Scientific Reasoning and Epistemological Commitments: 
Coordination of Theory and Evidence Among College Science Students

Ava Zeineddin,1 Fouad Abd-El-Khalick2

1Teacher Education Division, College of Education, Wayne State University, 277 Education Building, Detroit, Michigan 48202
2University of Illinois at Urbana-Champaign, Champaign, Illinois

Received 24 March 2009; Accepted 30 December 2009

Abstract: Reasoning skills are major contributors to academic and everyday life success. Epistemological commitments (ECs) are believed to underlie reasoning processes and, when considered, could do much in delineating the complex nature of scientific reasoning. This study examined the relationship between ECs and scientific reasoning among college science students. Prior knowledge (PK) was factored in as an intervening variable. Participants were 139 college students enrolled in two physics courses in a large Midwestern university. They completed an online questionnaire, which assessed their PK regarding buoyancy in liquids and EC to the consistency of theory with evidence. Responses to the online questionnaire were used to select 40 participants with varying levels of PK and EC. These participants were divided into four groups, each with 10 students, representing four conditions: High PK–High EC, High PK–Low EC, Low PK–High EC, and Low PK–Low EC. These groups allowed using a 2 × 2 factorial quasi-experimental design to examine the relationship between participants’ reasoning and ECs, accounting for their PK. The quality of participants’ reasoning was assessed during individual interviews, which presented them with four problem-solving tasks involving objects immersed in water. Two-way analysis of variance (ANOVA) indicated the absence of interaction between PK and EC. The results showed that the higher the ECs were, the higher the quality of reasoning was for comparable levels of PK. Additionally, it was found that PK impacted reasoning more strongly when ECs were weaker.

Keywords: scientific reasoning; epistemological commitments; prior knowledge; hydrostatics

Science education reform documents have long emphasized helping students develop scientific reasoning skills as a major goal for science education (e.g., American Association for the Advancement of Science, 1990, 1998; National Research Council, 1996). Educators believe that reasoning skills play an important role in students’ ability to develop scientific understandings and conduct scientific investigations (Kuhn, 1989; Lawson, 2005; Lawson, Banks, & Logvin, 2007; Samarapungavan, 1992; Schauble, Glaser, Duschl, Schulze, & John, 1995; Tyler & Peterson, 2003). Scientific reasoning skills mark the development of adolescent cognition (Klaczynski, 2000) and are often demanded for effective decision-making and problem-solving (Greenhoot, Semb, Colombo, & Schreiber, 2004; Hogan, 2002; Holyoak & Morrison, 2005; Overton, 1990; Williams, Papierno, Makel, & Ceci, 2004). Reasoning skills are also intrinsic to the processes of knowledge acquisition and conceptual change (Bransford, Brown, & Cocking, 2000; Burbules & Linn, 1988; Kuhn, 2004).

However, a large body of research in the fields of cognitive and developmental psychology and in science education indicates that students’ reasoning in both academic and everyday life contexts is inadequate (Hogan & Maglienti, 2001; Kuhn, Amsel, & O’Loughlin, 1988; Reif & Larkin, 1991; Schauble et al., 1995; Stanovich & West, 1997; Varelas, 1996; Vass, Schiller, & Nappi, 2000; Zimmerman, Raghavan, & Sartoris, 2003). Students’ reasoning is often described as “theory-motivated” (Klaczynski, 2000) and is marked by shortcomings or biases (Klaczynski, Gordon, & Fauth, 1997; Kuhn, 1989). In addition, research indicates that...
a complex interplay of factors (e.g., cognitive, motivational, contextual) influence students’ ability to reason scientifically. Of particular interest to the present study are students’ epistemological commitments (ECs), which are beliefs about the nature of knowledge and the processes of knowing. ECs could account for gaps that exist between ways of thinking applied in everyday life and in scientific domains (Hogan & Maglienti, 2001; Klaczyński, 2000; Reif & Larkin, 1991). Such commitments are also believed to underlie the flaws and biases in reasoning about personal beliefs (Klaczyński et al., 1997) or contradictory evidence (Chinn & Brewer, 1993). However, very few research studies have examined the relationship between ECs and scientific reasoning. Such is the purpose of the current study.

