Urban Middle School Students' Perceptions of the Value and Difficulty of Inquiry

William A. Sandoval · Aletha M. Harven

Abstract Following their participation in a guided-inquiry unit, 129 seventh-graders from five diverse urban middle schools were asked about their perceptions of specific inquiry tasks, from an expectancy-value framework. Students were asked to rate the interest value, utility value, and task difficulty of (a) data collection design; (b) explanation; (c) data analysis; and (d) citing evidence for claims. The utility of all tasks was rated highly, while interest ratings were moderate. Students perceived these tasks as moderately different from their usual work, and not especially difficult. No gender differences were found in students' ratings. Investigation tasks were rated as more interesting and useful than argumentation tasks. Students from lower SES schools found all tasks more useful and interesting than their peers in higher SES schools. Students' justifications for their ratings suggest they valued the utility of knowing how to back up their ideas with evidence.

Keywords Motivation · Interest · Science inquiry · Middle school · Biology

Introduction

One of the hopes behind inquiry-oriented reforms in science education is that when students learn to practice science as scientists do, they should become more motivated to learn science. The evidence for such hope is mixed, at

W. A. Sandoval (✉) · A. M. Harven
Graduate School of Education & Information Studies, University of California, 2335 Moore Hall, Box 951521, Los Angeles, CA 90095-1521, USA
e-mail: sandoval@gseis.ucla.edu

best (NRC 2005). Much of the research on affective aspects of science education has focused on students' "interest in science, attitudes about scientists, and attitudes toward the use of science" (Simpson et al. 1994, p. 216), but has made little connection to the theoretical literature on motivation (Osborne et al. 2003). Also, while Koballa and Glynn (2007) mention several interventions that seem to improve student attitude toward science through more interactive instruction, there has been little research that connects features of inquiry-oriented instruction to student motivation. This study is an attempt to make such a connection, by eschewing general assessments of science attitude, interest, or motivation to ask students directly for their perceptions of specific tasks of inquiry. We use the expectancy-value theory of motivation (Eccles and Wigfield 1995, 2002) to elicit middle school students' perceptions of their expectancies for success and the intrinsic value, two crucial factors behind motivated learning, of a set of inquiry tasks. While our study is exploratory, our findings indicate some of the features of inquiry tasks that allow students to perceive value from them and extend the limited research that has been done in this area.

Background

Research on affective aspects of science learning, including students' attitudes, interest, and motivation, has waxed and waned throughout the last several decades (Koballa and Glynn 2007). This research has tended to focus on students' attitudes toward science learning or science itself, or on their interest in learning science (Simpson et al. 1994). Research on students' attitudes toward science and motivation to learn science has increased over the last two decades, in part because

overall these attitudes are not very positive (Osborne et al. 2003). Another reason may be the warning that a "cold" model of conceptual change that excludes affective factors is not an adequate model of learning (Pintrich et al. 1993). Both Koballa and Glynn and Osborne and colleagues note the inherent relations between attitudes and motivation. At the same time, Koballa and Glynn show that constructs and theoretical frameworks for both attitudes and motivation are widely diverse.

Within motivation research there are a range of theoretical frameworks competing to account for individual differences in motivations to learn and achieve in school. Many of the leading frameworks were briefly reviewed by Pintrich and colleagues (Pintrich et al. 1993), and they suggest the range of influences that can potentially promote or inhibit any individual's motivation to learn in a particular situation. Echoing Pintrich and colleagues, Osborne et al. (2003) argue that research in science education should draw more explicitly from motivation frameworks, especially those that explore how situational factors influence motivation. As we describe below, we think expectancy-value theory (Eccles and Wigfield 1995, 2002; Wigfield and Eccles 2000) is such a framework.

The inquiry-oriented curricular reforms of the last three decades focus on learning outcomes, aiming for students to gain both deeper conceptual understanding and to learn practices of scientific investigation and reasoning. The evidence is they generally succeed in this (Minner et al. 2010; Schroeder et al. 2007). There remains scant evidence, however, that such reforms improve student motivation (NRC 2005). Reform efforts actually rarely study motivational outcomes (NRC 2005), and research from the learning sciences perspective that has produced many of these reforms has generally neglected to address motivational issues (Blumenfeld et al. 2006). Clearly, there are critical reasons to be concerned about motivational issues in science education, and the capacity for inquiry reforms to address them. Historically, girls are much less likely to sustain an interest in science through their years in school (Osborne et al. 2003; Simpson et al. 1994). Students of color are vastly under-represented in the sciences, both professionally and in course taking patterns in school. Students from disadvantage communities tend to achieve less well than students from higher socioeconomic status communities (Grigg et al. 2006), and these achievement differences might be attributable to differences in motivation (Graham and Taylor 2002), but this is far from clear. Much more research is needed that addresses the motivational consequences of engagement in inquiry curricula for students from varied backgrounds. It could be that inquiry-oriented curricula combat the "pedagogy of poverty" (Haberman 1991) common to urban schools (Thadani et al. 2010).

Pintrich et al. (1993) outlined four motivation constructs that influence the likelihood of individual conceptual change: goals, values, self-efficacy, and control beliefs. The work they review and more recent reviews (Blumenfeld et al. 2006; Eccles and Wigfield 2002) verify a number of intuitive claims about motivational beliefs. Students with a goal orientation of mastering subjects pursue deeper learning strategies. Students perform better on tasks when they value them, when they believe they can do well on them (self-efficacy), and when they believe they are in control of the outcomes of their own effort. Pintrich et al. pointed out, however, that a number of contextual variables in classrooms mediate such positive influences. For example, the structure of particular tasks or the authority structure in the classroom can induce students to see little value in the tasks they are asked to do, or to see their performance as beyond their control. They pointed out several factors that could moderate the influences of motivational beliefs on learning, including: the authenticity of school tasks for students, the level of autonomy students have, and the locus of authority in the classroom. In general, more authentic tasks, more student autonomy, and more student authority are associated with higher levels of motivation (Pintrich et al. 1993).

These features of classroom environments are features associated with inquiry pedagogy. It is surprising then that there appears to be little research in science education to explicitly examine how features of inquiry-oriented instruction might affect student motivation. Researchers have explored how aspects of classroom environments generally affect student motivation and achievement. Students are less motivated in classrooms that they perceive as being focused on ability (Anderman and Young 1994; Nolen 2003). Studies that have looked specifically at inquiry reform projects have used observational methods rather than the survey methods typical of motivation research, with one consequence being that they are limited to very small samples (Mistler-Jackson and Songer 2000; Patrick and Yoon 2004). Mistler-Jackson and Songer focused on six students from an implementation of their Kids as Global Scientists (KGS) project: two each that they categorized as having low, medium, and high motivation, using a combination of self-efficacy, control beliefs, values, and goal orientation scales. They report that the students with low and medium motivation liked KGS better than other science units, and specifically liked being able to spend time on the computer. Patrick and Yoon followed a similar strategy, following four students through a guided-inquiry unit. They found a range of variability in how these four students expressed their motivation, and it is quite difficult to discern any patterns in their limited sample. Further, neither of these studies links their interview and observational methods to typical motivation constructs. It

is unclear, for example, how liking to use the computer relates to value, goal orientation, self-efficacy, or control beliefs.

