and underattribution, are combined, what emerges is a model of prediction in which the explanatory burden in a multivariable context shifts from one single variable to another single variable over time.

Discussion

How do we explain the failure of students in the present work to fully utilize the causal knowledge they gained in the investigation and inference task in performing the prediction task? Why are the explicit causal attributions they made in the first task not reflected consistently in the second task, in which they are called on to apply this very same causal knowledge? We know that children younger than the age of those in this study can under certain conditions appropriately integrate information from at least two sources in an additive fashion (Anderson, 1991; Wilkening, 1982; Dixon & Tuccillo, 2001), so their failure to do so in the present causal context cannot readily be attributed to processing limitations, If it were, they might have simply ignored one of the variables and focused on integrating the other two. The results, however, do not support such a model.

We suggest that coordination of the effects of multiple variable to predict outcomes is a critical component of scientific thinking – certainly in terms of its practical import – but that it appears to be a distinct skill from that of identifying these effects by means of controlled experimentation, the set of skills typically emphasized under the heading of inquiry skills. Clearly, the present findings show, attainment of these fundamental inquiry skills entailed in identifying causal and noncausal effects in multivariable data is not sufficient to ensure ability to make multivariable predictions based on the knowledge these skills yield. We return to these points later.

Year 2

1. Further developing and consolidating skills of investigation, inference and prediction

Our objective in year 2 was to support further skill development within this cohort of students in a two-pronged effort, one devoted to consolidation and enhancement of inquiry skills and the other to further development of multivariable prediction skills. Twenty-eight of the 30 students we worked with in year 1 continued to the fifth grade and participated in the inquiry program; one student is omitted from the year 2 analysis due to prolonged absence. The now-fifth-grade students were joined by 7 new students who entered the school's fifth grade at the beginning of the school year.

Method

We introduced a new program, *Avalanche Hunter,* to maintain students' interest, initially identical in structure to *Earthquake Forecaster* and *Ocean Voyage*, but later elaborated **to incorporate effects of variable sizes (specifically, one variable had twice as large an effect as the others).** Wind type, snow type, cloud cover, soil, and slope are the five binary variables having potential causal effects on avalanche risk. Students attended a 40-min class that met twice a week for most of the school year. The class was introduced to students explicitly as a class in inquiry. The year began with individual assessments, after which students did activities in pairs or occasionally small groups.

Initial individual assessment. We were interested not only in the extent to which students maintained their skills from the preceding school year but also in the extent to which they applied these skills consistently, a dimension that the year 1 assessments captured in only a limited way as they were confined to a single session. We thus asked each student to engage in two sessions of individual assessment with *Avalanche Hunter*, during the first two class meetings of the school year. (The prediction module was not introduced until later, after students had worked with *Avalanche Hunter* long enough to have identified the causal and noncausal variables.)

Inquiry practice. At sessions following the two initial sessions devoted to individual assessment, students began working in rotating pairs on *Avalanche Hunter*. The coach introduced the task as similar to ones returning students were familiar with from last year. Students were divided into two working groups that met in adjacent classrooms, one consisting of students who had showed a predominance of valid strategy use in the pretest and the other of those who had not (to allow the respective agendas to be differentiated if it proved necessary). New students were in the latter working group, and for the first few sessions were not paired with one another.

Two components were added to the activity of both groups, both ones that had not been present in year 1. One was the introduction, after several weeks of work with *Avalanche Hunter,* of "claim sheets" designed to enhance students' awareness of the forms of evidence that support a claim. Pairs were instructed to complete a claim sheet as soon as they were certain that they had determined the role of a particular feature. **A claim sheet prompted the student to enter a claim at the top of the sheet: "______(feature name)_______ MAKES A/NO DIFFERENCE." (The student is instructed to cross out "A" or "NO."). A second prompt read simply "EVIDENCE:" below which the student was instructed to indicate the evidence that provided the basis for the claim.**

Claim sheets, when completed, were placed by the pair in the appropriate one of five large brown envelopes secured to a class bulletin board, each labeled with one of the five features. This practice continued across sessions, with the sheets and envelopes available at each session. Pairs were encouraged to go to the envelopes and review their classmates' claim sheets once they had contributed one of their own. At a point at which all pairs had had the opportunity to complete a claim sheet for each variable, one of the coaches took the sheets out and reviewed them with the class. She noted to the class that most pairs had come to the same conclusion regarding whether that feature did or didn't make a difference, even though they had not done exactly the same experiments to reach that conclusion. **All the claim sheets in the "Slope Angle" envelope, she noted, for example, claimed that this feature made a difference, and all had based their claim on evidence comparing outcomes for steep and gentle slope angles. But some had made this comparison under conditions of low cloud cover, others high cloud cover, and so forth for the remaining features.**

The second new component added to the activity was the introduction of a probe by the coaches, delivered first to the class as a whole and then later reinforced on an individual basis if a coach observed a student exhibiting the concept confusion in question. Inclusion of this probe was motivated by our examination of patterns of performance in students' year-1 work on the prediction module. This examination revealed a pattern observed on the part of the mostly low-performing students on this task who indicated that a variable contributed to the prediction only if its level was the one associated with higher risk. If its level was the alternative one, the student indicated that the variable played no role in the prediction. This was the case, note, even though the student in the investigatory phase had isolated the variable as "making a difference." This approach, note also, contributes to the inconsistency in causal attribution across predictions that was observed.

We decided that this was a conceptual confusion that could be addressed directly and we therefore introduced it as a "reminder" presented by one of the coaches, that if a feature makes a difference it always makes a difference, whether the level is high or low, present or absent. An absent feature, the coach noted, has as much to say about what happens as a present one**. In other words, for example, a steep slope increases the risk we would predict (compared to having no information about slope angle), but in just the same way a gentle slope decreases the risk we would predict.** Although we anticipated this intervention to have its major effect on performance in the prediction module, the intervention was confined to sessions involving the main investigatory activity, since it was pertinent to the fundamental question of what it means for a variable to "make a difference."

Work with *Avalanche Hunter* continued **twice weekly** from late October to early December **(12 class sessions),** by which time almost all students had achieved a high degree of mastery of investigative and inference strategies, although, as detailed below, they still showed less than 100% consistent optimal strategy usage.

Prediction practice. Once students had made claims regarding all five features, they began the prediction module. Some students moved on to the prediction module sooner than others, but all students undertook predictions on multiple occasions, with every student completing between 6 and 13 reiterations of the prediction module **(typically two per session).** One other change in procedure from year 1 was to have students undertake their prediction work in pairs. This procedure, which had worked well in the investigation sessions in getting students to articulate their thinking, we anticipated might enhance the need to justify predictions to one another and decrease the likelihood of ignoring one or more of the effective variables. Students' solitary prediction work in year 1 having enabled us to eliminate peer influence in assessing inconsistency, we wished to introduce peer collaboration for its potential beneficial effects. Although we cannot make a clean experimental comparison between the novice year 1 group and the largely identical group revisiting the task in year 2, anecdotally the social condition appeared beneficial. Pairs did engage in frequent debate regarding their predictions, often revising them during discussion. **One member of the pair, for example, would initiate a risk-level decision for the case, which the partner questioned ("Why do you say that?"), prompting the pair to engage in discussion regarding how the different feature levels influenced the prediction and, typically, in the process, to revise the prediction at least once and sometimes repeatedly.**

Introduction of variable effect size. Following winter vacation, when inquiry classes resumed in mid-January, a new form of *Avalanche Hunter* was introduced, one in which one variable (cloud cover) had twice as large an effect as the other causal variables, and students were asked to indicate whether any of the variables were more important than any others. (Students were cautioned that these were a new set of findings for a different location and that the results might not be the same as those they had observed earlier.) Students showed little difficulty in applying their skills to this modified problem and recognizing that effects need not all be of equal size. **In other words, they concluded, "cloud cover makes more of a difference than anything else – you really have to pay attention to it."**

Individual year-2 posttest assessment. Following completion of the double-effect *Avalanche Hunter* problem, in March there took place for all students a phase of individual assessment, returning for this purpose to the simple form of *Earthquake Forecaster* and the associated prediction module. The purpose was to assess how much progress each student had made individually, in the absence of peer influence. Students individually required between two and three sessions to complete two cycles (both investigation and prediction modules).

Results

Inquiry skills. On the year 2 pretest, of the 27 students continuing from the previous year, 10 performed at ceiling (consistently performing controlled comparisons and drawing valid inferences at both sessions), 9 showed some controlled comparison and valid inference (but did not show these strategies consistently), and 8 showed no valid strategy usage. Performance of the group as a whole thus did not present as strong a picture as had been the case at the end of the preceding school year, although skills were largely maintained. Unsurprisingly, the 7 new students showed negligible skill (2 of the 7 made one valid inference each).

At the mid-year-2 individual posttest, performance was strikingly improved. All 34 students used controlled comparison and made valid inferences a majority of the time. Students fell into the following groups:

1. 15 students: Exclusively optimal strategies, across all sessions with complete consistency.

2. 8 students: Optimal performance at two of the three sessions, deviating at only one session either by making an uncontrolled comparison or failing to make an appropriate inference (most often only one of these two). (In 5 of these 8 cases, the student had been a consistent high performer and the student's performance at the initial posttest session was consistently at ceiling level and deviated only at the second or third session, when we suspected the student may have begun to consider interaction effects.)
3. 11 students: Mixture of valid and invalid strategies, although all of these students used valid strategies more often than invalid ones.

The seven students who had not begun their inquiry activity until fifth grade were roughly equally distributed across the above three groups, rather than overrepresented in any subgroup.

Performance of students on *Earthquake Forecaster* at this posttest, it should be noted, was overall equivalent to their performance on the parallel *Avalanche Hunter* that they had been working on immediately prior to the individual posttest. Over the three-year intervention, *Earthquake Forecaster* until the final few weeks was used only as an individual assessment tool and hence students had less practice with it than they did with *Avalanche Hunter* in year 2 or *Ocean Voyage* in year 1. They nonetheless showed no difficulty in transferring the skills they had developed over time (and in the social context of pairs) to this less familiar individual assessment.

Prediction skills. Here, the portrayal of attainment is much more modest. We focus on the 18 students reported earlier who showed mastery of investigation and inference strategies in year 1, expecting that having these strategies in their repertories would support the development of the multivariable prediction strategies that remained weak at the end of year 1.