**Background**

**Scientific Reasoning**

Scientific reasoning is complex in nature (Lawson, 1982; Schunn & Anderson, 1999). Lawson (1982, 2005) regarded formal or advanced reasoning largely as hypothetico-deductive in structure and consisting of a number of interrelated aspects or schemata that function independently depending on the situation or task. These aspects include control of variables, and probabilistic, proportional, and correlational reasoning skills. Waters and English (1995) maintained that scientific reasoning involves both inductive and deductive processes. Hogan and Fisherkeller (2005) defined scientific reasoning as “the practice of thinking with and about scientific knowledge” (p. 95). Overton (1990) and Holyoak and Morrison (2005) regarded reasoning as a specific type or branch of thinking that involves drawing inferences from initial premises and is closely related to judgment, decision-making, and problem-solving. Nevertheless, despite all attempts to delineate the nature of scientific reasoning and how it relates to other cognitive constructs, its essence currently remains debatable.

Recently, the work of D. Kuhn (1989, 2004) and coworkers (e.g., Klahr & Dunbar, 1988; Kuhn, Schauble, & Garcia-Mila, 1992) has been very influential in the study of reasoning. According to Kuhn (2004), scientific thinking entails more than the strategies of controlling variables and inductive causal inferences, which have been dominant in reasoning studies. Scientific thinking or reasoning is a conscious, purposeful knowledge-seeking process that is social in nature. More specifically, it is “any instance of purposeful thinking that has the objective of enhancing the seeker’s knowledge” (p. 372). It is, therefore, a process that people go through in order to revise their ideas and build new understandings. The heart of this reasoning process is the coordination of theory and evidence, which does not only mean revising the theory in light of the evidence, but differentiating between, and contemplating on, both. Successful theory–evidence coordination requires questioning existing theories, seeking contradictory evidence, and eliminating alternative explanations (Kuhn, 1989, 2004). The ability to evaluate evidence independently of prior beliefs is a basic component of effective or advanced thinking (Klaczyński et al., 1997; Stanovich & West, 1997). Consideration of alternative hypotheses has been regarded by other researchers (e.g., Lawson, 1992; Norman, 1997) as the most salient aspect of advanced reasoning. Students’ reasoning usually lacks the characteristics of successful theory–evidence coordination and is hence inadequate. Students’ reasoning is usually weakened by their inability to distinguish between their theoretical ideas and empirical evidence. Students rarely reflect on their prior conceptions and readily dismiss inconsistent information. According to Kuhn (1989), shortcomings in theory–evidence coordination arise from students’ tendency to think with their theories rather than about them. Students’ reasoning thus becomes “selective” or “self-serving” (Klaczyński & Narasimham, 1998) as they refrain from falsifying their beliefs and approach tasks with the inclination of finding verification rather than disconfirmation.

Moreover, traditional studies of scientific reasoning have dealt with either the strategies (e.g., designing experiments, formulating hypotheses, or evaluating evidence) or the knowledge employed by learners (Zimmerman, 2000). This distinction of what reasoning entails has been strongly criticized. Schauble (1996) noted that both experimentation strategies and prior knowledge (PK) affect scientific reasoning because “prior knowledge guides observations, as surely as new observations lead to changes in knowledge” (p. 103). Lawson et al. (2000) contended that both declarative knowledge and general hypothesis-testing skills are needed for successful reasoning performance. Koslowski (1996) argued that reasoning strategies about covariation cannot occur without reference to prior beliefs. Dunbar (2000) argued that isolating knowledge from processes is “highly artificial” and does not capture the real nature of scientific reasoning.
Kuhn’s (2004) coordination of theory and evidence represents an integrative framework of reasoning that incorporates PK and reasoning processes. Kuhn argued that Inhelder and Piaget’s (1958) view of scientific reasoning as a system of “logic-driven” operations undermines the role of PK and context in reasoning. In contrast, Kuhn underscores the impact of personal theories in the activity of reasoning: “Theories are integral to knowledge-seeking at every phase of the process” (p. 373). In this study, we adopted Kuhn’s (2004) integrative framework of reasoning as the intentional separation and coordination of theory and evidence, which leads to continuous revision of theories and progressive generation of new knowledge and understandings. The present study utilized an integrative approach to studying scientific reasoning by contextualizing reasoning tasks in a specific scientific domain, thus, factoring in participants’ PK.