Lynch and her colleagues (Lynch et al. 2005) looked at the effects of an inquiry curriculum, *Chemistry That Applies* (CTA, State of Michigan 1993) on student motivation and cognitive engagement. Their motivation construct focused solely on goal orientation, using the typical contrast between a mastery or performance orientation. Similarly, they dichotomized cognitive engagement as either basic, meaning that students did the minimum to sustain their involvement in classroom activities, or advanced, to include deep processing strategies and high levels of self-regulation. Their quasi-experimental study included more than 1,500 students, and found that CTA produced significant gains in basic levels of engagement and mastery goal orientation. They did not find differences in advanced engagement, somewhat surprisingly. Part of the explanation lies in high pretest scores on this scale as reported by students. It suggests that engagement may need to be measured through means other than self-report.

The study we report here seeks to extend these recent findings relating motivational constructs to aspects of the classroom environment. We are specifically interested in whether or not students value inquiry tasks. We use the term value broadly, to include how interesting, difficult, and useful students find these tasks. Motivation research suggests, "perceptions of the value of a task do not have a direct influence on academic performance but they do relate to students' choice of becoming cognitively engaged in a task or course and to their willingness to persist at the task" (Pintrich et al. 1993, p. 184). Further, students' perceptions of the values of particular tasks should be amenable to change, as tasks can be more or less explicitly framed in relation to student interest and values.

Expectancy-Value Theory

Our approach here is grounded in Eccles and Wigfield's expectancy-value theory of achievement motivation (Eccles and Wigfield 1995, 2002; Wigfield and Eccles 2000). Eccles and Wigfield propose that students' motivation to achieve in any particular subject is a combination of their expectancies of success and the subjective value that they place on that subject. There are three main constructs in their theory. *Ability beliefs* refer to the beliefs that people hold about their ability in a particular domain. These beliefs are not proposed to be general, but are instead specific to particular domains—such as the belief that one is good at science. *Expectancies* for success are beliefs about one's likelihood to successfully perform a task or, more broadly, to achieve in a domain. Within particular domains, ability beliefs and expectancies for

success are highly correlated (Wigfield and Eccles 2000). When students think they are good at something, they expect to succeed at it, and when they do not believe they are good at something, they do not expect to succeed. A disturbing finding from Eccles' and Wigfield's research and research in this area generally is that as children progress through school their ability beliefs and expectancies for success decline (Eccles and Wigfield 2002). The third main construct in expectancy-value theory is *subjective values*. These are the values that achievement in a particular domain holds for an individual. Eccles and Wigfield consider three such values: *attainment* value is the importance one places on doing well on a task or in a subject; *intrinsic* value is the enjoyment one has from doing a task, and is obviously similar to interest; and *utility* value refers to how useful a task is perceived to be in helping an individual achieve some goal they want (Wigfield and Eccles 2000). The arrangement of achievement goals may be hierarchical, in that one may value doing a task today because it will help them achieve a later goal (Wentzel 2000).

According to this theory, students' motivation to achieve in science is connected to the combination of expectancies and values they hold about science. From our point of view, what one means by "science" is often ambiguous, and certainly can be expected to influence students' attitudes and motivation. Researchers in motivation seem to assume the term "science" refers equally well to school science and professional science. In this study we avoided asking students' their opinions about science generally, and instead asked them about specific tasks they participated in during a guided-inquiry instructional unit. On the assumption that attributions of expectancy and value are situational, students' perceptions of these can be expected to vary across tasks, and asking students their perceptions of specific tasks is more likely to generate usable knowledge than generalized surveys about science.

This study thus contributes to an understanding of student motivation and science learning in two ways. First, by looking at specific inquiry tasks rather than at science as a monolithic domain, we can begin to understand how features of inquiry instruction can influence students' situational motivation—something amenable to change. Second, our method and sample extend previous case studies in this area while connecting our findings to the research base on motivation.

Method

Our aim in this study was to find out how students perceived specific tasks of inquiry, after participating in those tasks during a 3 week guided-inquiry unit on plant biology

and adaptation, called *Sensing the Environment* (Griffis and Wise 2005). Following the completion of this unit, students were asked to rate the value of four inquiry tasks, using a procedure and materials described below.

We pursued four specific questions. First, how do students perceive the value of specific inquiry tasks in relation to their usual schoolwork? By value we mean both the interest and utility students perceive in these tasks, as well as the difficulty of these tasks. We are interested in gauging both the level of perceived value and students' reasons for their perceptions. Second, do students' perceptions differ by task? Do they find some tasks more interesting or useful (or harder) than others? Third, do boys and girls differ in their perceptions of the value of inquiry tasks? Given the history of gender differences in science attitude and motivation, it seems likely that boys and girls would perceive these tasks differently. Fourth, do the perceptions of task value differ by school or community? This is a way of asking whether there might be differences in perceptions of value related to ethnic or economic differences among students.

Participants

Participants included 129 seventh grade students (64 boys, 65 girls) from five schools in a large city in the western United States, participating in a field test of *Sensing the Environment*. This is a convenience sample. Teachers were recruited for the field test, and all participating teachers volunteered to teach the materials in at least one of their normal track science classes. Table 1 shows that the schools served economically and ethnically diverse student populations. Because participating students at each school were enrolled in normal track science classes, we assume they are a fair representation of the rest of their school. The communities where the schools are located run the gamut from poor, urban neighborhoods to relatively affluent beach communities. A range of attitudes and motivation toward science might be expected between schools.

Teachers at each school followed lessons plans developed by the research team, following training on the materials.

Instructional Context

Sensing the Environment (Griffis and Wise 2005) is an example of guided inquiry. The materials were developed as part of the educational mission of a Science & Technology Center funded by the National Science Foundation. The work of this center focuses on the deployment of remote sensing networks to support scientific research in a number of fields, including ecology. The aim of *Sensing the Environment* was to provide middle school students the opportunity to work with such sensing data as a resource for learning core science topics. The unit studied here asked students to consider the question, why do plants look different? The unit was comprised of ten lessons that took approximately 15 class periods to complete.

The unit followed an approach to guided inquiry that broke the unit down into two distinct phases: *staging activities* comprised of lab activities to help students explore foundational concepts behind the driving question (e.g. photosynthesis); and an *investigation* where they used these ideas to help them make sense of a complex data set (cf., Reiser et al. 2001). The focus of the investigation was for students, working collaboratively, to construct an explanation that could answer the driving question, using the data they gathered as evidence.