This group (now reduced from 19 to 18 in number, due to attrition) did show improvement by the end of year 2, compared to their performance at the end of year 1. The mean number of variables (of 5) for which these students showed inconsistency (in causal attribution) at the individual assessment at the end of year 2 was 1.49 (down from 2.03 at the end of year 1). Eleven of the 18 continued to show some inconsistency (for causal variables, noncausal variables, or both) at the year 2 assessment. Only 7 students were perfectly consistent in their causal attributions. (Performance of students not among the original 18 was even lower and showed no improvement from year 1.)

The other aspect of understanding additive multivariable causality that we examined in year 1 is the number of causal variables that students successfully integrated in making their prediction judgments. In this respect we saw continued underattribution of causality, with little progress from year 1 to year 2. Among the 18 students, 7 correctly implicated all 3 causal variables in their prediction judgments, 5 implicated an average of two variables and 6 an average of less than two variables. This performance is equivalent to that reported at the end of year 1 for these students. (Again, performance of students not among the original 18 was even lower and showed no improvement from year 1.)

Discussion

Even allowing for some inconsistency due to inattention or emergence of a new idea during a prediction session, these figures reflect less than solid mastery of this fundamental aspect of scientific understanding involving effects of multiple variables – that (barring interaction effects) causal effects operate in a consistent manner across occasions and multiple causal effects may operate simultaneously. Despite their progress in other inquiry skills and the substantial practice they had engaged in, a majority of students continued to show significant weaknesses in applying the knowledge they had gained in their investigatory work to situations in which application entailed integrating distinct units of knowledge that had been acquired individually. These weaknesses, we believe, warrant further attention and investigation.

Our data do confirm, however, that extended engagement with problems involving inquiry is effective in developing rudimentary inquiry skills – notably, the skills of identifying an addressable question **(the causal role of a specific feature),** seeking informative data via controlled comparison, and drawing appropriate conclusions of causality and noncausality – in fifth-grade students. (Because their number is small and because we don't know if they would have done as well out of the company of their more experienced peers, we cannot draw any firm conclusions regarding the seven students who began their inquiry activity as fifth graders, but their performance suggests the possibility that initiation at this age may be as productive as beginning earlier.)

At the same time, despite the skill students displayed, our findings indicate that many students continue to show variability in strategy usage long after the more effective strategies have appeared in their repertories – a finding consistent with the sizeable body of microgenetic research (Kuhn, 1995; Siegler, 2006) in which variability has been found to be the norm rather than the exception. These findings contradict any model that posits skill acquisition at a single point in time, following which the skill can be assumed to be in place.

The nature of the successful and unsuccessful strategies that young students apply in very simple inquiry contexts like the ones we employed has been described in detail in our own (Kuhn, Schauble, & Garcia-Mila, 1992: Kuhn et al., 1995) and others' previous research (Schauble, 1990, 1996; Klahr, 2000). **Elementary-school students have great difficulty in designing effective investigations and in utilizing their results in ways that are clearly distinguished from expectations and are informative. These skills require practice in order to develop and remain securely in place. In the present work, we devote special attention to the aspects of inquiry that precede and follow the design and interpretation of experiments, especially the identification of inquiry goals (Kuhn & Dean, 2005) and the application of conclusions in new contexts (for example, in prediction). We are most interested in how all of these rudimentary inquiry skills can be built upon and used by students in effective ways – ways that go well beyond mastery of the control-of-variables strategy – and the challenges that are encountered in realizing this potential.** Hence we postpone more detailed description and illustration of students' approaches until we reach the latter phases of the work.

2. Introducing more complex forms of data: probabilistic and interactive effects

Method

Following the posttest assessment, students remained in the two working groups noted earlier (under Year 2 *Inquiry practice*) for the remainder of the school year. The individual posttest confirmed that the more advanced group had mastered the foundational skills involved in investigation and inference, having been classified in subgroups 1 or 2, and were ready to move on, while the remaining students we believed could benefit from more practice at their current level. (Two students switched working groups at this point, to better fit these categorizations.)

To maintain their interest, we introduced different content – the *Ocean Voyage* program from year 1– to the lower performing group, and they continued their work. The higher performing group also worked with *Ocean Voyage*, but in their case a more advanced version of *Ocean Voyage* was introduced. It was the first of two further problems presented during the final months of year 2, both of which represent elaborations of the structure of the database. Other aspects of the program remain the same. **Students worked on one of these forms of *Ocean Voyage* twice weekly for the remainder of the school year, from early March until the end of May.**

These elaborations enhanced the challenge of students' inquiry by introducing more complex forms of evidence. Both of these enhancements are important, we believe, in more closely approximating the kinds of data students are likely to encounter in more naturalistic investigatory contexts. Data in such contexts, as well as involving multiple variables, are likely to be both probabilistic and interactive, rather than yield to any simple, determinate solutions.

Probabilistic effects. The first more complex problem we presented introduces the probabilistic characteristic. The outcome for a particular constellation of variable levels is not constant but rather takes the form of a distribution with one outcome (a particular voyage distance) most frequent but adjacent outcomes of lesser and greater distance also occurring but with lower frequency. (Specifically, the variable of captain's age - young or old – is associated with a distribution of outcomes, rather than a single consistent outcome. The most frequent outcome - 60% of instances - is level 1 for the young captain and level 2 for the older captain. However, in 20% of instances, the young captain yields a level 0 outcome and in 20% a level 2 outcome. Similarly in 20% of cases the old captain yields a level 1 outcome and in 20% a level 3 outcome.) Thus, comparison of any two instances may be misleading. Instead, students must generate multiple instances and compare these distributions (e.g., for young and old captain) in order to identify the effect.

Continuing to work in pairs, students initially had a great deal of difficulty with the probabilistic problem. Often, the comparison of two specific instances yielded a difference in outcomes, but then if repeated did not, paving the way for students to recognize that the simple controlled-comparison strategy that had served them well to this point no longer yielded consistent results. Students were left with the confusing situation that sometimes a variable appears to make a difference and sometimes does not – a conclusion that many of them initially drew. After working on the problem for three class sessions, students appeared discouraged, and several objected to the coaches that this problem was impossible to solve.

At this point, one of the coaches made a suggestion. Difficult problems, they said, sometimes can be tackled if the problem-solving workload is divided up. The coaches suggested that the class divide into five teams of 3-4 students each. Each team was then asked to focus on analyzing the role of a specific feature. The choice of feature by each of the teams was agreed, so that each feature was covered by one of the groups. Each team was then coached to concentrate just on their feature and determining whether it made a difference. It was suggested that since they weren't getting the same results each time, they should perhaps run multiple trials of a given instance and possibly a pattern would emerge**. Students readily followed these suggestions and soon began to make progress on the task and feel more capable in engaging it, with all students experiencing success in identifying the effect of the feature their group focused on within a session or two. In a final whole-class discussion, one of the coach reminded students that they would not have been able to reach these common conclusions had they compared only single trials of each variable combination.**

Interactive effects. The interaction problem was introduced to all students during the final few weeks **(4-6 sessions, depending on group)** of the year-2 school year. It was presented in the context of *Earthquake Forecaster* and entailed an interaction between two of the three causal variables. (The interacting variables are snake activity and gas level. Snake activity has an effect on risk only when gas level is heavy.) Students again worked in pairs, as previously, within the two work groups. This time, however, **from the outset** coaches encouraged students to compare their conclusions with those of others. This led to the discovery of discrepant conclusions. **In other words,** as in the probabilistic problem, sometimes a variable seemed to make a difference but other times it did not, **depending on which of the specific levels of each of the variables was examined.** Coaches suggested to students that they "double check" conclusions about the effect of a variable, trying it out to see if it still held when a different level of another variable was employed. Students worked with the interaction database up to 4 times depending on the time available to them. (A number of students had commitments outside the classroom at this time of year.) This amount of time proved inadequate for most pairs to make much headway with the problem and we therefore postponed further work on interactions to year 3.

Results

No further assessment of individual achievement was undertaken in year 2, beyond the spring assessment of investigative and prediction skills described above, as we intended to continue work with both probabilistic and interactive effects in year 3 and therefore felt it would be premature. Although we judged the time spent with interaction effects inadequate to draw any firm conclusions, students' work with the probabilistic problem was revealing. Although methods varied somewhat across teams, each of the teams was successful in recording the distribution of outcomes that a particular instance yielded over repeated trials and comparing that distribution to the distribution produced by an appropriate comparison instance, in order to reach a conclusion regarding the causal or noncausal status of the feature they were investigating. All relied on some at least intuitive measure of what the central tendency was for each distribution. **Most students were satisfied with a characterization of central tendency in terms of the mode: "This is the risk level you get most of the time for this case." A few, however, tried to characterize the set of outcomes (for a particular variable constellation) in more precise quantitative terms, noting the range and/or undertaking to calculate an arithmetic mean or median.** (The concept of mean, **we were informed,** had been introduced in their math class.) They then went on to compare whether the central tendency for one variable level differed from that for the other variable level. **Coaches made no suggestions regarding record-keeping method, which tended to involve making a simple list of outcomes, not always well labeled for future reference. (This component of the inquiry process undergoes further development during Year 3, as we shall see.)**

This activity extended through four class periods, with most of the final period devoted to each group's presenting the conclusion and supporting evidence for their feature. Interestingly, one of the groups had had time to go on to investigate a second feature, and their conclusion regarding this feature conflicted with that of the other group working on this feature. This discrepancy required scrutinizing the respective supporting evidence of each group and resolving the discrepancy, an undertaking not only the two groups but the entire class became engaged in.

Although we felt unable to draw any firm conclusions regarding students' work on interactions, a few pairs in the more advanced work group did appear to make some headway with the problem, voicing an "it depends" conclusion regarding whether a particular feature made a difference. The typical conclusion, however, observed both in students' conversation and on their claim sheets, was that a feature "sometimes" makes a difference and sometimes doesn't **– a conclusion that, to our considerable concern, appeared to cause them no concern whatsoever.** We therefore resolved to continue work on interactions during year 3.

Year 3

The objectives for year 3 were multiple. First, we wished to continue engagement with more complex databases involving probabilistic and interactive effects, especially as time had been inadequate in year 2 for investigating interaction effects and not all students had encountered probabilistic effects. This time, however, students worked with a database that included both kinds of effects (probabilistic and interactive) at once.

Second, we wanted at this point to experiment with allowing students more freedom in designing and executing their own inquiry, hopefully without compromising their skills. Had the foundation we had provided in Years 1 and 2 equipped them to independently engage in their own inquiry without the scaffolding provided either by the software or adult direction? In addition to exploring it here, at the end of year 3, as reported in the final section, we conducted an explicit test of this question, comparing the inquiry students with a slightly older group who had not participated in the inquiry program.