We maintain the necessity of incorporating PK when exploring the nature of scientific reasoning. Reasoners usually employ their extant knowledge even when tackling problems from unfamiliar domains. Coordination of theory and evidence inevitably entails contemplating on and expanding one’s own ideas in order to “make sense” of the data at hand. However, in this sense-making process, reasoners vary in the degree of depth of which prior ideas are challenged or questioned, if at all. In many cases, prior ideas negatively influence or bias reasoning processes as individuals tenaciously adhere to their beliefs even after encountering contradictory evidence. We think that the extent to which prior beliefs can bias reasoning is related to, or even determined by, one’s ECs. For instance, the commitment of checking consistency between knowledge claims and empirical data, when apparent, should spark the reasoning processes of reflecting on preexisting beliefs and seeking more plausible ones, especially when encountering inconsistent evidence. Clearly, both PK and ECs need to be examined when studying the quality of scientific reasoning.

Epistemological Commitments and Reasoning

ECs or beliefs act as metacognitive tools that activate and guide reasoning processes (Kuhn, 2004; Smith & Wenk, 2006). According to some researchers, epistemological beliefs could, in some cases, outweigh PK in impacting reasoning (Greenhout et al., 2004). As with scientific reasoning, there is little agreement on the meaning or definition of the construct of “epistemological beliefs” or “personal epistemology” and whether it is domain-specific or domain-general. More specifically, there is “a lack of conceptual clarity about the elements or dimensions that constitute individual epistemological theories or beliefs” (Hofer & Pintrich, 1997, p. 111). Researchers have conceptualized personal epistemology in a variety of ways: as a cognitive developmental structure (King & Kitchener, 1994; Kuhn, 1991; Perry, 1970), as a system of independent beliefs (Schommer, 1990), as a system of interrelated theories (Hofer & Pintrich, 1997), as a system of fine-grained context-specific resources (Hammer & Elby, 2002), or as a metacognitive control process (Hofer, 2004; Kuhn, 2004). In this study, we endorsed the more dominant research perspectives on personal epistemology that regard it as a coherent system of cognitive structures (e.g., King & Kitchener, 1994). We used the terms “epistemological commitments” and “epistemological beliefs” to reflect the coherent and integrated system of personal epistemology.

Two intrinsic elements of personal epistemology emerge across the various perspectives: (a) the nature of knowledge, including aspects of certainty and simplicity of knowledge, and (b) the nature or process of knowing, including the source and justification of knowledge (Hofer, 2004; Hofer & Pintrich, 1997). Simple, fixed, or right/wrong conceptions of knowledge delivered by authority figures has been labeled as dualistic (Perry, 1970), prereflective (King & Kitchener, 1994), or absolutist (Kuhn, 1991). Transitional beliefs of knowledge reflecting doubt in certainty and the role of authority have been referred to as multists (Kuhn, 1991; Perry, 1970) or quasi-reflective (King & Kitchener, 1994). More sophisticated beliefs reflecting a relative, contextual, and uncertain conception of knowledge constructed by the knower have been described as relativist (Perry, 1970), reflective (King & Kitchener, 1994), or evaluative (Kuhn, 1991). While multists recognize the role of evidence in justifying knowledge claims, it is the evaluativists that demonstrate successful differentiation and coordination between theories and evidence: “Reflective thinkers consistently and comfortably use evidence and reason in support of their judgments” (King & Kitchener, 2004, p. 9). Reflective thinkers do not readily dismiss evidence inconsistent with preexisting theories and are thus believed to hold more sophisticated system of personal epistemology.