The unit started by showing students a photograph of an area from a southern California coastal mountain range. The photograph (a version of which is shown in the right side of Fig. 1) depicted a sunny hillside falling into a small creek bed, and rising to a smaller, shadier hillside. The hillsides and creek bed represent three distinct microclimate areas, where small but observable differences in temperature, humidity, and photosynthetically active radiation (PAR, referred to in the curriculum as "light intensity") lead to differences in the plants that grow in each of the three areas. Students looked at this photograph and

Table 1 Demographics of participating schools

	African-American	Asian-American[a]	Caucasian	Latino	Free or reduced lunch	N	
						Girls	Boys
School 1	25	8	9	57	75	9	7
School 2	65	4	23	7	36	19	18
School 3[b]	1	8	71	5	3	10	8
School 4	1	10	74	14	13	10	19
School 5	33	–	–	65	43	17	12

Ethnicities are reported as percentages of student body. N shows the number of students from each school participating in this study

[a] Includes Asian, Filipino, and Pacific Islander

[b] 15% of students reported mixed ethnicity or did not respond

Fig. 1 Online investigation tool for *Sensing the Environment*
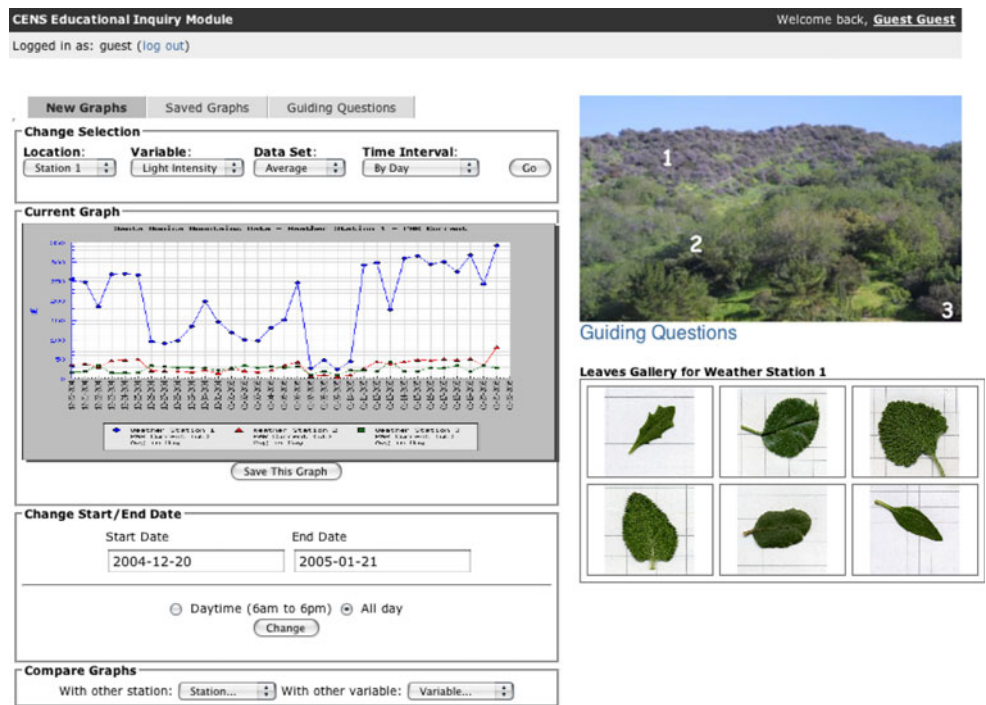
were asked by their teacher what they saw. After a range of observations was elicited, students were asked what questions they had about these observations and were guided by teacher prompts to pose a single question for the classroom: why do plants look different? This was the driving question for the unit.

Staging activities were framed around answering questions that followed from and could help answer this driving question. These sub-questions were written into the materials, but teachers were encouraged to elicit them, or similar questions, from students. The intent was for students to take a role in directing the course of their activity. Staging activities focused on the process of photosynthesis, then the process of transpiration, and finally labs that examined how climate variations might affect rates of transpiration (e.g., as leaves increase in surface area, would they transpire more quickly or more slowly?). These staging lessons were designed to scaffold students' learning of ideas and concepts relevant to the topic that could help them in their subsequent open-ended investigations, based on research showing the importance of such knowledge to effective inquiry (Zimmerman 2000). Yet, these were not typical cookbook labs where procedures and purposes were given to students. Teachers were provided, through training and lesson guides, with discursive prompts to encourage students' active contributions to the framing and conduct of activities. For example, after the initial activity posing the driving question, teachers led students in a discussion of what plants do that might lead them to take different forms. As predicted, students mentioned photosynthesis as an important activity of plants. Students were asked to draw a model of their understanding of photosynthesis. The variability in these models led naturally to questions about how the "ingredients" of photosynthesis get into leaves, which led to specific lessons where students explored these questions. This design follows closely on research-based recommendations for lab activities to be sequenced clearly and coherently in instruction, for students to be involved in the framing of the purpose of such activities, and for reflective discussion to be a central feature of such labs (NRC 2005).

The online investigation asked students to answer the driving question by exploring data collected by a sensor network deployed in the mountains they looked at to start the unit. As shown in Fig. 1, students chose a site and variable measured at that site, then a date range and aggregation factors for the variables (average, high, and low) and time scale (monthly, daily, weekly, hourly). Finally, students could choose to compare either the same variable at a different site or a different variable at the same site. This last restriction prevented students from making totally confounded queries. As can also be seen in the figure, the online tool presented photographs of leaves from plants found around each site, including their surface areas. Consequently, students had the data that could help them directly link differences in leaf size to differences in temperature, humidity, and light intensity. (Available leaf samples included both different species at each location and within-species differences between locations. Our aim was not that students specifically learn the theory of natural

selection, in which case within-species variation would be important. Rather, we wanted them to learn the general principle that morphological structures are related to environmental conditions; in this case, managing the photosynthesis–transpiration compromise in different microclimates. This decision was driven largely by the data available to use through the sensor network.)

Students had to make a number of choices about what data to look at and how to interpret resultant graphs. It was possible to generate hundreds of different graphs given the data set, meaning that a simple browsing of the data was impossible. Students, therefore, had to engage in central aspects of scientific inquiry, including the generation of tentative explanations and plans to find data that could evaluate those explanations. Conversely, explorations of the available data could generate candidate explanations for subsequent test. This investigative task was thus quite open-ended, with multiple plausible explanations for the driving question.

In summary, students' inquiry was guided in this unit in specific ways, and open-ended in other ways that are consistent with Chinn and Malhotra's (2002) analysis of research-based inquiry efforts and epistemological authenticity. They pointed out that most such efforts focus on issues of data collection and analysis. Our particular emphasis was on arguing with data: analyzing and interpreting data in order to advance a particular causal explanation for the driving question. Students did not pose their own independent questions, nor did they design their own experiments. They did, however, consensually frame the questions that they pursued as a class. While they did not choose the variables available for sampling in the online environment, their investigation demanded that they make decisions about which variables were important to look at, which comparisons could be informative, and over what time range such comparisons should be made. They also had to notice that the environmental variations could be connected to leaf variations. Finally, they had to organize their interpretations of specific data into coherent written arguments. This design is motivated by research suggesting the difficulty of students framing productive inquiry questions without substantial domain knowledge (Krajcik et al. 1998; Zimmerman 2000) and of the importance for lab activities to be embedded into a coherent sequence of instruction whose purposes are clear for students and include ample reflective discussion (NRC 2005). This design is also consistent with other online inquiry environments shown to improve students' scientific reasoning (e.g., Linn and Hsi 2000; Tabak and Reiser 2008).