Third, we wanted to explore integrating the activity with the students' school curriculum and assessing how this might affect their performance. Are the same skills and challenges evident in this more content-rich context?

The time available to work with students in year 3 unfortunately was significantly reduced over what it had been in year 2. The school began a program to prepare upper-grade students to be competitive for entrance into selective high schools, as a result of which greater emphasis was placed on improving the skills assessed in standardized tests. As a result of the reduced time we had to work with students, we limited our objectives to the three indicated above and postponed a fourth – designing and implementing activities that would further develop the multivariable integration and prediction skills we had found to be still weak. Despite the importance of better understanding these skills and the challenges they pose, we thought it most important to examine **the feasibility of affording students greater independence in their inquiry activities and of integrating these activities with the academic curriculum.**

During year 3, our analytic focus shifted entirely to the group level, and we did not conduct individual skill assessments. We were more interested by this point in how students are able to apply their skills and in examining skill levels exhibited by the group as a whole and what kinds of activities the class would be successful in engaging as a function of the skills they had developed.

At the beginning of year 3, the school implemented a planned increase in the size of the sixth-grade student body, while a few students left the school at the end of fifth grade. As a result, we worked with the 39 students who made up that year's sixth grade, 11 of whom were new to the school. In composing the **nine** four- or five-person teams described below, we took care to integrate new students, such that the majority of teams had only one new student. (Three teams had two new students; one team had no new students.) We did not do assessments specifically to ascertain skill levels of the 11 new students, compared to those exhibited by the 28 continuing students, and focus our analysis, as noted, on the skill levels exhibited at the group level. Yet we observed no indications of the new students having difficulty being absorbed into their respective teams and participating in the activity or certainly of undermining the group work.

Method

The cart problem

Students at this school studied an "integrated" curriculum – one in which the same themes extended across all of their classes. During one trimester of the sixth-grade year, the theme was the Renaissance. Many different aspects of this theme were explored. In their humanities classes, for example, students read and discussed, viewed a professional production of, and acted out scenes from *Romeo and Juliet.* An aspect of the Renaissance theme most directly relevant to inquiry was the idea of the Renaissance as a time of exploration and innovation. Religious and government leaders were no longer accepted as the ultimate authority on all matters, and the very idea of empirical investigation to produce knowledge gained acceptance.

A particular aspect of experimentation and discovery that students focused on in their science classes was a study of biomechanics that focused on the design of simple machines during the Renaissance period and the manner in which human and machine interacted to form a biomechanical system. Based on a book of drawings of Renaissance machines, housed in the Smithsonian Institution but available on the internet, we adapted a depiction of a primitive cart for transporting materials, producing the sketch of Rafael and his cart that appears in figure 6.

As part of their Renaissance curriculum, over a three-week period students spent five 55-min class periods engaged in a problem we designed having to do with Rafael and his cart. The problem presented to students was one of Rafael and his helpers having to clear a pile of stone from a location so building could begin there. Rafael undertook to experiment with how different features of the carts he had available to use for the task affected their efficiency in getting the job done. The four features he varied in his experiments, as shown in figure 6, were the handle length, the wheel size, the bucket size, and the bucket placement. The specific outcome variable was the number of stone-moving trips Rafael could make in a one-hour time period. **Although students initially showed some tendency to search for "the fastest cart," consistent with an engineering rather than analysis orientation (Schauble, Klopfer, & Raghavan, 1991), coaches reminded them if necessary that the goal was to find out the effects of the various features on efficiency. Although this might have proven a challenge for less experienced students, these students in their third year of inquiry class exhibited little difficulty adopting the appropriate orientation.**

Consistent with the relevant physical principles, wheel size has no effect but the other three variables do, two of them interactively (handle length makes a difference only in the far bucket placement, with the long handle superior; it has no effect with the near placement, which is always superior to far). A probabilistic element was introduced, designed around the bucket size variable (which is causal, with smaller bucket superior) i.e., 20% of outcomes for each bucket size are increased one level (from what they would be based on straightforward addition of all effects) while another 20% are decreased one level. Hence, effects can only be identified by executing multiple trials for a given instance to ascertain typical (most likely) outcome.

Students worked on the problem in nine teams of four (or occasionally five, depending on attendance). The data were available to them via a computer interface that allowed them to select a level of each of the four variables and then view the outcome for that particular instance. This program, it should be emphasized, provides much less structure in comparison to *Earthquake Forecaster* and the similar programs that students had worked on as fourth and fifth graders. In contrast to those simulations, the cart program functioned simply to make the database (of instances and outcomes) available. It was up to the team to decide how to proceed in choosing instances to examine, keeping data records, and making inferences. Once the feature levels were selected, an outcome could be generated immediately, and students became adept at designating feature levels and generating an outcome in under 10 seconds.

Procedure

The only structure imposed on the team's work was one of suggesting roles for the four team members. One member of each team was chosen to be leader. The team then chose which members would fill the remaining roles, which were described to the students. One role was that of scribe. The scribe had a laptop computer available with a word-processing file to keep notes that carried over from session to session. A third student was the data manager, in charge of accessing the computer program (on a different laptop) and executing the group's instructions regarding the instances they wished to call up and examine. The fourth student was the evaluator, whose job it was to watch over everything that was going on and to make suggestions for improving the group's work. These roles were not monitored once the group's work began, however. All teams adhered to them to some degree, but there was variability in how strictly they were adhered to. Our anecdotal observations, however, did not suggest that strict adherence to these roles was a factor in teams' success**. Some teams, we observed, worked very productively with more flexible roles that varied across time.**

After one of the adult coaches introduced the problem, using figure 6, the teams were asked to generate hypotheses regarding the effects of the different variables and then to agree on a plan for their work. They were reminded of the value of keeping records of their work. Once they were ready, they began to access data. During the teams' work, adult coaches limited their role to one of asking questions when a group seemed to lose focus or need direction. The following are representative of questions and suggestions made by the coaches:

Talk to your teammates.

What was the question you were asking?

What is your goal?

Do you agree with one another?

What are you finding out now?

What are you comparing?

Is there anything else important here?

Would you always get that result?

What should you do next?

What have you found out up to now?

Are there any conclusions you can come to?

Are you satisfied with that conclusion?

Are you entirely sure of it?

Each of these prompts proved effective on various occasions in furthering a team's progress. An additional form of prompt from coaches was warranted when a team prematurely indicated that they had solved the problem, i.e., that they knew whether a variable made a difference or didn't make a difference, after comparing only two instances. In this case, the coach cautioned them not to "jump to conclusions" and to double check their results. This prompt led them to discover the variability in outcomes and then to develop a way to characterize the distribution of outcomes for a particular cart. Means, medians, and modes had been a topic in their math class, **and we observed some teams employ them.** If asked, coaches suggested the mode as the most useful of the measures of central tendency for the present purpose.

A related challenge occurred when students prematurely concluded what the effect of a variable was, having examined it in relation to only one constant constellation of the remaining variables. In this case, the coach asked them whether they could generalize, i.e., would you always get this result at all levels of the other variables? This question often led them to seek replication of their findings at other variable levels.

Beginning to examine interaction effects, however, often created the further challenge of retaining focus on the appropriate variable. For example, if undertaking to examine the effect of bucket size at a different wheel size, once students began to replicate the bucket size effect at the new wheel size, they sometimes lost focus and began to contemplate instead the effect of wheel size. In these cases, the coach attempted to restore their focus with a "What is your question?" probe.

At the next-to-last session, students were told they should conclude their investigations at this session as the final session would be devoted to writing a research report on what they had found. These reports, they were told, were to be given to their classroom teachers for evaluation (which they were). An outline was posted on the classroom Smartboard as a basis for the research report. It contained four main headings, listed sequentially, and several subheadings: (a) Purpose? (Hypotheses?); (b) Method? (c) Conclusions? (Evidence for conclusions?); and (d) Comments (Hypotheses correct? Explanations for findings? Other comments?)

To ensure that all students engaged actively in the report writing, each team was divided into two sub-teams of two (or occasionally three) students each, with each sub-team writing an independent report based on their team's work. The plan was to have the two sub-teams exchange reports and compare them, but this plan proved impossible to carry out as the report writing took a full class period and no further class periods were available for this activity.

Results

Although there was much else of interest in our videotapes of the teams' work and the cumulative scribe notes of each team, the focus of the analysis we present here is on these 18 research reports as the culmination of the group's multi-year progress in developing inquiry skills and hence indicative of the level of skill the group had achieved. This was especially the case, we believed, as students received minimal instruction regarding how to write these research reports and virtually no guidance while they were writing them.

A few of the pairs, notably those who included a great deal of detail in their reports, indicated they had not finished the report when the class period ended. Accordingly, we do not focus on the exhaustiveness of students' reports (in the sense of addressing the effects of all of the variables), but rather on what the report does contain, especially regarding how the group saw as the purpose and goal of the activity, how they went about identifying effects of variables and the manner in which they supported the claims they made. Analysis of these reports identified an array of strengths and weaknesses that was in many ways surprising. These are summarized in table 4. **The skills and sub-skills that appear there were identified, based on analysis of a portion of the reports, as exhaustive of the inquiry skills we saw exhibited in the reports. This analysis was then extended to the entire set independently by two researchers. (A negligible number of disagreements were resolved by discussion.)**

We begin discussion of table 4 with a few comments of clarification. **First, the level of analysis is the research report prepared by a pair of students; hence we cannot draw conclusions regarding individual skill.** Regarding the three method skills (replication, comparison, control), all students did in fact demonstrate all three skills in their work, although, as reflected in table 4, not every report included reference to each of these skills as a part of their team's work method. On the few occasions when an uncontrolled comparison was made, members of the team quickly noticed and corrected the error (an occurrence that was recounted in one report). Hence, we do not regard the less-than-maximum numbers in the table as reflective of significant weaknesses in the groups' mastery of these basic inquiry methods. A final comment: in the last row in table 4, pertaining to concern about interaction effects, students did not have to actually discover the interactive effect (although 5 of the 6 reports did specify it) to be classified as "adequate" in this category; it was sufficient to have voiced an awareness that the effect of a variable might not be replicable at other levels of the remaining variables.