Moreover, students’ epistemological views of the nature and justification of knowledge influence their learning orientations (Edmondson & Novak, 1993; Hammer, 1994, 1995; Tsai, 1998). Researchers
(e.g., Dunbar, 1993) believe that ECs or beliefs impact learning styles because they control the goals and reasoning strategies that students employ. Epistemological understandings or standards are thus considered the underpinnings of scientific reasoning (Kuhn, 2004). Weinstock and Cronin (2003) reported that epistemological level (absolutist, multiplist, or evaluativist) was the main predictor of the quality of juror-reasoning in comparison to educational level, age, and gender. The researchers concluded that epistemological understandings underlie informal, everyday reasoning, and general thinking skills. Hogan and Maglienti (2001) investigated the role of epistemological standards in scientific reasoning by comparing the reasoning of scientists and nonscientists. Findings revealed that scientists consistently relied on empirical evidence to judge knowledge claims, whereas students relied more on their personal views. These results confirm that ECs help explain differences in reasoning abilities.

Yet, the relationship between scientific reasoning and ECs needs much more exploration. Considering the different terms, meanings, and empirical methods used across studies dealing with both reasoning and epistemology, it would be very difficult to make specific and solid conclusions about the relationship between the two constructs. Furthermore, only two studies, Hogan and Maglienti (2001) and Weinstock and Cronin (2003), examined the direct relationship of epistemological beliefs and reasoning, with the latter dealing with “informal” reasoning. More importantly, none of the studies encountered in the literature investigated the relationship between ECs and scientific reasoning while taking into consideration the influence of PK as an intervening factor. As such, the present study represents a contribution in this regard.

**Purpose and Research Questions**

This study aimed to explore the relationship between epistemological beliefs and the quality of scientific reasoning among college science students in the domain of hydrostatics. Specifically, the aim was to examine the impact of one EC, namely commitment to consistency of the presented theory with evidence, on the quality of students’ reasoning processes (as revealed by degrees of success in theory–evidence coordination). The guiding research questions were:

1. What range of conceptions or understandings do college science students’ hold regarding Archimedes’ Principle or buoyancy in liquids?
2. To what extent are students committed to the consistency of theory with evidence?
3. What is the quality of college science students’ reasoning as they tackle problems related to Archimedes’ Principle?
4. To what extent does commitment to consistency of theory with evidence impact the quality of participants’ reasoning, taking into account their prior knowledge?

**Method**

Using a $2 \times 2$ quasi-experimental factorial design, this study investigated the impact of the EC to consistency of theory with evidence, on the quality of undergraduate students’ scientific reasoning related to buoyancy in liquids. Participants’ reasoning served as the dependent variable. Their EC served as the independent variable. Learners’ prior conceptions of buoyancy served as the intervening variable. PK impacts learners’ thinking in a certain domain, and hence it was crucial to elucidate the nature of participants’ PK and conceptions relevant to the considered material.

**Participants**

Participants were 139 undergraduates, 50 female (36%) and 89 male (64%), enrolled in two physics courses, PHYS 102 and 212 (with a total enrollment of 1,102 students), during Fall 2007 at a large Midwestern university. Participants’ ages ranged from 18 to 23 years with an average of 19.7 years ($SD = 1.07$). Participants had an average GPA of 3.10. Participants enrolled in PHYS 102 majored in the life sciences, preprofessional health programs, agriculture, or veterinary medicine while those enrolled in PHYS 212 majored in engineering, chemistry, physics, or mathematics. PHYS 102 and 212 students have already taken the prerequisite courses PHYS 101 and 211, respectively, which addressed the physics content under investigation, that is, Archimedes’ Principle and buoyancy in liquids. Therefore, all participants have had...
formal instruction in the targeted physics concepts. In addition, the majority of participants in both courses took at least one high school (62%) and one college (73%) physics class. Table 1 presents a summary of participants’ biographical and background information.