Materials

We developed a written survey specifically for this study, drawing from expectancy-value theory (Eccles and Wigfield 1995). We were interested in students' perceptions of four tasks central to the unit: (1) designing queries to find the data they needed, (2) writing an explanation, (3) analyzing data, and (4) using data as evidence in their explanations. We focused on three of the sub-scales from Eccles and Wigfield's (1995) original questionnaire: intrinsic interest, extrinsic utility, and task difficulty. We selected these three because we believe they may be particularly amenable to change through instruction and they could be assessed without undue burden on students. We also wanted to know if students found these four tasks to be different from their typical science work. Students responded on a five-point Likert scale to each of the following questions: (a) was [the task] different from what you normally do in your science class? (1-not very different to 5- very different), (b) compared to what you normally do in your science class, how interesting was [the task]?, (1-not very interesting to 5-very interesting), (c) is it useful to [do the task]?, (1-not very useful to 5-very useful), and (d) compared to the work you normally do, how hard was [the task]?, (1- not very hard to 5-very hard). Students also were asked to justify their ratings in space provided after each question.

Our question about difference does not come from expectancy-value theory. We included it because we wanted to know whether or not students perceived these tasks as any different from what they usually do in their science classes. Advocates of inquiry believe that such tasks are atypical of most science instruction, but it is not at all clear that students believe this (Sandoval 2005). Although our prompt stems for interest, utility, and difficulty asked students to compare each task to their usual activities, our scales were presented in absolute terms and we found that students responded as such. Consequently, we use the difference ratings as comparative and the other ratings as straightforward evaluations of the tasks.

Data Collection and Analysis

Students completed the survey following the completion of the unit, in their regular classrooms. The survey took about 10 min to complete. Students were asked to answer all of the questions quietly on their own. Surveys were administered and collected by researchers. The surveys yielded two sources of data: Likert scale ratings of difference, interest, utility, and difficulty; and text justifications of those ratings.

Students rated each dimension on a scale from 1 to 5, with 5 always being the highest value (e.g., most interesting, or most difficult). Ratings were entered into SPSS 16 for analysis. Missing data was rare, with less than one tenth of one percent of ratings missing. Missing ratings were excluded casewise for each analysis. Specific

statistical analyses are described in the results section in relation to the question they were intended to examine.

Students' justifications for their ratings were transcribed verbatim. As students wrote justifications for each of the four dimensions on each of four tasks, the result was approximately 2,000 free text responses of 1–2 sentences each. Our approach was to use findings from motivation research to generate *a priori* coding themes to apply to the dimensions of interest, utility, and difficulty. On the interest dimension we looked for indicators of intrinsic value; for example, students enjoying the opportunity to look at real data. We also looked for expressions of interest related to classroom factors that are known to influence motivation. For example, students may have found tasks interesting because of their increased autonomy, or less interesting if the tasks felt authoritative. On the utility dimension we looked for expressions of utility value, such as being able to achieve success in school from knowing how to do these things, or the benefit from knowing how to do them later in life. Considering the difficulty dimension, we looked for expressions of students' ideas about their own ability to do these tasks. For the difference dimension, we took a different approach, since this dimension was not derived from motivation research. Instead, we expected that students might consider the unit different because it was on a new topic, or it included a different kind of work (e.g., working on the internet), or that it was not different at all. Themes for difference justifications, therefore, were derived from students' responses.

The themes we developed are presented in Table 2. We expected that any of the themes could potentially appear as justifications for any dimension. For instance, students might rate their interest in a task highly because of its difference from their usual science work, or its utility value. To clarify the *restatement* code, it could mean one of two things. Students commonly gave justifications that simply verbalized their rating. For example, an interest rating of 1 (the lowest) might be explained by, "it was boring." Or a difficulty rating might be explained by "it was easy" or "it was hard." These justifications add no insight into the reason behind the initial rating. We included the very small number of blank responses, less than one percent of the sample, in this code too, simply to avoid creating a separate code for them.

A subset of 200 justifications, fifty randomly selected from each dimension, was coded independently by each of the authors. Inter-rater agreement ranged from 74 to 82% across dimensions, with kappa values ranging from .65 (utility) to .90 (difficulty), indicating good to very good agreement between raters (Kraemer 1982). Disagreements between the two coders were resolved through discussion. All the remaining justifications were coded by the first author. We use examples of justifications here to illustrate the kinds of reasons students' gave for their ratings.

Results

We organize our presentation of results in terms of the questions we asked. First, we explore patterns in students' rating of the four inquiry tasks along each of the four dimensions of value. Then we examine the kinds of

Table 2 Coding themes applied to rating justifications

Theme	Description
Autonomy	Any expression of value that cites doing things for oneself, making choices and decisions, not being told what to do
Authority	Any expression of value that cites teacher or other authority as its source
School utility	Expression of value that cites the utility of a task for school, either now or in the future
Future utility	Expression of value that cites the utility of a task in the future after school, either for work explicitly or just "in life"
Positive expectancy	An expression of value based on an expectation to do well on a task
Negative expectancy	An expression of value based on an expectation to do poorly on a task
Personal value	Any expression of value, or lack, that does not fit one of the above categories
Different topic	An expression of value based on the topic of study being different from what has been studied before
Different work	An expression of value based on the work being atypical
Same	An expression of lack of value that cites the task as being common, usual, or normal for science class
Computers	Expression of value, or lack, due to working with computers or the internet
Collaboration	Expression of value, or lack, from working in groups
Restatement	Justification is just a restatement of the rating. Blank justifications and the rare uninterpretable justification were included in this category

Some themes can have positive or negative valence. Themes may appear on multiple dimensions. The term *value* in the description column includes any of the dimensions of interest, utility, or difficulty

justifications that students provided and their relation to high and low ratings.

How do Students Perceive Inquiry Tasks?

The first question of our study is simply how do students perceive the value of these four inquiry tasks, and their difference from typical schoolwork? Table 3 summarizes task and dimension ratings for all students. Overall, students perceived these tasks as moderately different from what they usually did. They did not find them especially interesting or difficult, but they did find them useful. All tasks were rated most highly on the utility dimension and the overall utility rating was high.

The range and distribution of ratings varied considerably across the four dimensions of difference, interest, utility, and difficulty. Difference ratings were distributed roughly equally across the five possible ratings, within a range of 16% (1) to 24% (3 and 5). Ratings for interest, utility, and difficulty all showed noticeable skews. Seventy-one percent of interest ratings were 3 or lower, as were 75% of difficulty ratings. In the other direction, 70% of utility ratings were 4 (25%) or 5 (45%).

Value Ratings by Task

The next step in our analysis was to determine whether or not ratings within dimensions vary by task. Do students find some tasks more interesting, useful, or difficult than others? (We treat ratings of difference separately, below.) We conducted repeated measures analyses of variance for each dimension independently, using task as the within-subjects factor, with pairwise contrasts to see which task differences contributed to any main effect. Ratings of *interest* differed significantly by task, $F(3, 129) = 13.79$, $p < .001$; pairwise comparisons showed, as suggested in Table 3, that students rated each of design and analysis as more interesting than explanation and evidence use. *Utility* also differed by task, $F(3,129) = 6.00$, $p = .001$; with analysis and evidence use each being rated more useful than design and explanation. Finally, *difficulty* ratings also

Table 3 Observed mean ratings (and standard deviations) on each dimension of each task across the whole sample

	Difference	Interest	Utility	Difficulty
Design	3.40 (1.11)	3.02 (1.31)	3.80 (1.13)	2.78 (1.35)
Explain	3.23 (1.38)	2.38 (1.38)	3.86 (1.18)	2.85 (1.32)
Analysis	3.29 (1.48)	2.96 (1.41)	4.18 (1.11)	2.36 (1.24)
Evidence	2.78 (1.48)	2.36 (1.27)	4.15 (1.07)	2.32 (1.33)
OVERALL	3.18 (1.39)	2.68 (1.37)	4.00 (1.14)	2.58 (1.33)

Significant differences are noted and discussed in the text

varied by task, $F(3, 129) = 9.33$, $p < .001$; with design and explanation each rated as more difficult than analysis and evidence use.