What are the most important findings to come out of table 4? We see that students continue to exhibit occasional weaknesses in clearly differentiating hypotheses (explanations of why an effect might be expected) and evidence, and in coordinating the two. Yet the two most notable weaknesses reflected in table 4 are the relatively low number of reports that included awareness of potential interactions (33%) and, more important, the low number that included an indication of the objective of the activity or the question to which an answer was being sought. Of these two weaknesses, the low awareness of interactions is less surprising. Previous work (Kuhn et al., 1992, 1995) indicated that even after extended investigation students of this age, and even young adult students, infrequently considered the possibility of interaction effects. In the present work, we sought to prompt students' consideration of such a possibility (Would you always get that result?), but it did not always lead to further exploration. Even though all students came to understand that replication of the same instance produced variable results, they did not all take the next step to recognize that replication of the same effect (i.e., the same *relationship* across two sets of instances) might not always produce the effect.

Interestingly, even when a team did include exploration of possible interactions in their inquiry, this exploration did not always make its way into the reports. Among the four (of nine) teams who did engage in such exploration, in two cases both of the sub-teams included discussion of interaction in their reports. In the other two cases, even though both of these two teams had in fact identified the interaction during their investigation, only one of the respective sub-teams in each case reported on it. The other sub-team reported on the individual variables as main effects. During their work itself, as noted earlier, students found it challenging to maintain their focus on the appropriate variable while they were investigating its effects at different levels of other variables. All of the four teams who considered interaction effects in fact exhibited this wavering of intent at some point.

While we expected interaction to remain challenging, an unanticipated **and we believe critical** finding was how often students failed to include in their reports any indication of the objective of their investigation. In the brief outline described earlier that we provided students as a guide for their reports, "Purpose" was the first heading listed. Moreover, in the more structured context of the computer simulations that these students had worked on the two preceding years, the question(s) that were to be the objective of the activity had to be explicitly identified by the student at the outset of each cycle, before the student could proceed. Yet, in the present reports fully half of the reports overlooked the "Purpose" cue in the outline, even though almost without exception the reports included content relevant to each of the other headings in the outline. Although many reports simply ignored the "Purpose" cue, a few made unsuccessful attempts to address it. These examples of unsuccessful attempts, shown in table 4, offer insight into the conceptual challenges that this aspect of inquiry entails. Even though it was addressed and seemed to disappear during the inquiry itself, notable is reemergence of the engineering focus on the part of a few pairs in their reports of the activity – the objective is to produce the best outcomes, rather than understand the factors that underlie outcomes. Even among those table 4 examples that do not revert to an engineering focus and maintain an appropriate analysis focus, however, what is notable, and what unites these unsuccessful examples, is a failure to identify the variable, or feature, as a conceptual unit warranting investigation and analysis.

This finding is consistent with the claim we made at the outset of this paper regarding the importance of students' achieving and maintaining awareness of the objective of inquiry activity. But we appear to have underestimated the continuing challenge that this awareness poses. We cannot, of course, say that those students who did not articulate this awareness in their reports lacked any such awareness (and it is unlikely that they did). Yet the reports they produced underscore that such awareness is fragile and warrants continued support. We return to this point in our general discussion.

On the positive side, students' reports reflected two important accomplishments. First, as an overall group students showed themselves quite able to conduct effective investigations in the absence of the strong scaffolding they had experienced in Years 1 and 2. They had mastered essential elements of scientific method, including replication and control of variables, and seldom showed any deviation from them or were quickly corrected by peers if they did. They were largely successful in coordinating the rich ideas they initially displayed regarding mechanisms with the data they generated. And, finally, they were able to implement these skills in a content-relevant context of their academic curriculum.

Comparative Analysis of Final Skill Levels at the End of Year 3

Although we regarded the research reports as the major product documenting the cognitive outcomes of year 3, we wished to include one further assessment in which we probed the degree to which students could transfer their skills to a different context, in particular one of their own choosing **and one that was not as highly structured as the year 1 and year 2 tasks, in the sense of guiding students through the inquiry process.** In addition, we wished to undertake a comparison of the skill levels students had achieved with those of students who had had no involvement in the program. We report on fulfillment of these goals in this final section.

Once they appreciate the objectives of inquiry, a critically important step in the development of inquiry skills, we believe, is for students to identify their own research topics and formulate the questions they want to answer with respect to those topics. Only in so doing can students begin to recognize the power and versatility of these skills. Embarking on this step was one of the major objectives of this phase of the work.

Do you have a "Does It Make a Difference" question you would like to find out an answer to?

Your question should be of this form:

Does _________ make a difference in _________?

Here is an example:

Does <u>time of day</u> make a difference in <u>how well people solve crossword puzzles</u>?

Notice this is a question you would be able to go out and get <u>evidence</u> to find an answer to. A question you choose should be this kind of question.

Think of a question like this that you would like to investigate and write it below:

Does _______________________________
make a difference in ___________________________?

As a vehicle for formulating their own research questions, students were presented with the sheet that appears in the box above and asked to work in teams to contemplate and complete it. A few groups required some assistance in **phrasing their question in an empirically addressable form (from questions, for example, such as "Does the team make a difference in how well they play?")**, but all managed to do so without great difficulty. Following group discussion, it was agreed that the entire class would pursue a question that had been posed by two different groups (with perhaps some communication between them) and attracted great interest:

Does playing video games increase your reading speed?

We would have liked to incorporate an activity in which students collected their own data, for example by administering a measure to assess the reading speeds of students in their own or another class and assessing whether these students played video games. Unfortunately, the reduced time we had available with students in year 3 made this potentially time-consuming activity not feasible, especially as we wished to compare students' performance to performance on this task by another group of students who had not participated in the inquiry program (and for whom such an activity would therefore not have been appropriate). We therefore provided a data base (presented as information from sixth-grade students at another school), but, in accordance with the first objective of this phase (assessing transfer), we undertook to represent it in as different a format as we could from the computer format the students had become familiar with, so as to assess their ability to transfer their skills to a superficially different task.

Main sample

Students in their inquiry class again worked in teams of 4 or 5, but they did not access data via computer, nor did they make any use of their laptop computers (which they used in much of their academic work) during the activity. Instead, the data were made available in the form of "student record cards." Each card contained the name of a student (first name and last initial, e.g., Juanita A.) and the student's score on a test of reading speed, presented as WPM (which it was explained to students stood for "words per minute") and assuming one of four values (100, 150, 200, or 250). In addition, information was available regarding the student's status on five binary variables:

a. whether the student played video games
b. whether the student watched TV
c. whether the student read on weekends
d. whether there was a computer in the home
e. whether the students' parents read at home

It was mentioned that information on these other factors was available, in addition to whether or not the student played video games, and that the team might want to take it into consideration. Each team was given access to a large envelope containing 110 such record cards in random order. The structure represented in the data set included the complicating characteristics of interaction and probabilistic outcomes that students had encountered in their earlier work. Specifically, presence of weekend reading raised reading scores, but only in the absence of TV, and parental reading raised scores but the effect was probabilistic rather than uniform and constant. The remaining variables had no effect.

Students were given no specific instruction as to how to approach the task, other than to try to answer their question for this population of individuals. (It was pointed out that results could be different for a different population.) Teams worked on the task for a total of one hour and were asked to indicate their conclusion(s) on "claim sheets," like those used earlier, that asked students to make a claim and provide supporting evidence.

The teams clearly faced a formidable task, to begin with simply in imposing some order on the database, and there are a number of ways they might have proceeded. We thus found it notable that, without conferring, each of the teams began with the same strategy: separating the record cards into two piles, those in which the person named on the card played video games and those in which the person did not. Within the two piles, a team typically did more sorting, placing together cards signifying the same or similar status on the remaining variables. Eight of the nine teams then employed what we will refer to as a "subtractive" strategy: comparing instances that were identical except for presence or absence of video-game-playing. One team, for example, indicated on a claim sheet that video games make no difference, citing as their evidence

VTPWC 100

TPWC 100

Another team used the same strategy but presented as their evidence the comparison of VC (video games and computer) to V alone.

Furthermore, all teams but one included multiple cases of the same type of record in their comparison. (There were either 3 or 4 instances of each unique type in the data base.) This strategy parallels the one used in the cart activity in which outcome for the same cart type was accessed multiple times. Students applied in the present context the understanding they had gained that the outcome might not always be the same, and all but one team made reference to modes for specific kinds of cases as what they were comparing rather than individual cases. (One team wrote explicitly, "We think 150 is the mode because 2 of the 3 cards were 150.")

Because of the limited time available, not all teams applied their strategy to drawing conclusions regarding all four of the remaining variables. All but one team, however, did evaluate the role of at least some of the other variables, and three teams did so for all of them. Two teams took the further step of addressing interaction effects, undertaking to replicate the effect (of weekend reading) at multiple levels of other variables.

Only one team displayed an approach significantly different from what has been described.

This team began in the same way, with the initial sort into two categories (video games and no video games). But then, rather than further division based on other variables, this team simply generalized across these variables, concluding that video games make no difference because "the majority" had an outcome of 100 WPM in both categories. (Given that levels of the remaining variables were approximately equally distributed across the two categories, this approach gets the team to a correct conclusion, but it is of course an unreliable method in general since these distributions need not be equal.)

To be added to this generally positive picture of skill development, is the to us rather surprising finding that two of the nine teams showed inconsistent, and faulty, strategies as they went on to investigate effects of the remaining variables. Both teams had clearly shown the subtraction method at the beginning of their work. One team, for example, show clear mastery of this method in their investigation of video games. As evidence for their claim that video games make no difference, they report:

> To confirm about V we did VTC to TC. Both modes were 100. While doing this, we also found out that T does not matter [by comparing VTC and VC].

They appear to go astray, however, when they begin to contemplate the effect of more than one variable at a time, and on a separate claim sheet, they report that "computer and weekend reading make a difference," offering as evidence that "The mode for V was 100 and the mode for VCW was 200." On another claim sheet they draw on both correct and incorrect methods in providing evidence for the same claim (that weekend reading does not matter): "The mode of TP is 150 and TPW is also 150. To confirm our answer, we used TCW and the mode was 100."

A second team showed a similar mixture of methods, although they never considered interactions. After using the subtractive method to evaluate two different variables, they claimed that computer does not matter because "We got TPW and the mode was 150 and the mode was 150 for computer so it doesn't matter." The examples provided by these two teams, then, suggest a conclusion that has now become familiar from microgenetic research (Kuhn, 1995; Kuhn et al., 1995; Siegler, 2006): Even what appear to be very straightforward and very effective strategies are not immediately consolidated, and ineffective strategies may co-exist in a strategic repertory for extended periods.