**Content**

The chosen physics content for the present study is floatation, buoyancy in liquids, and Archimedes’ Principle. As early as elementary school, students begin to explore the everyday phenomenon of how objects behave in water. In middle school, they study the concept of density and use it to explain sinking and floating, or buoyancy in water. However, researchers (e.g., Camacho & Cazares, 1998; Hardy, Jonen, Moller, & Stern, 2006; Libarkin, Crockett, & Sadler, 2003; She, 2002) have reported that many middle and high school students hold naïve understandings of the phenomenon of buoyancy. Students attribute the sinking and floating of objects either to their size (large objects sink while small ones float), shape (e.g., a round object sinks while a boat-shaped object floats), weight (heavy objects sink and light ones float), or material (e.g., an object that is metal-like sinks). Moreover, students usually attend to one variable (e.g., weight of the object) to the exclusion of other relevant variables (such as volume of the object or density of the liquid). Even university students’ reasoning and conceptual understanding of buoyancy and Archimedes’ Principle have been described as lacking (Loverude, Kautz, & Heron, 2003). Students’ difficulties in reasoning and solving problems about buoyancy partly stem from their inadequate understanding of, and differentiation among, the underlying concepts of mass, weight, volume, density, force, and pressure (Loverude et al., 2003).

Buoyancy involves several underlying concepts. An understanding of buoyancy in liquids requires an understanding of the concepts of (a) density (relation of mass to volume), (b) liquid pressure and how it differs with depth, (c) liquid displacement, and (d) objects in equilibrium. Moreover, an understanding of the effects of buoyancy in liquids, that is, the sinking and floating of objects, requires consideration of the whole system of object and liquid. This consideration entails comparing the density of the object relative to that of the liquid and/or the weight of the object relative to the buoyant force exerted by the liquid.

**Procedure**

All 1,102 students enrolled in the two participant physics courses were invited to participate in the study. Students who indicated their willingness to participate were provided the electronic link to the online questionnaire, which was designed to assess their PK of hydrostatics and the sought EC. The questionnaire was kept available online for 6 weeks. A total of 139 students completed the questionnaire. Based on performance on the online questionnaire, a sample comprising 40 students (29%) was purposefully selected from this pool of 139 participants to sit for 1-hour individual interviews. This sample reflected variance in

<table>
<thead>
<tr>
<th>Table 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>Participant background information</td>
</tr>
<tr>
<td>Variable</td>
</tr>
<tr>
<td>Sex</td>
</tr>
<tr>
<td>Female</td>
</tr>
<tr>
<td>Male</td>
</tr>
<tr>
<td>Total</td>
</tr>
<tr>
<td>Class</td>
</tr>
<tr>
<td>Freshman</td>
</tr>
<tr>
<td>Sophomore</td>
</tr>
<tr>
<td>Junior</td>
</tr>
<tr>
<td>Senior</td>
</tr>
<tr>
<td>Physics rating</td>
</tr>
<tr>
<td>“Only Okay”</td>
</tr>
<tr>
<td>Average</td>
</tr>
<tr>
<td>Good</td>
</tr>
<tr>
<td>Very good</td>
</tr>
<tr>
<td>Excellent</td>
</tr>
</tbody>
</table>
terms of PK and EC. Four groups were identified: (a) High PK and High EC, (b) High PK and Low EC, (c) Low PK and High EC, and (d) Low PK and Low EC. Mean scores were considered as cutoff criteria for classifying high and low performance on the measures of PK and EC. Scores above the mean were considered relatively higher and scores below the mean were considered relatively lower. However, participants with the highest and lowest scores who fitted into the four cells of the $2(PK) \times 2(EC)$ matrix were selected for interviews. Ten students were interviewed from each group. Almost all students who were contacted did sit for the interview.