The design task, then, was rated as more interesting and more difficult. Explanation was less interesting, less useful, and more difficult. Analysis was more interesting and more useful, and less difficult. Finally evidence use was less interesting and difficult, but more useful. The writing tasks, explanation and evidence use, tended to be less interesting, while the more informal interactive tasks of design and analysis were more interesting. At the same time, the analysis and use of data as evidence were seen as most useful.

Do Ratings Vary by Gender?

A visual inspection of mean ratings of boys and girls on each task within each dimension suggested potential gender differences. These were explored by running the same repeated measures ANOVAs as above, with gender as a between-subjects factor. We found no significant effects of gender on ratings of interest, utility, or difficulty; nor did we find any interactions between task and gender on any of these dimensions.

Do Ratings Vary by SES?

We wondered whether students' ratings might vary according to their socioeconomic status, as reflected in their schools' demographics. We did not choose a sample to test this question deliberately, so our analyses on this question should be regarded as tentative. At the same time, an inspection of the demographics of our sample schools suggested an important opportunity to explore the question, especially given the dearth of such research from inquiry-oriented efforts.

We split our sample into two groups of schools based on the racial/ethnic composition of the student body and the number of students receiving free or reduced lunches. Three of our schools (1, 2, and 5) were comprised of at least two-thirds majorities of African-American and/or Latino students. These schools were also located in communities of medium to high poverty (using definitions from NCES 2009). On the other hand, Schools 3 and 4 served predominantly White students in communities in or near the lowest quintile of poverty defined by NCES (the cutoff is 11% free/reduced lunch). Given the history of achievement differences in science between White students and students of color (Grigg et al. 2006; O'Sullivan et al. 1997) this seems a justifiable split for beginning to explore potential SES differences in students' perceptions of inquiry task. We labeled Schools 1, 2, and 5 as the "low SES" group and Schools 3 and 4 as the "high SES" group. We stress that our use of these

Table 4 Dimension ratings for each task by low and high SES schools

	Difference		Interest		Utility		Difficulty	
	Lo	Hi	Lo	Hi	Lo	Hi	Lo	Hi
Design	3.23 (1.22)	3.65 (1.01)	3.27 (1.36)	2.73 (1.17)	4.10 (1.01)	3.33 (1.22)	2.83 (1.39)	2.58 (1.29)
Explain	3.37 (1.33)	2.90 (1.51)	2.67 (1.54)	2.04 (0.99)	4.15 (1.14)	3.53 (1.14)	2.76 (1.37)	2.75 (1.32)
Analysis	3.30 (1.46)	3.23 (1.54)	3.39 (1.35)	2.25 (1.23)	4.50 (0.99)	3.60 (1.22)	2.36 (1.26)	2.37 (1.25)
Evidence	2.67 (1.49)	2.90 (1.46)	2.50 (1.33)	2.15 (1.16)	4.42 (0.91)	3.67 (1.18)	2.34 (1.38)	2.18 (1.20)
OVERALL	3.14 (1.40)	3.17 (1.42)	2.96 (1.44)	2.29 (1.17)	4.29 (1.02)	3.53 (1.19)	2.57 (1.37)	2.47 (1.28)

terms is relative and used only to distinguish these groups in our sample. Given the wide variability of poverty in our "low SES" group, any significant differences between these two groups deserve further study.

We conducted repeated measures ANOVAs for interest, utility, and difficulty using task as a within-subjects factor and SES grouping as a between-subjects factor. There was a significant main effect of SES on *interest*, $F(1, 129) = 11.56$, $p = .001$, and on *utility*, $F(1, 129) = 28.96$, $p < .001$. Task and dimension ratings by SES group are shown in Table 4. Notice that the students in low SES schools rated all of the four tasks more interesting and more useful than their peers in high SES schools. It is also worth noting that there were no differences in difficulty ratings between the groups.

Perceptions of Difference

One might wonder whether or not the value that students perceive in these tasks is related to how different they seem from what the students typically do in their science classes. We might also wonder whether or not perceptions of difference are associated with the socioeconomic status of the schools students attend, as inquiry-oriented instruction may be perceived as more different than the usual instruction in the more disadvantaged schools if, in fact, they tend to receive the "pedagogy of poverty" (Haberman 1991). As with the other dimensions, we ran a repeated measures analysis of variance with difference as a within-subjects factor and SES as a between-subjects factor. There was a main effect of task, $F(3, 129) = 7.78$, $p < .001$. Post-hoc comparisons showed that evidence use was perceived as less different than all of the other three tasks (see Table 3 for reference). There were no main or interaction effects for SES, however, suggesting that regardless of the affluence of the school students attended they perceived the difference of these tasks similarly.

Justifications for Ratings

While the Likert scale ratings provide a sense of the level of perceived value and difficulty of each of the four tasks

we asked about, we also wondered what reasons students gave for their ratings and whether these reasons would be consistent with prior motivation research. We derived our coding scheme (Table 2) to apply across all four dimensions was asked about, while presuming that some themes may be more apparent in justifications for particular dimensions. Because of our small sample size and the relatively large number of themes identified in students' responses we have not attempted to statistically associate justification themes to particular ratings or tasks. Instead, for each dimension we can provide frequencies of the kinds of themes expressed and examples of what students said. In what follows, we refer to low ratings as ratings of 1 or 2, and high ratings as ratings of 4 or 5. All justifications are quoted verbatim, without corrections to spelling or punctuation. Percentages reported for justifiation frequencies are rounded to the nearest whole number. We present only those themes that appeared in more than.

Interest

The majority of justifications students provided for interest ratings, 62%, were simply to re-state their rating. For low ratings, restatements included justifications such as, "It was boring," "It was kinda boring," or "It was not interesting because I didn't like it." Occasionally, we coded justifications as restatements when they did not fit clearly into another theme and were rare, such as, "It took a lot of time," or the apparently sarcastic, "We just analyzed data! Wow!" to explain a rating of 1. Similarly, high ratings often evoked simple re-statements like, "it was interesting," or "it was kinda fun."

Other justification themes given in at least two percent of responses included *autonomy* (10%), *same* (8%), *negative expectancy* (8%), and *personal value* (5%). Autonomy justifications were more common for high ratings, and included general ideas like, "It was interesting because we did it by ourselves," and more science-specific ideas like, "it was pretty interesting to actually feel like a scientist and collect data." Many autonomy justifications referred to the gathering and analysis of data, of having to figure out the data for themselves, as what made the tasks interesting.

Similarly, expressions of personal value were more commonly given for high ratings, and generally referred to the value of getting to learn new things (e.g., "it was interesting because I got to learn new things," or "you got to see what the plant did'), and the value of checking your own knowledge, as in "you get to see how much you really know and how well you take in information."