Comparison sample

Here we report the results of the assessment we undertook of a group of students who had not participated in any inquiry activity, in order to compare their inquiry skills with those participating in our inquiry program. These were students at the same school but one grade above those participating in our program. These seventh graders were divided into two class sections, only one of which was available to us on these occasions, but the numbers of participants were adequate for our purposes, **which was to show that these older students had not developed the inquiry skills that enabled the sixth graders to succeed on the task.** We presented these students with the reading-speed task described above that inquiry students had engaged in as their final activity of that year. The approaches they took to the task, compared to those exhibited by the inquiry students, are of considerable interest**. (Given our purpose here, we did not undertake to investigate what kinds of intervention might have led to success on the part of this older group.)**

None of the three seventh-grade teams (of 4-5 students each) working on the reading-speed task exhibited approaches anything like those described above as characteristic of the sixth-grade inquiry students. Specifically, none initially sorted the record cards on the basis of their status on the video-game (or any other) variable, and none applied the subtractive strategy, or even a comparative strategy, to compare two kinds of cases.

The one seventh-grade team who got closest to examining cases having a certain variable constellation made the claim that weekend reading makes a difference. As evidence they state, "Somebody who reads on the weekend and plays video games has a 200WPM average," followed by a list of 11 names that appear on record cards of individuals whose reading speed is indicated as 200 WPM. This approach in fact illustrates one taken by all the seventh-grade teams: Instead of sorting the record cards by status on one or more of the variables, as the sixth-grade inquiry students did, they sorted the cards by outcome, putting together all those cards that showed a common reading speed. The team noted here followed their list of 11 names with the conclusion:

> Some have no weekend reading and video games, they have a higher reading than someone who just plays video games, some have lower if they just play the computer and much lower than if they just read and play video games.

In contrast to this team, the other seventh-grade teams undertake some quantitative analysis. One team's work appears in figure 7. This team went on prepare a second sheet pertaining to video games, claiming they make a difference, which appears in figure 8. On this sheet, confined to cases of 150 WPM, they appear to draw conclusions regarding all variables, suggesting that those variables showing percentages above 50% have an effect and those with percentages below 50% do not.

Finally, the third team uses a different quantitative strategy (figure 9), first dividing reading speeds into two categories, fast and slow. The four similar percentages that appear in the figure we might anticipate would constitute the support for their conclusion of no relationship. But instead they make the comparison reflected in their final sentence, one that in fact suggests a relation between variable and outcome inverse to the one that was anticipated on theoretical grounds and hence is interpreted by them as an absence of relationship.[1]

General Discussion

The single most important conclusion that we believe follows from the data we have presented here is that over the three years of their participation, students in our inquiry program developed skills they otherwise would not have during this period. The approaches demonstrated by the seventh-grade comparison group on the reading-speed task were thoughtful, intelligent approaches to making sense of the data presented to them. These seventh graders were motivated and largely able students who were energetic in drawing on numerous logical and mathematical skills at their disposal to address the question at hand. But the strategies these students applied were not ones capable of fulfilling their objectives. They did not have the strategies they needed to be effective.

The students one year younger who had participated in the inquiry program present a different picture. The work that has been described here indicates that the rudimentary skills of inquiry that were the focus of the early phases are acquirable by middle-grade students by means of extended engagement in an inquiry environment. They are also readily amenable to incorporation within the context of an academic curriculum. And, once developed, they are extendable in some fairly advanced ways.

The feasibility and productivity of engaging elementary-school students in inquiry activities have been suggested in earlier work (see National Research Council, 2007, for

review of approaches). The present undertaking adds two things to such work. One is close examination of how inquiry skills develop over a very extended period of engagement during the critical middle-elementary years and the series of challenges that are encountered in the process. The other is examination of the development of more advanced inquiry skills, in the senses first of interpreting more complex forms of data and second in carrying out an inquiry activity without specific guidance through its sequential phases (as students did in the Cart and Reading-speed problems).

This is not to say that the skills of the students we worked with were solidly in place at the end of their third year of work. The fragility of these skills was evident in the numerous ways we have described and summarize here. The continuing cycling of inquiry skills in the national curriculum standards across elementary and secondary grades would appear not to be ill-conceived. Students need to engage their skills on a regular basis, as they gradually gain stronger command of them at both the performance level and the meta-level (Kuhn, 2001a, 2005; White & Frederiksen, 2005). They are not learned once, by any method, and then reliably available thereafter as needed.

Consistent with earlier work (Kuhn et al., 1992, 1995, 2000; Kuhn & Dean, 2005; Dean & Kuhn, 2007; Schauble, 1990, 1996). the present findings provide further evidence of the effectiveness of dense engagement with problems requiring these skills in developing the elementary investigatory and inference skills basic to inquiry (*Two-instance Comparison* in table 2 and the three skills labeled *Comparison*, *Control*, and *Replication* in table 4), The present findings also point to the effectiveness of scaffolding that is introduced and then gradually relaxed (Reiser, 2004; Pea, 2004)**. In the present work, students were initially guided through each phase of the inquiry process by the highly structured computer interface designed for this purpose. In the final (reading speed) activity, students received no guidance at all with respect to how to undertake the task.**

Although the various kinds of software and human supports we introduced, including prompts to encourage reflection and analysis (“What do the findings tell you?” “Is there anything else that could be making the outcomes different?”), appeared productive, we cannot of course isolate which of them were more or less effective in promoting development. But this fact does not undermine achieving our goal in this work: to support the development of concern here in order to observe and better identify its characteristics.

Nor, we should make clear, do we wish to imply that we are advocating the specific multi-year sequence of activities engaged in by the students we studied as a model curriculum for the development of inquiry skills. We see the contribution of the present work to lie rather in the identification and description of a core set of such skills and the challenges that acquiring these skills pose. As we stated at the outset, knowledge of this sort stands to be useful in formulating educational objectives with respect to inquiry and designing appropriate curricula. The scope and specifics of such curricula, and how they are to be integrated with other curricular objectives, will be decisions that need to take into account specific circumstances and most often be made at the local level. Nonetheless, it is hard to envision an inquiry curriculum that would proceed very far without concerning itself with the fundamental skills that have been our topic here. Designers of inquiry programs for middle-school students have sometimes assumed that these fundamental skills are already in place – that students will immediately understand what it means to examine evidence and to "find out." The work we have presented here suggests strongly that these assumptions are unwarranted.

What develops?

Formulating an objective. In addition to illustrating the slow, uneven course of mastery, with lingering appearance and reappearance of less-advanced strategies, the particular findings of most value to come out of the present work, we believe, are the continuing difficulties that students exhibited with respect to awareness of inquiry objectives, **which are of course the foundation on which all further inquiry strategies and meta-strategies rest.** The year 1 and year 2 findings are **consistent with our own and others' earlier work (Kuhn & Dean, 2005, Kuhn et al., 1992, 1995, 2000; Schauble, 1990, 1996; Schauble et al., 1991) in indicating that the purpose of inquiry that students need to identify is that of *analysis* –** in this case finding out how a particular multivariable system operates – rather than the engineering focus of producing favorable outcomes **(see Lehrer & Schauble, 2006, and Zimmerman, 2007, for review of research).**

Consistent with the findings of Kuhn and Dean (2005), identifying a question appears to play a key role in making the rest of the inquiry cycle productive. In the probabilistic version of *Ocean Voyage* in year 2, for example, students floundered until they were helped

to formulate a specific research question. Like other components of the inquiry process, this skill is not one a student learns once and has mastered. Rather, as we saw in students' research reports on the Cart problem, this focus could waver, even when it had appeared to be secure, and especially in this case in which the problem content made it easy to slip back into an engineering mode. Retaining a focus on analytic objectives **appears to be** a challenge that novice investigators continue to face for some time.

Once students in the context we presented them identify an appropriate goal for their inquiry, the challenges facing them have by no means all been met. First, there remains the challenge of decontextualization (Stanovich, 2004) in making sense of evidence, a requirement, as we noted earlier, that has been the subject of misunderstanding in the science education literature. Students cannot undertake rigorous analysis of evidence **unless they can read the data,** i.e., represent and reflect on it independent of their own expectations. By year 2, all of our students were able to do so consistently.

Interpreting evidence and drawing conclusions. Once students are able to attend to the evidence, they still may not get very far in drawing valid inferences based on it since initially they tend to limit themselves to interpreting single, isolated cases (Level 2 in table 2), thus not affording themselves the opportunity for comparison that is essential to valid analysis and interpretation (represented in table 4 by the three successive dimensions labeled *Claims…*, *Conclusions…*, and *Coordination…*) Also essential to this achievement, our year 3 findings suggested, is identifying the variable, or feature, as a conceptual unit warranting investigation and analysis. Related to this achievement is the recognition that if a feature makes a difference it always makes a difference (whether the level is high or low, present or absent). An absent feature thus has as much to say about what happens as a present one.

Predicting outcomes. **Second only to identifying inquiry objectives, the other skill that remained the most significant challenge for students was integrating information about multiple variables to make predictions.** While identifying the goal of inquiry proved more fragile than we expected and, given its foundational status, warrants continued support, prediction (of outcomes influenced by multiple variables), in contrast, has been largely neglected as a scientific inquiry skill. In this case, further research is warranted regarding the challenges that mastery of this skill poses and how best to support students in meeting them **(Kuhn et al., in press).** To be able to integrate the effects of multiple variables on one or more outcomes is a skill of significant practical as well as theoretical import. Issues of cognitive load, as well as meta-level comprehension, are implicated. The means of achieving this goal, most likely, will lie in the kind of scaffolded engagement with problems requiring the skill that we have found effective here.

Representing and communicating findings. Turning to a more positive outcome, the quality of the year 3 research reports, which were produced with minimal adult guidance, is worthy of note. Although these reports reflect a number of continuing conceptual difficulties and are far from perfect, the progress that these middle-school students showed toward authentic scientific writing bears emphasis. Difficulties in conveying the results of inquiry activities in writing, even in the most elementary respect of note-taking during investigation, have been the consistent report of studies that have examined these skills (Garcia-Mila & Andersen, 2007; Keys, 1999; Sandoval, 2003). The achievements of the students reported on here in formally representing and communicating their inquiry activities in written form are thus notable, especially in demonstrating an absence of age-related constraints on developing such skills.

Beyond the basics

Once students had achieved the development observed in year 1 and the beginning of year 2, this foundation allowed them to grapple successfully with more sophisticated (and authentic) kinds of data bases (involving probabilistic and interactive effects) introduced later in year 2 and in year 3. There do not appear to be any strong constraints that limit what might be achieved at this age in the development of inquiry skills. Our year 3 work shows that these achievements can readily be extended into richer content that is integrated into students' academic curriculum even though some of the basic skills we have identified – most notably, a focus on objectives and the integration involved in predictions – remain fragile. The work on interactive effects at the end of year 2 and in year 3 also highlights the continuing challenge of identifying and retaining a focus on objectives (in the case of interactions, as we observed, at issue is a focus on which variable one is trying to find out about).