Instruments

Online Questionnaire. The questionnaire included two main parts: Part I presented the physics pretest, which aimed to assess participants’ PK of the relevant physics content. Part II included the epistemology scenarios, which aimed to examine participants’ EC to consistency of theory with evidence. Each part took about 20 minutes to be completed. The questionnaire was completed online in one 40-minute session. It was designed in a “forward-only” task-by-task manner, which did not allow respondents to go back to and change their initial answers. As explained below, this latter arrangement was important given the design of the epistemological scenarios. Moreover, to control for possible effects associated with the order of appearance of tasks on the questionnaire, different versions of the questionnaire with different ordering of the tasks were used.

PART I. The physics pretest included 10 short-answer questions (see Appendix A for illustrative examples of the questionnaire items). Four items were adapted from Hewitt (1998), two items from Libarkin et al. (2003), one from M. Selen (personal communication, August 2007), and the remaining items were developed by the researchers with guidance from content experts. Two physics professors and a science educator reviewed the test to ensure its face and content validity. The items were revised in accordance with the reviewers’ suggestions for improvement.

The physics pretest assessed participants’ conceptions of the phenomenon of buoyancy in liquids and the relevant parameters of density and buoyant force. More specifically, the items addressed ideas regarding: (a) factors affecting floatation (two items), (b) water displacement (three items), (c) the relationship among mass, weight, volume, and density (two items), (d) the effect of density of liquid relative to that of submerged object (one item), and (e) the relation between buoyant force and the weight of liquid displaced (two items). One item was open-ended and aimed to directly explore students’ ideas regarding sinking and floating. Another item checked respondents’ simple recall of the statement of Archimedes’ Principle and was included to provide a general picture of whether or not participants could accurately remember Archimedes’ Principle. This latter item was not scored so that not to penalize participants for failure to retain information. Two of the items specifically assessed whether respondents ascribed to naive ideas regarding the effect of size and shape of immersed objects on sinking and floating, respectively.

PART II. The epistemology scenarios included a scenarios-based instrument developed by the researchers to assess the degree of EC of a respondent to the consistency of theory with evidence. This instrument deals in particular with epistemological beliefs pertaining to the role of evidence in justifying knowledge claims. The design of the instrument is based on the assumption that ignoring contradictory evidence is indicative of a less sophisticated and more naive epistemological stance.

Furthermore, tapping into students’ ECs, which are mostly implicit and underlie their thinking, is a challenging task. While these commitments are believed to impact their reasoning, learners are not necessarily aware of these commitments or able to explicate them. Attempting to access such commitments through various forms of convergent-type questions might not be sufficient. A possibly more valid alternative would be to infer these commitments from student actions or responses to carefully designed scenarios. Such approach was adopted in the present study.

The epistemology scenarios were three: The “Coyote,” “Forest Fires,” and “Mass Extinction” scenarios. The texts for the first two scenarios were adapted from Mishra and Brewer (2003) and the third from Abd-El-Khalick and Lederman (2000a) and Abd-El-Khalick (2005). The approach used in the scenarios was developed by the researchers. In each scenario, participants read an introductory passage that presented a certain scientific theory along with supporting and convincing evidence. Participants were asked to rate their
degree of belief in the theory on a 10-point scale (with “1” referring to a “completely false theory” and “10” referring to a “perfectly correct theory”). Next, participants were provided with two sets of additional information relevant to the theory. In the case of an “experimental” scenario, the two sets of additional information presented evidence that was inconsistent with the theory (i.e., designed to detract from the validity of the theory). In the case of a “control” scenario, the additional information was relevant to the theory, but was designed to neither support nor contradict the theory. The scenarios were counterbalanced: Two versions of each of the three scenarios were produced. One of the versions was “experimental” and the other “control.” Appendix B presents an illustrative example of the scenarios showing the experimental version.

Each participant received, in different order, two experimental and one control scenario. In all three scenarios, participants were asked to re-rate their belief in the theory after reading each set of additional information. The assumption was that participants with strong/high commitment to the consistency of theory with evidence would demonstrate a relatively weaker belief in the theory in light of the additional inconsistent evidence (for the two experimental scenarios), as compared to those with weak/low commitment to consistency of theory with evidence. In other words, ratings were expected to decrease for the experimental scenarios following each set of contradictory evidence and to remain the same for the control scenarios. As noted above, the online questionnaire was designed so that after submitting their ratings, participants were not able to go back and change them. This arrangement was crucial to the logic of the scenarios-based instrument undertaken in this new approach to assessing ECs.