Negative expectancy and sameness justifications were more commonly given for low ratings. Sameness justifications included statements like, "we do this about every day," or "we do it in labs." Note that sameness was a reason for not being interested, it was never given for a high interest rating. Students' expressions of negative expectancy most commonly cited writing as what they did not find interesting, as in "I don't like write a lot" or "I'm not the best writer," although these justifications also cited other aspects of the tasks that students found hard, like "I didn't know what to do," or "it was hard to do in such a short time."

Utility

Recall that utility ratings were skewed toward high ratings, with very few low ratings at all. Utility justifications included the fewest number of restatement justifications, at 31%. The substantive justifications for high ratings fell primarily into three themes: *personal value* (28%), *school utility* (18%), and *future utility* (17%). A common sentiment expressed in *personal value* justifications was the idea that it is important to have data to back up one's claims, "people can't prove you wrong with facts," or "we need data to support our statements." These justifications asserted a general importance or value to the tasks without any explicit reference to their value in school or in future jobs. Justifications in terms of school utility included statements about the utility of knowing how to analyze data and express claims. Sometimes these justifications just asserted that it would be necessary to know how to do these tasks later in the current science class, or in later classes; other times they asserted that "other teachers may want us to do this," leaving it unclear if they meant other science teachers or other subjects. Occasionally, students acknowledged school utility even while not feeling it personally, as in "I'm not planning on becoming a scientist, but for later school years, I need to know how to" [collect data].

Justifications of the future utility of knowing how to do these tasks included statements that asserted a future value outside of school. These justifications seemed mainly of two types: a general assertion of later utility ("One day you might need to know this," or "We will need it in future years"); or a claim of job utility ("most jobs like a la[w]yer you need it," "there will be a lot of that in most careers," or "you always have to write explanations in real life").

Difficulty

Forty-seven percent of difficulty justifications were *restatements*. *Negative expectancy* justifications (14%) included a range of reasons why students found the work hard, "I thought it was a little bit hard because it was a lot of information," "It was confusing and it wasn't explained very well," "my normal work I do is take notes and I study and then I take a test. It was hard for me to do my own experiments," "I'm not very smart in science so to explain myself about this subject was hard." As with negative expectancy justifications for interest ratings, students often mentioned that writing was hard or they did not like to write. Students also pointed to the novelty of the kind work as a source of difficulty (*different work*, 4%): "we don't normally do an explanation," or "very hard cause I never did a experiment like that."

The most common substantive reason for low difficulty ratings were *positive expectancy* justifications (13%). These included general claims of ease like, "this was easy for me." These justifications also included specific assertions about the ease of particular tasks: "you only have to gather and analyze data," "You just plug in evidence," or "The data was right in front of your faces." The other common justification for a low difficulty rating was that the tasks were the *same* (10%) as what students usually did, such as "we do it all the time," "Not very hard because we practiced to do an explanation," or "because in science class you always support your claim with data." Both sameness and positive expectancy justifications, then, suggest students found these tasks easy because they believed they were familiar tasks in the science class. Finally, about 6% of the justifications students gave mentioned using the *computer* made things easier, "the computer helped a lot," or that it was "not very hard when you have the internet."

Difference

The most common justification for difference ratings was *restatement* (46%), simple assertions that "it was not different," or "we never do this." Thirty percent of the justifications asserted differences in the type of work that students did in this unit from their usual work. Of these, 9% simply said that using computers was different, whereas the other 21% cited some specific difference in activity. These included references to doing their own research ("we don't usually research like that," or "normally we take notes"), gathering data ("yes, because we haven't gathered data

before," "we don't really use data," or "we don't usually use graphs for working"), and writing explanations ("we don't usually write essays," or "I wasn't used to writing explanations"). The novelty of writing essays was a prominent reason students gave for high difference ratings. Also common in this category were claims that "we usually use our textbooks."

On the other hand, 21% percent of students said these tasks were the *same* as what they usually did, saying "we do similar things a lot," "we usually have to explain things," or "we normally use data to support our claim."

Discussion

There is a general understanding of the features of classroom environments that support increased student motivation: autonomy, appropriate challenge, personal value of the tasks, supportive teachers, and mastery goal orientation in the classroom (Anderman and Young 1994; Pintrich et al. 1993; Ryan and Patrick 2001). Popular conceptualizations of inquiry-oriented science implicitly suggest that these features are a part of inquiry. This study explored whether or not students perceived value in a selection of inquiry tasks for some of the above reasons. Our aim is to understand how specific learning tasks and activities can be organized to promote motivation (and related constructs like interest and engagement). Ultimately, the aim is not just that students are interested in learning science or have fun, but that they see that science has real value in their lives in and out of school, whatever their aspirations might be. We reiterate this framing of our study to properly contextualize our findings. Obviously, our findings say as much about our particular instantiation of inquiry as they do about students. We focus our discussion here on how features of our curriculum may have influenced students' ratings of these tasks, for better and worse. In doing this, we emphasize what we see as important findings in relation to extant research.

This study aimed to answer four broad questions. How do students perceive the value of specific inquiry tasks in relation to their usual schoolwork? Do students' perceptions differ by task? Do boys and girls differ in their perceptions of the value of these tasks? Do perceptions of these tasks vary by ethnic or economic community?

Patterns in Perceptions of Value

Overall, students perceived these tasks as useful, moderately interesting, but not especially difficult. While utility ratings were high on all four tasks, they were significantly higher for the tasks of data analysis and evidence use. What students found most useful about these tasks was having

the ability to make sense of data that they could the use to back up their ideas. Students here were particularly articulate about the utility of these various tasks for success in school and out, now and in the future. They talked about the value of having data to back up your ideas, and the value of knowing how to get and interpret data for that end. Both boys and girls perceived this value, and this suggests an opportunity for science educators to focus on the value of being able to do particular kinds of things, as opposed to alleging the value of knowing particular sorts of stuff or of an ambiguous "science" generally.

Girls did not differ from boys in their ratings of these tasks, and we think that is an important result. The lack of difference in boys' and girls' ratings of these tasks stands in stark contrast to the bulk of research on motivation and interest in science (Koballa and Glynn 2007; Osborne et al. 2003). We think there are two plausible reasons for this. One could be that our curricular approach is particularly egalitarian in the opportunities it provides to both boys and girls to find something of interest, in contrast to typical school science. While this may be the case, we have no data to evaluate such a claim one way or another. We might expect, were this so, that ratings for difference might be high overall, or higher for girls than boys to indicate that the girls perceived such opportunities as different from their usual science. We did not find this. A second reason for a lack of gender differences could be how we asked our questions. Research on science motivation or attitude routinely asks students how they feel about "science," which leaves it open to the respondent to conjure up whatever image of science is in their heads. As reviewed by Osborne and colleagues, the images that girls typically conjure up are not especially positive. We may have sidestepped that issue by asking about specific tasks. We think this is the likely explanation for the lack of gender differences observed here. Further research could explore more fully the bases of boys' and girls' perceptions of value. It could also be the case that the lack of gender difference is disciplinary—women are well-represented in biology compared to other science disciplines. There may be something about biology that girls find appealing. Contrasts with other science subjects would be a useful extension of this work.