It remains to note that for all of the skill they acquired, the students we followed over three years hardly became experts in scientific method. The activities they engaged in only indirectly addressed two essential aspects of scientific method, its epistemological foundations (Kuhn, 2005; Kuhn et al., in press; Metz, 2004, Sandoval, 2005) and its argument aspect (Kuhn, 1993; Kuhn et al., in press; Lehrer, Schauble, & Petrosino, 2001; **Newton, Driver, & Osborne, 1999**), that allow claims to enter the realm of scientific discourse. Like skills of inquiry, skills of argument are ones that must be engaged and developed; they are far from intuitive (Kuhn & Udell, 2003, 2007; Kuhn, Goh, Iordanou, & Shaenfield, in press).

Nor did we undertake to develop students' awareness and understanding of the epistemology of science as a discipline. Although epistemological matters did become salient at times during the course of their activities, we focused instead on "epistemology in action" (Sandoval, 2005; Metz, 2004) – in other words, on students' meta-level understanding of the knowledge-building nature and status of their own scientific activities. Reflected in this approach is our view that what have been referred to as "personal epistemology" (Hofer & Pintrich, 1997, 2002) and "scientific epistemology" (Sandoval, 2005; Smith, Maclin, Houghton, & Hennessey, 2000) and treated almost entirely independently are not in fact distinct entities (Kuhn et al., in press). Both have to do with understanding of how knowledge is constructed, whether by untutored children or professional scientists.

Nonetheless, unless students become conversant in the fundamentals that have been the focus of the present work, there is nowhere to go in the world of science. Equally important, one's effectiveness in realms outside of science is likely to be constrained as well. As one of us (Kuhn, 2005) has argued elsewhere, the significance of the cognitive skills examined here extends well beyond the discipline of science narrowly conceived.

To further the likelihood that young students like those we worked with here might be disposed to enter the world of professional science, their engagement with scientific inquiry must be sustained and broadened. It is arguably best to introduce them to a scientific way of thinking in contexts that involve matters already of interest to them. If they become convinced of the value of a scientific approach, the kinds of questions they are interested in applying it to are likely to expand. Helping them to see first-hand the value of that approach thus seems the first order of business. **Toward that end, the sustained engagement over time that we examined here appears both necessary and feasible.**

References

Anderson, N. (1991). *Contributions to information integration theory: Vol. III. Development*. Hillsdale NJ: Erlbaum.

Dean, D., & Kuhn, D. (2007). Direct instruction vs. discovery: The long view. *Science Education, 91,* 384-397.

Dillenbourg, P. (Ed.). (1999). *Collaborative learning: Cognitive and computational approaches*. Oxford: Elsevier.

Dixon, J., & Tuccillo, F. (2001). Generating initial models for reasoning. *Journal of Experimental Child Psychology, 78*, 178-212.

Duschl, R., & Grandy, R. (2005). Reconsidering the character and role of inquiry in school science: Framing the debates. Unpublished manuscript, Rice University.

Ford, M. (2005). The game, the pieces, and the players: Generative resources from two instructional portrayals of experimentation. *Journal of the Learning Sciences, 14,* 449-488.

Fortus, D., Hug, B., Krajcik, J. et al. (2006). Sequencing and supporting complex scientific inquiry practices in instructional materials for middle school students. Paper presented at the annual meeting of the National Association for Research in Science Teaching. San Francisco, April.

Garcia-Mila,, M., & Andersen, C. (2007). Developmental change in notetaking during scientific inquiry. *International Journal of Science Education, 29,* 1035-58.

Hofer, B., & Pintrich, P. (1997). The development of epistemological theories: Beliefs about knowledge and knowing and their relation to learning. *Review of Educational Research, 67,* 88-140.

Hofer, B., & Pintrich, P. (Eds.). (2002). *Epistemology: The psychology of beliefs about knowledge and knowing*. Mahwah NJ: Erlbaum.

Kelly, G. J., Druker, S., & Chen, C. (1998). Students' reasoning about electricity: Combining performance assessments with argumentation analysis. *International Journal of Science Education*, 20 (7), 849-871.

Keselman, A. (2003). Supporting inquiry learning by promoting normative understanding of multivariable causality. *Journal of Research in Science Teaching, 40*, 898-921.

Keys, C. (1999). Language as an indicator of meaning generation: An analysis of middle school students' written discourse about scientific investigations. *Journal of Research in Science Teaching, 36,* 1044-61.

Klahr, D. (2000). *Exploring science: The cognition and development of discovery processes*. Cambridge: MIT Press.

Klahr, D., & Nigam, M. (2004). The equivalence of learning paths in early science instruction: effects of direct instruction and discovery learning. *Psychological Science, 15*, 661-667.

Krajcik, J., Blumenfeld, P., Marx, R., Bass, K., Fredricks, J., & Soloway, E. (1998). Inquiry in project-based science classrooms: Initial attempts by middle school students. *Journal of the Learning Sciences, 7,* 313-350.

Kuhn, D. (1993). Science as argument: Implications for teaching and learning scientific thinking. *Science Education, 77*(3), 319-337.

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

Kuhn, D. (2001a). Why development does (and doesn't) occur: Evidence from the domain of inductive reasoning. In R. Siegler & J. McClelland (Eds.), *Mechanisms of cognitive development: Neural and behavioral perspectives*. Mahwah: Lawrence Erlbaum Associates.

Kuhn, D. (2001b). How do people know? *Psychological Science, 12*, 1-8.

Kuhn, D. (2002). What is scientific thinking and how does it develop? In U. Goswami (Ed.), *Handbook of childhood cognitive development*. Oxford: Blackwell.

Kuhn, D. (2005). *Education for thinking*. Cambridge: Harvard University Press.

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

Kuhn, D., Black, J., Keselman, A., & Kaplan, D. (2000). The development of cognitive skills to support inquiry learning. *Cognition and Instruction, 18 (4),* 495-523.

Kuhn, D., & Dean, D. (2004). Connecting scientific reasoning and causal inference. *Journal of Cognition & Development, 5*(2), 261-288.

Kuhn, D., & Dean, D. (2005). Is developing scientific thinking all about learning to control variables? *Psychological Science, 16,* 866-870.

Kuhn, D., Garcia-Mila, M., Zohar, A., & Andersen, C. (1995). *Strategies of knowledge acquisition*. (with commentaries by Sheldon White, David Klahr & Sharon Carver). *Society for Research in Child Development Monographs, 60* (4), Serial no. 245.

Kuhn, D., Goh, W., Iordanou, K., & Shaenfield, D. (in press). Arguing on the computer: A microgenetic study of developing argument skills in a computer-supported environment. *Child Development*.

Kuhn, D., Iordanou, K., Pease, M., & Wirkala, C. (in press). Beyond control of variables: What needs to develop to achieve skilled scientific thinking? *Cognitive Development*. [Special issue, The Development of Scientific Thinking, B. Sodian & M. Bullock, eds.]

Kuhn, D., Schauble, L., & Garcia-Mila, M. (1992). Cross-domain development of scientific reasoning. *Cognition and Instruction, 9*, 285-332.

Kuhn, D., & Udell, W. (2003). The development of argument skills. *Child Development, 74(5),* 1245-1260.

Kuhn, D., & Udell, W. (2007). Coordinating own and other perspectives in argument. *Thinking and Reasoning, 13*, 90-104.

Lafon, P., Chasseigne, G., & Mullet, E. (2004). Functional learning among children, adolescents, and young adults. *Journal of Experimental Child Psychology*.

Lehrer, R., & Schauble, L. (2006). Scientific thinking and scientific literacy: Supporting development in learning contexts. In K. A. Renninger & I. Sigel (Eds.), *Handbook of Child Psychology. Vol. IV*. Hoboken NJ: Wiley.

Lehrer, R., Schauble, L., & Petrosino, A. J. (2001). Reconsidering the role of experiment in science education. In K. Crowley, C. Schunn & T. Okada (Eds.), *Designing for science: Implications from everyday, classroom, and professional settings* (pp. 251-277). Mahwah NJ: Lawrence Erlbaum.

Metz, K. (2004). Children's understanding of scientific inquiry: Their conceptualization of uncertainty in investigations of their own design. *Cognition and Instruction, 22*, 219-290.

National Research Council (1996). The National Science Education Standards. Washington DC: National Academies Press.

National Research Council, National Academy of Science. Committee on science learning. (2007). Taking science to school: Learning and teaching science in grades K-8. Washington DC: National Academies Press.

Newton, P., Driver, R., & Osborne, J. (1999). The place of argumentation in the pedagogy of school science. *International Journal of Science Education, 21,* 553-576.

Osborne, G., Erduran, S., & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching, 41,* 994-1020.

Patronis, T., Potari, D., & Spiliotopoulou, V. (1999). Students' argumentation in decision-making on a socio-scientific issue: implications for teaching. *International Journal of Science Education, 21,* 745-754.

Pea, R. (2004). The social and technological dimensions of scaffolding a d related theoretical concepts for learning, education, and human activity. *Journal of the Learning Sciences, 423-452*.

Reiser, B. (2004). Scaffolding complex learning: The mechanisms of structuring and problematizing student work. *Journal of the Learning Sciences, 13,* 273-304.

Reiser, B., Smith, B., Tabak, I., Steinmuller, F., Sandoval, W., & Leone, A. (2001). BGuILE: Strategic and conceptual scaffolds for scientific inquiry in biology classrooms. In S. Carver & D. Klahr (Eds.), *Cognition and instruction: Twenty-five years of progress*. Mahwah NJ: Erlbaum.

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

Sandoval, W. (2005). Understanding students' practical epistemologies and their influence. *Science Education, 89,* 634-656.

Schauble, L. (1990). Belief revision in children: The role of prior knowledge and strategies for generating evidence. *Journal of Experimental Child Psychology, 49,* 31-57.

Schauble, L. (1996). The development of scientific reasoning in knowledge-rich contexts. *Developmental Psychology, 32,* 102-119.

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

Siegler, R. (2006). Microgenetic studies of learning. In D. Kuhn & R. Siegler (Eds.), (W. Damon & R. Lerner, Series eds.), *Handbook of Child Psychology: Vol. 2. Cognition, Perception, and Language.* (6th edition). Hoboken NJ: Wiley.