Individual Interviews. The interviews assessed the quality of reasoning of participants in each of the four cells in the $2^{(PK)} \times 2^{(EC)}$ matrix. In addition, interviews were used to validate questionnaire results regarding participants’ EC to consistency of theory with evidence. The subsample of 40 participants sat for individual interviews, which were conducted by the primary author in a private room. The interview was divided into two parts. The first part included physics tasks and lasted for about 50 minutes. The second part included an epistemology task and lasted for about 10 minutes.

In the First part, participants examined, in a counterbalanced order, a set of four hypothetical reasoning scenarios (see Appendix C) involving a system of water and immersed objects. Three of the four scenarios were used in a previous pilot study conducted by the researchers (Zeineddin & Abd-El-Khalick, 2008). Each scenario is described in a brief text associated with illustrations. Each scenario differs in terms of what is asked of participants and the sort of “observations” available to them in the form of illustrations. Participants were instructed to read each scenario on their own, take their time to think, write down their answer, and then inform the researcher that they are ready to discuss it. The researcher then asked participants to read, think-aloud, and explain their written justifications.

In the order presented in Appendix C, the first scenario, Floating Block of Wood, was adapted from Hewitt (1998) and edited by two physics professors. The second scenario, Five Blocks, and the third scenario, Three Suspended Cubes, were adopted from Loverude et al. (2003). The fourth scenario, Metal and Wooden Block System, was provided by J. Mestre (personal communication, August 10, 2007). These scenarios assessed the same fundamental conceptions included in the physics pretest but at a deeper level. Moreover, the structure of the in-depth interview was such that students’ lines of reasoning were revealed through the think-aloud segment in which the researcher asked probing questions. The following questions were typically asked: Why do you think that the forces are equal? On what factor or factors are you basing your judgment? What is your own understanding of buoyant force? How do you differentiate between these variables? Why do you think that the effects will cancel out? Did you think of another possibility? Can you give an example? Do you agree with the observation? Do you think it is possible that if we did this experiment then we would observe the result shown here? What do you mean by this idea? What do you predict would happen if we made this change [e.g., cutting the strings] to the system? Can you compare the buoyant forces in each figure?

After completion of the four reasoning scenarios, the researcher provided interviewees with the statement of Archimedes’ Principle. Interviewees were asked to read the Principle aloud and comment on it. Then, they were asked to use Archimedes’ Principle and reflect back on their original answers. More specifically, they were asked whether having the theory of Archimedes at hand would reinforce their initial answers or would induce them to make changes or corrections. Interviewees were asked to describe and
justify any corrections they made. This procedure was undertaken to assess the impact of awareness of Archimedes' Principle (i.e., the theory at hand) on students' reasoning. It could be argued that bringing the principle to the attention of participants at the outset of the interview might have impacted their reasoning through giving them a strong clue about what theory to use to tackle the problems at hand. It was crucial to have participants choose the sort of theoretical ideas they wanted to bring to bear on the problems. On the other hand, participants were eventually provided with the principle and given an opportunity to revise their answers in order to rule out the possibility of achieving seemingly high levels of reasoning through the straightforward application of Archimedes’ Principle, as well as assessing the extent to which participants either held to, or were willing to change, their initial reasoning.

In the second part of the interview, each participant was randomly given one of the three epistemology scenarios (experimental version) that they previously answered when they took the online questionnaire. Participants completed the epistemology task on their own. The researcher, then, asked participants to provide justifications for their ratings or answers. This procedure was undertaken as a check on the validity of the epistemological scenarios. All interviews were audiotaped and transcribed for analysis.