The other striking result here is the difference in interest and utility ratings between students at lower and higher SES schools. Students at the three lower SES schools in our sample were much more likely to rate utility highly (a rating of 4 or 5) than students at higher SES schools. A tempting explanation for these differences is that our instructional materials contrast with the usual "pedagogy of poverty" (Haberman 1991) of these lower SES schools. If this were the case, students did not seem to perceive that difference. Students did perceive these tasks as moderately

different from their usual schoolwork, but difference ratings did not vary between the higher and lower SES groups. It may be that the children in the lower SES schools valued this difference more highly, that they somehow saw it as enabling greater autonomy in their activity, and hence higher interest and utility. There is an accruing body of evidence that typically low-performing urban students achieve especially strong learning gains from well-structured inquiry instruction (NRC 2005). Our findings suggest another benefit of inquiry for these students—they find it valuable. An important area for future research would be to explore these findings with a larger sample of schools and more robust methods of observation. We emphasize two points from our own analyses. First, it is clear that students from the lower SES schools perceive value in these tasks, and this suggests that providing more, and perhaps more varied, opportunities for authentic inquiry is likely to improve such students' engagement with science. Second, it is also clear that these students did not perceive these tasks to be more challenging than they could handle, which should give the lie to perceptions that ambitious instruction is inappropriate for underserved, disadvantaged students. We consider our findings regarding SES differences to be tentative. We did not specifically sample to test for such differences, an obvious limitation to any claims we might make. Our findings suggest research on such differences is warranted, specifically research that could identify the group or individual factors that might underlie such differences.

The patterns of ratings, and their justifications, raise several questions. Why did students rate all tasks as useful? Why did they rate data collection and analysis as more interesting than explanation and evidence use? Why did they rate data collection and explanation as harder than analysis and evidence use? The high utility ratings on all tasks appears, from students' justifications, to reflect an idea that knowing how to do these things is helpful in school or out of school. Students commonly asserted that it was important to know how to do these tasks in order to do well in their current science class, or in future science classes or levels of school. Students also commonly noted that they thought they would need to know how do these things later in life, particularly in the workplace. Taking such justifications at face value, as a group these students appear to believe that knowing how to gather your own data and make sense of it to explain something is an important set of skills for future success, in school or out. Alternatively, they at least recognize that it is a socially valued skill. More cynically, students may simply have been saying that they recognize that they are expected to see such tasks as useful. Discriminating between these possible explanations requires probing students' justifications more deeply than we could in this study.

While interest ratings on all tasks were moderate, students found the tasks of data collection and analysis to be more interesting than writing explanations or using evidence in those explanations. Explanation was also rated as being the most difficult of the four, and low interest ratings for explanation were also justified by assertions that it was hard or that students did not like to do it (perhaps because it was hard). Students' justifications for interest ratings suggest they found data collection and analysis more interesting because of the autonomy and challenge involved in figuring things out for themselves. Students had the freedom to explore the online environment in any manner they wished, with the only constraint that they could not construct confounded queries (i.e., compare two different variables at two different sites). They also, obviously, could construct any interpretation of the data that made sense to them. On the other hand, the explanation task required students to make the effort to articulate those interpretations in writing. Many students cited the difficulty of writing for their lack of interest in the explanation task, and it is not surprising that interest would be tempered by difficulty. It seems likely that students had little experience writing their own explanations in these classes, based on what we know of typical science instruction (see NRC 2005) and from the commonness of that claim to justify high difficulty ratings. The materials we developed were designed to scaffold students in constructing explanations, but in terms of linking data they found as evidence for the claims they wanted to make. Writing, per se, was not specifically targeted for support. Our findings suggest addressing this difficulty may be important for raising some students' interest and engagement.

Motivation, Interest, and Inquiry

Our findings suggest how specific features of inquiry learning environments may influence known situational factors of motivation. The tasks of data collection and analysis are primarily tasks of investigation, whereas the tasks of explanation and evidence use are tasks of argumentation, and in our case, primarily writing tasks. What we see in the overall pattern of ratings and justifications in our sample is higher interest in the tasks of investigation compared to the tasks of argumentation. As would be predicted by motivation research, students' perceptions that these tasks offered the autonomy of figuring something out for themselves underlie high ratings of interest. Related to this, there is little evidence here that students thought the challenge in such autonomous investigation

was too difficult. Students appeared to highlight the analysis of data as the crux of their experience, explicitly noting the value of being able to have evidence to back up your own ideas.

Within the terms of motivation research, students perceived these investigative tasks as being at an appropriate level of challenge and offering autonomy. There is some evidence that students perceived these tasks as authentic, with students noting the value of working with "real data." Also, those students who rated these tasks as different from their typical schoolwork commonly cited doing their own research as the reason (along with using the computer). We do not wish to make too strong a claim to authenticity, as we realize the ways in which our approach limited some aspects of authentic scientific investigation. Our main point is that, not surprisingly, students who found these tasks interesting and useful gave reasons that were expected from prior research on motivation.

The more important point lies in understanding what aspects of our findings may generalize to other models of inquiry. Our approach to structuring these units emphasizes the construction of explanations from data. Other approaches might emphasize the asking of questions, or experimental design. Such differences in approach would lead to different kinds of tasks being designed and enacted, and probably different interpretations from students. Clearly, we cannot simply say that students found inquiry useful, or somewhat interesting. Our investigative tasks, particularly collecting data, were purposely constrained. Students did not define variables, they did not take their own measurements. Instead, they had to decide how to generate meaningful (to them) queries from the sensor database. Thus, several choices that might contribute to feelings of autonomy were not available to them. It may well be the case that inquiry interventions that promote such choices would generate higher interest. A question worth pursuing is whether the difficulty of posing questions and designing experiments would moderate students' interest, or whether the autonomy of designing investigations would outweigh the challenges involved. We suggest the approach we have taken here is a fruitful avenue for collecting information about such perceptions from large numbers of students, and that such data can potentially inform the design of inquiry learning environments.

For those interested in promoting argumentation in science instruction, our results are somewhat cautionary. Students appear to see these argumentation tasks as the most difficult, apparently because of the difficulty of writing. One alternative may be to provide students alternatives to writing explanations. For example, other efforts at supporting argumentation are structured around debates over competing simple claims (Bell and Linn 2000; Clark and Sampson 2007). While these approaches engage

students in evaluations of data, they do not really promote students' articulation of their own thinking (students typically choose alternative articulations of claims), nor are they necessarily viable for a number of science topics of interest, as causal theories in science are often complex (Perkins and Grotzer 2005). Moreover, helping students learn to write is not only an important aspect of understanding scientific argumentation, it is an important vehicle for literacy development for ethnic minority students (Lee and Fradd 1998). We conclude then, that interventions like ours that emphasize complex writing may have to support writing, per se. Of course, one way of doing so is just to provide students with repeated opportunities to construct explanations over the school year, and perhaps they would come to find it less difficult.

Conclusions

This study is obviously exploratory, and limitations in our sample and methods caution against overly strong claims. Still, to our knowledge, we are the first to go beyond very small-scale case studies (Mistler-Jackson and Songer 2000; Patrick and Yoon 2004) to explore motivational questions as they specifically pertain to inquiry-oriented instruction. Clearly an exploratory study such as this one raises a number of questions that deserve further study. It would obviously be valuable for these findings to be replicated. Beyond this, it would be quite valuable to combine the survey method used here with closer observations of classroom instruction and interaction, and with interviews with students that could explicate their ratings justifications. Such methods would better elucidate and contextualize students' perceptions and be better able to link them to features of inquiry instruction. The important issue is that assessments of motivation and interest will be more useful if they focus on students' perceptions of specific tasks or practices of science, rather than querying students' perceptions of an ambiguous "science" that they do not know.