Smith, C., Maclin, D., Houghton, C., & Hennessey, M. (2000). Sixth-grade students' epistemologies of science: The impact of school science experiences on epistemological development. *Cognition and Instruction, 18*, 349-422.

Stahl, G., Koschman, T., & Suthers, D. (2006). Computer-supported collaborative learning. In R.K. Sawyer (Ed.), *The Cambridge handbook of the learning sciences.* Cambridge: Cambridge University Press.

Stanovich, K. (2004). *The robot's rebellion*. Chicago: University of Chicago Press.

White, B., & Frederiksen, J. (2005). A theoretical framework and approach for fostering metacognitive development. Educational Psychologist, 40, 211-223.

Wilkening, F. (1982). Integrating velocity, time, and distance information:
A developmental study. *Cognitive Psychology, 13,* 231-247.

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

Zimmerman, C. (2007). The development of scientific thinking skills in elementary and middle school. *Developmental Review*, *27,* 172-223.

Footnote

1. This difficulty in distinguishing between an inverse relationship and absence of a relationship is not unusual at this age level (Kuhn, Amsel, & O'Loughlin, 1988; Lafon, Chasseigne, & Mullet, 2004).

Table 1

Causal Structure Represented in *Earthquake Forecaster*

Feature	Effect	Outcome
Soil type (igneous or sedimentary)	Non-causal	Both yield identical outcomes
S-wave rate (fast or slow)	Non-causal	Both yield identical outcomes
Water quality (poor or good)	Causal	Good indicates one unit of greater risk
Snake activity (high or low)	Causal	High indicates one unit of greater risk
Gas level (heavy or light)	Causal	Heavy indicates one unit of greater risk
Outcome alternatives: Extreme, High, Medium, or Low earthquake risk		

Table 2

Levels of Pretest Performance by Fourth Graders on Initial Inquiry Problem

Level	Description	Illustration	Frequency
Level 1. Exclusively theory-based	These students do not yet differentiate theory from evidence. All of their inferences are based purely based on their prior theories about the phenomena, and they at most make reference to fragments of the evidence to illustrate these views.	This case has high risk because lots of gas gets something started in the earth.	14
Level 2. Some evidence based inferences but no two-instance comparisons	On at least some occasions, these students attempt to interpret the data they access, but they limit their interpretations to single instances that are not compared to any others.	This case has good water, heavy gas and lots of snake activity but I think it's the gas that's making it high-risk.	11
Level 3. Some two-instance comparison	These students carry out comparisons between cases but comparisons are not controlled (i.e., differ with respect to only one variable) and therefore lead to faulty inferences.	The first case had all good things and low risk. This one has all bad things and high risk. So heavy gas has something to do with the high risk.	05
Level 4. Some valid two-	These students carry out some controlled comparisons,	This one has all the same things as the first	

instance comparison	leading to valid inference, but also some uncontrolled comparisons that do not allow for valid inference.	one, except the gas is light, and the risk went down. So the gas has something to do with the risk.	00
Level 5. Consistent valid 2-instance comparison	These students carry out consistently controlled comparisons between two instances and draw appropriate inferences.	Consistent use of Level 4 comparisons and inferences.	00

Table 3

Fourth Graders' Progress during Year 1

	Posttest Level					
	Level 1	Level 2	Level 3	Level 4	Level 5	Total number of students
Minimal pretest competency (Level 1)	3 (21%)	1 (7%)	3 (21%)	2 (14%)	5 (36%)	14
Some pretest competency (Levels 2 or 3)	2 (13%)	0	2 (13%)	2 (13%)	10 (63%)	16
Total number	5 (17%)	1 (3%)	5 (17%)	4 (13%)	15 (50%)	

Table 4

Summary of Skills Exhibited in Students' Research Reports

Skill	Number of reports (of 18) adequately exhibiting skill	Examples from student reports
Objective or question	9	*Adequate*: "The purpose of this investigation was to figure out what features would make a difference in the amount of trips you could make between the pile of rocks and the worksite you were heading to." *Inadequate*: "The purpose of the project was to build a cart that can carry rocks and make the most amount of trips in an hour." *Inadequate*: "The purpose of this project was to see how many trips Rafael could make in an hour" *Inadequate*: "We had to pick which pieces will make a difference on the amount of trips and figure out what combination worked the best."
Supported hypothesis(es)	12	*Adequate*: "It's better to have a short handle because if you're closer to the weight it won't be harder to pull." *Inadequate*: "We thought the smaller bucket would be better."
Prediction(s)	16	*Adequate*: "We thought the smaller bucket would be better." *Inadequate*: Hypothesis: Does wheel size make a difference?"
Method: Replication	15	*Adequate*: "Our method was to redo the same case over and over."

		Inadequate: No mention of replication
Method: Comparison	15	*Adequate:* "We then did a big bucket to run against the small bucket." *Adequate*: "Our method was to make this original case. Then we would compare most of the cases to this one." *Inadequate*: "Doing a lot of cases like selecting different features to figure out which matter and don't matter."
Method: Control	14	*Adequate*: "We would see the results for a certain cart and then do it over and only change the level of the feature we were investigating." *Inadequate*: "From comparing the two wheel sizes we realized that both modes were 8."
Claims supported by evidence	17	*Adequate*: "From comparing the two wheel sizes we realized that both modes were 8. Therefore it doesn't make a difference." *Inadequate*: "We did a case 15 times. Conclusion: bucket size makes no difference"
Conclusions well distinguished from hypotheses	15	*Adequate:* "We have evidence from different cases for our conclusions" *Inadequate*: "Conclusions: Bucket size matters because when small it's lighter and easier to pull. So you can complete one more trip."
Coordination of conclusions and hypotheses	15	*Adequate:* "The hypothesis was wrong. We thought wheel size made a difference but it didn't." *Inadequate*: No mention of how conclusions relate to hypotheses
Concern with the possibility of interaction effects	6	*Adequate:* "Our first conclusions were quick ones without that much evidence. After doing this we wondered if another variable affects another variable. We found out that handle matters when the placement is far, but not if it is near. Our evidence was multiple cases of the same

		thing so we could find out the mode. The modes for handle matters when it was far in position, but earlier in our quick results we thought it didn't matter but that was when the positioning was near. So we checked 31 times to make sure we understood therefore giving us the results of handle matters when the placement is far but not if it is near." *Inadequate*: "We found handle length makes no difference."

Figure 1. Find out screen.

Figure 2. Case request screen.

Figure 3. Results and conclusions screen.

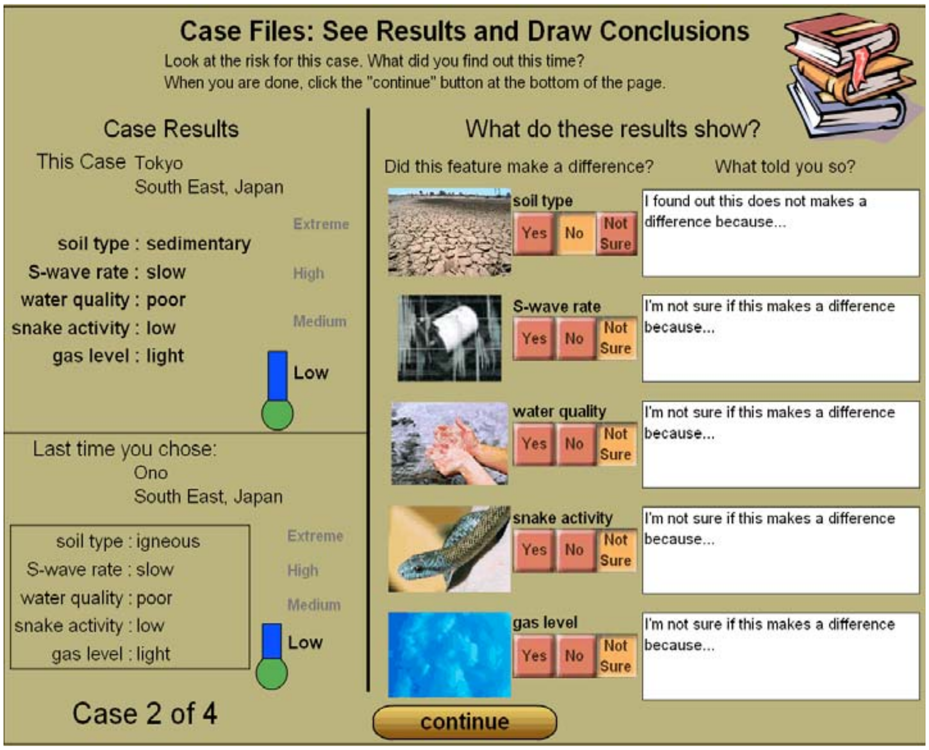


Figure 4. Notebook screen.

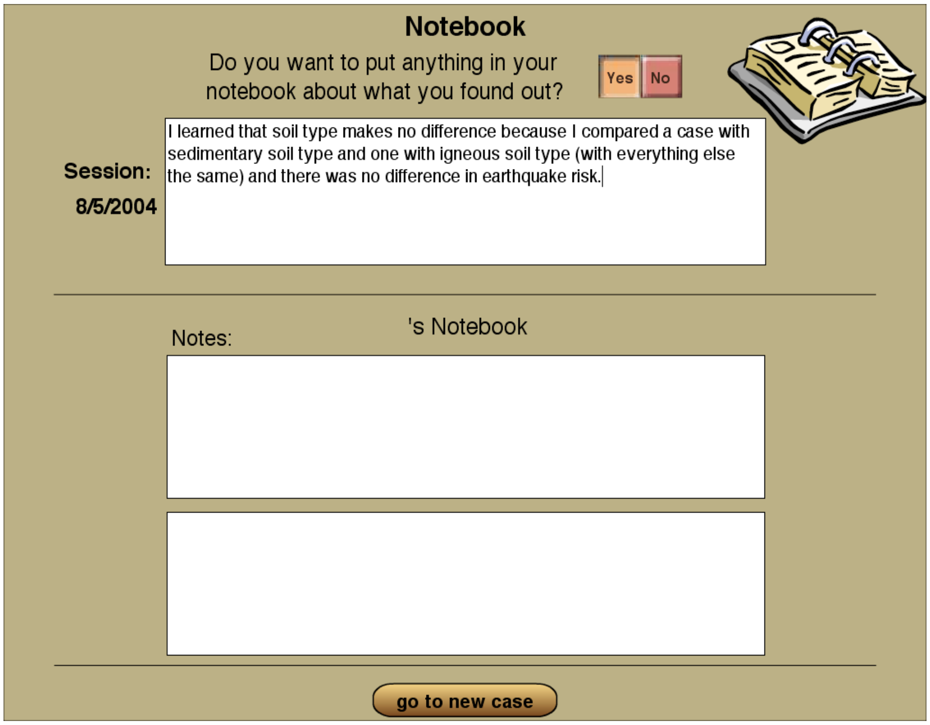


Figure 5. Prediction screen.