With such caveats in mind, our findings suggest specific implications for inquiry oriented learning environments. One is that students find investigation tasks interesting and useful, and increasing autonomy in such tasks may further increase students' perceptions of value. A second is that argumentation tasks, currently much in vogue in the science education research community, probably have to closely attend to issues of interest and difficulty. This might mean providing scaffolds to support writing, or possibly it means making the purpose of such argumentation more inherently interesting, or both. We also think that engaging students in thinking about the value of particular kinds of tasks—knowing how to do certain things and think

in certain kinds of ways—may be a means of increasing girls' and historically underserved students' interest and motivation in science. Future research can productively examine how attributes of inquiry-oriented instruction can affect students' perceptions of the value of engaging in such tasks, and learning and doing science.

References

Anderman EM, Young AJ (1994) Motivation and strategy use in science: individual differences and classroom effects. J Res Sci Teach 31(8):811–831

Bell P, Linn MC (2000) Scientific arguments as learning artifacts: designing for learning from the web with KIE. Int J Sci Educ 22(8):797–817

Blumenfeld PC, Kempler TM, Krajcik JS (2006) Motivation and cognitive engagement in learning environments. In: Sawyer RK (ed) The Cambridge handbook of the learning sciences. Cambridge University Press, New York, pp 475–488

Chinn CA, Malhotra BA (2002) Epistemologically authentic inquiry in schools: a theoretical framework for evaluating inquiry tasks. Sci Educ 86:175–218

Clark DB, Sampson V (2007) Personally-seeded discussions to scaffold online argumentation. Int J Sci Educ 29(3):253–277

Eccles JS, Wigfield A (1995) In the mind of the actor: the structure of adolescents' achievement task values and expectancy-related beliefs. Personal Soc Psychol Bull 21(3):215–225

Eccles JS, Wigfield A (2002) Motivational beliefs, values, and goals. Annu Rev Psychol 53:109–132

Graham S, Taylor AZ (2002) Ethnicity, gender, and the development of achievement values. In: Wigfield A, Eccles J (eds) The development of achievement motivation. Academic Press, San Diego, pp 121–146

Griffis K, Wise J (2005) Sensing the environment. Center for Embedded Networked Sensing, University of California, Los Angeles, CA

Grigg W, Lauko M, Brockay D (2006) The Nation's report card: science 2005. National Center for Education Statistics, Washington, DC

Haberman M (1991) The pedagogy of poverty versus good teaching. Phi Delta Kappan 73:290–294

Koballa TR Jr, Glynn SM (2007) Attitudinal and motivational constructs in science learning. In: Abell SK, Lederman NG (eds) Handbook of research on science education. Lawrence Erlbaum Associates, Mahwah, NJ, pp 75–102

Kraemer HC (1982) Kappa coefficient. In: Kotz S, Johnson NL (eds) Encyclopedia of statistical sciences. John Wiley and Sons, New York

Krajcik J, Blumenfeld PC, Marx RW, Bass KM, Fredricks J (1998) Inquiry in project-based science classrooms: initial attempts by middle school students. J Learn Sci 7(3&4):313–350

Lee O, Fradd SH (1998) Science for all, including students from non-English-language backgrounds. Educ Res 27(4):12–21

Linn MC, Hsi S (2000) Computers, teachers, peers. Lawrence Erlbaum, Assoc, Mahwah, NJ

Lynch S, Kuipers J, Pyke C, Szesze M (2005) Examining the effects of a highly rated science curriculum unit on diverse students: results from a planning grant. J Res Sci Teach 42(8): 912–946

Minner DD, Levy AJ, Century J (2010) Inquiry-based science instruction—what is it and does it matter? Results from a research synthesis years 1984 to 2002. J Res Sci Teach 47(4): 474–496

Mistler-Jackson M, Songer NB (2000) Student motivation and internet technology: are students empowered to learn science? J Res Sci Teach 37(5):459–479

NCES (2009) The condition of education (Publication., from US Deptartment of Education: http://nces.ed.gov/programs/coe/2009/

Nolen SB (2003) Learning environment, motivation, and achievement in high school science. J Res Sci Teach 40(4):347–368

NRC (2005) America's lab report: investigations in high school science. National Academy Press, Washington, DC

O'Sullivan CY, Reese CM, Mazzeo J (1997) NAEP 1996 science report card for the nation and states. National Center for Education Statistics, Washington, DC

Osborne J, Simon S, Collins S (2003) Attitudes towards science: a review of the literature and its implications. Int J Sci Educ 25(9):1049–1079

Patrick H, Yoon C (2004) Early adolescents' motivation during science investigation. J Educ Res 97(6):319–328

Perkins DN, Grotzer TA (2005) Dimensions of causal understanding: the role of complex causal models in students' understanding of science. Stud Sci Educ 41:117–166

Pintrich PR, Marx RW, Boyle RA (1993) Beyond cold conceptual change: the role of motivational beliefs and classroom contextual factors in the process of conceptual change. Rev Educ Res 63(2):167–199

Reiser BJ, Tabak I, Sandoval WA, Smith BK, Steinmuller F, Leone AJ (2001) BGuILE: strategic and conceptual scaffolds for scientific inquiry in biology classrooms. In: Carver SM, Klahr D (eds) Cognition and instruction: twenty-five years of progress. Lawrence Erlbaum, Mahwah, NJ, pp 263–305

Ryan AM, Patrick H (2001) The classroom social environment and changes in adolescent's motivation and engagement during middle school. Am Educ Res J 38(2):437–460

Sandoval WA (2005) Understanding students' practical epistemologies and their influence on learning through inquiry. Sci Educ 89:634–656

Schroeder CM, Scott TP, Tolson H, Huang T-Y, Lee Y-H (2007) A meta-analysis of national research: effects of teaching strategies on student achievement in science in the United States. J Res Sci Teach 44(10):1436–1460

Simpson RD, Koballa TR Jr, Oliver SJ, Crawley FE III (1994) Research on the affective dimension of science learning. In: Gabel DL (ed) Handbook of research on science teaching and learning. Macmillan, New York, pp 211–234

State of Michigan (1993) Chemistry that applies. State of Michigan, Lansing, MI

Tabak I, Reiser BJ (2008) Software-realized inquiry support for cultivating a disciplinary stance. Pragmat Cognition 16(2):307–355

Thadani V, Cook MS, Griffis K, Wise JA, Blakey A (2010) The possibilities and limitations of curriculum-based science inquiry interventions for challenging the "pedagogy of poverty". Equity Excell Educ 43(1):21–37

Wentzel KR (2000) What is it that I'm trying to achieve? Classroom goals from a content perspective. Contemp Educ Psychol 25:105–115

Wigfield A, Eccles J (2000) Expectancy-value theory of achievement motivation. Contemp Educ Psychol 25:68–81

Zimmerman C (2000) The development of scientific reasoning skills. Dev Rev 20:99–149