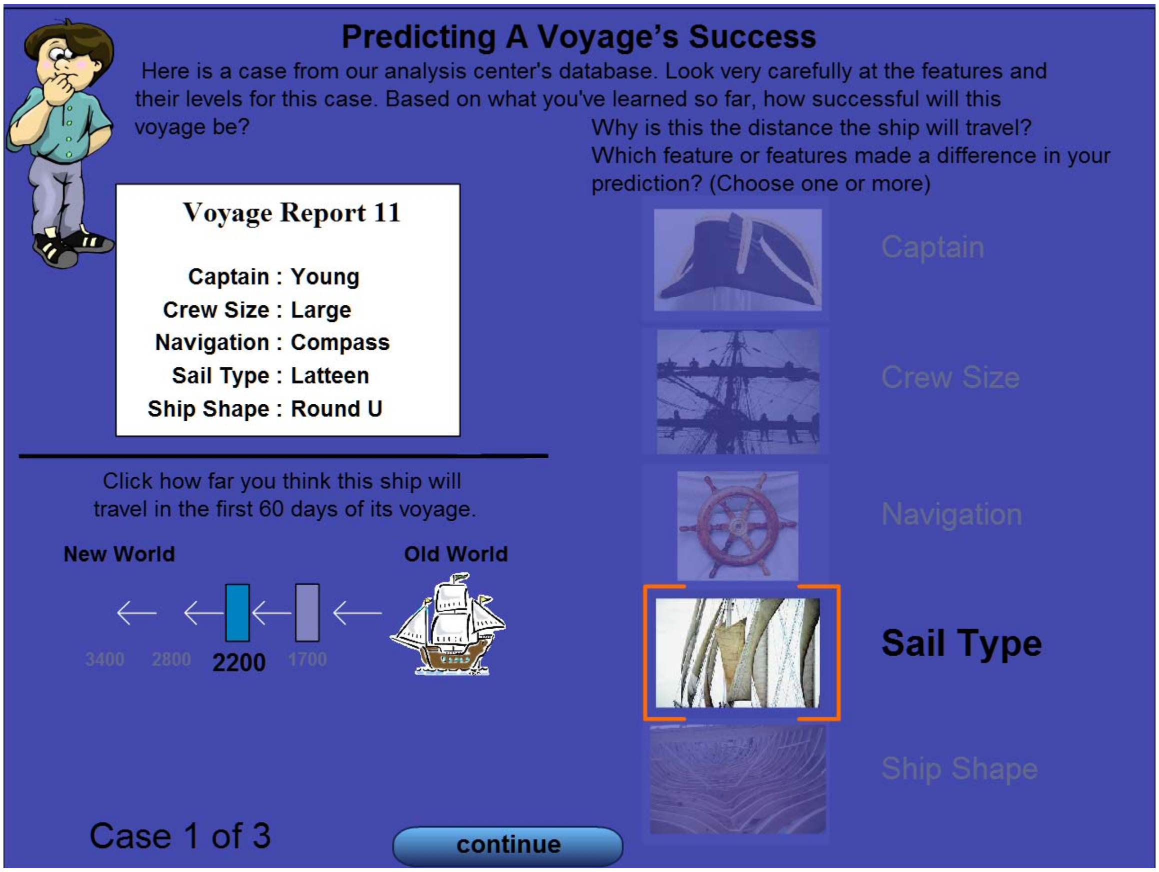


Figure 6. Rafael and his cart.

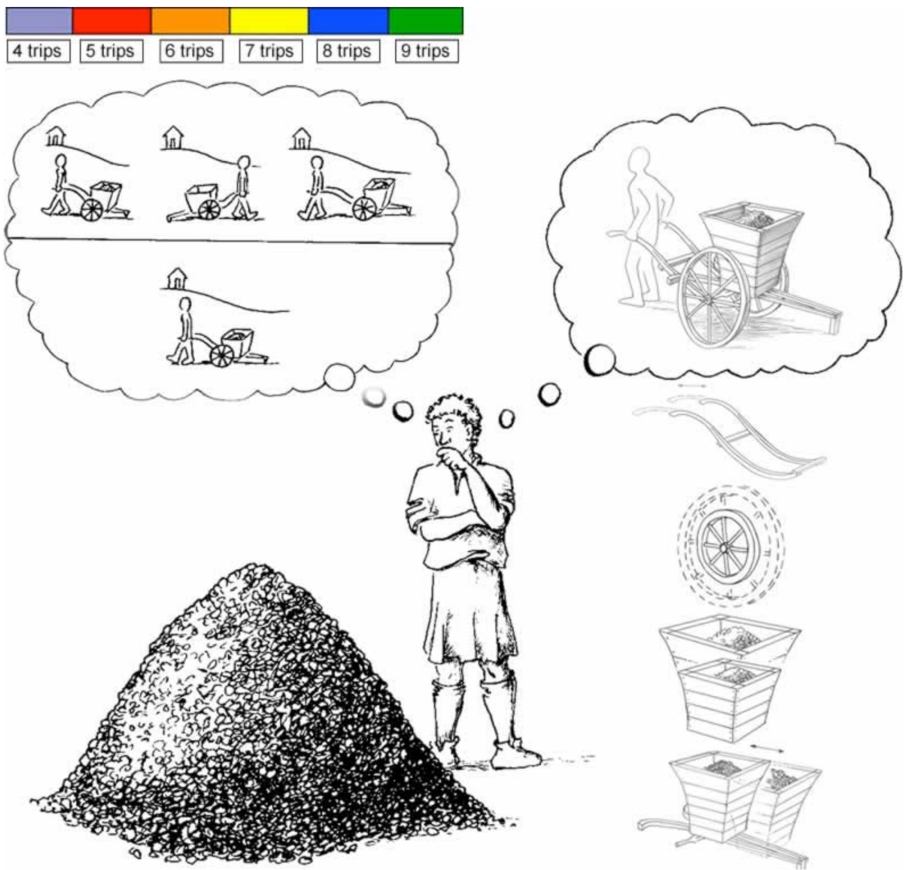


Figure 7. Seventh-grade student claim sheet

What is your claim?

Video games does/doesn't *(circle one)* make a difference in reading speed.

What is the evidence for your claim?

WPM = 100

25 out of the 54 kids

$\frac{25}{54}$ kids do not play games

but...

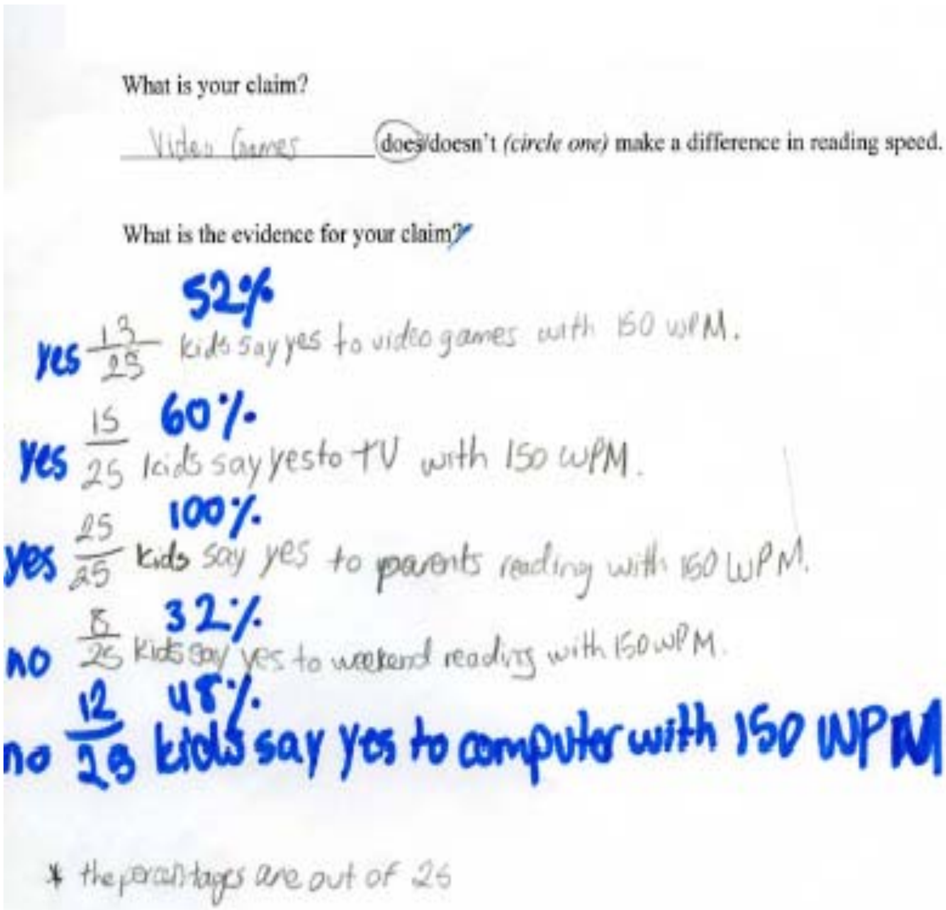
$\frac{29}{54}$ kids play video games

$\frac{25}{54} = 43\%$

$\frac{29}{54} = 53\%$

this is a 10% difference

Figure 8. Seventh-grade student claim sheet



What is your claim?

Video Games does/doesn't *(circle one)* make a difference in reading speed.

What is the evidence for your claim?

yes $\frac{13}{25}$ 52% kids say yes to video games with 150 wPM.

yes $\frac{15}{25}$ 60% kids say yes to TV with 150 wPM.

yes $\frac{25}{25}$ 100% kids say yes to parents reading with 150 WPM.

no $\frac{8}{25}$ 32% kids say yes to weekend reading with 150 wPM.

no $\frac{12}{25}$ 48% kids say yes to computer with 150 WPM

* the percentages are out of 25

Figure 9. Seventh-grade student claim sheet

What is your claim?

video games does/doesn't (circle one) make a difference in reading speed.

What is the evidence for your claim?

Yes Video fast speed	Yes Video slow speed	no video Fast speed	no video slow speed
29/111	31/111	27/111	24/111
26%	27%	24%	22%

According to our data we don't think vidio games has an effect on Reading pace.

The evidence shows that the largest amount or fast paced readers was in the category where they played vidio games.