Data Analysis

The data were analyzed by the primary author. The second author conducted blind checks on randomly selected samples of the interview data. Analyses were performed in two stages. The first stage focused on participants' PK and EC. Analysis of responses to each of the physics pretest and epistemology tasks was conducted and compared. Four conditions pertaining to the four cells in the $2 \times 2$ matrix of PK (high vs. low) and EC (high vs. low) were identified. The second stage of analysis focused on the quality of participants' reasoning under each of the four conditions. Next, two-way analysis of variance (ANOVA) and multiple regression analyses were conducted.

For the physics pretest, a total numerical score was calculated for each participant. A rubric was used to score nine items on a 2-point scale, such that a score of “2” represented an accurate response, “1” represented a partially accurate, and “0” represented an inaccurate response. The possible scores of the physics pretest ranged from 0 to 18 points.

The epistemology scenarios were analyzed and scored based on the existence of specific patterns in participants' responses. Judgments regarding the strength of the EC of consistency of theory with evidence were based on the notion that it is hardly unlikely for a respondent to maintain the same degree of belief in a given theory when provided twice with contradictory evidence, if the respondent has a relatively strong EC of consistency of theory with evidence. Therefore, a response with a decreasing pattern or ratings of belief in theory after each set of inconsistent evidence for an experimental scenario was regarded as revealing relatively strong EC of consistency of theory with evidence. In contrast, a response that remains the same or fluctuates up and down (or vice versa) for experimental scenarios was considered as indicative of a relatively weaker EC of consistency of theory with evidence. As for the control scenario, in which the additional evidence was neither consistent nor inconsistent, only same ratings or close to same (differing only by 1 point) were regarded as revealing the expected pattern of maintaining theory belief. In such cases, the score on the control scenario contributed positively to a respondent’s total score. In all other cases, the score on the control scenario contributed negatively or decreased the overall score. Based on these criteria, each of the three epistemology scenarios (two experimental and one control) was scored and a total score was calculated for each participant. It is the total score that we believe captured the approximate strength of the sought EC.

Finally, for the reasoning scenarios, the quality of reasoning among participants in the four groups on each of the four physics scenarios was judged according to the presence and degree of accuracy of six dimensions, which were explicitly or implicitly emphasized by Greenhoot et al. (2004), Hogan and Maglienti (2001), and Tytler and Peterson (2003). These dimensions are: (a) accurate conceptualization of the task (e.g., identification of what is given—such as “observations” illustrated in a given scenario—and what is required), (b) consideration of all relevant variables, (c) accurate interpretation and/or application of relevant theoretical ideas (e.g., Archimedes’ Principle), (d) consideration of alternative or competing explanations, (e) reaching accurate or supported inferences or conclusions, and (f) depth of conceptual “processing” (e.g., sequencing an argument; connecting evidence with a conclusion; synthesizing evidence, inference, and theory).
These six dimensions are aligned with the framework on reasoning adopted in the present study (cf. Kuhn, 1989, 2004). In essence, as noted above, this framework bridges domain-specific and domain-general perspectives on reasoning by integrating the explication and assessment of a participant’s PK and reasoning processes. The rubric’s six dimensions seek to externalize participants’ reasoning activity as they engage with coordinating their PK and existing theories with the evidence provided in the scenarios. Three assumptions underlie the link between the desired mode of reasoning and the aforementioned six dimensions. The assumptions are that successful reasoning is characterized by the (a) consideration of alternative or rival hypotheses (fourth dimension) in conjunction with evoking one’s existing or preferred explanatory or predictive hypothesis or theory (third dimension), (b) use of evidence to support ideas and/or reach accurate conclusions (fifth dimension), and (c) application of appropriate processes to assess alternative hypotheses and/or draw inferences (first, second, and sixth dimensions). When considered from this perspective, it could be seen that the six dimensions are collectively indicative of the extent of a participant’s success in coordinating theory with evidence. Holistically, the dimensions display how participants used their PK and the evidence presented in each scenario to reach (hopefully, accurate) inferences or conclusions.

A rubric (see Table 2) was developed based on the above dimensions. Each of the four reasoning scenarios was scored using the rubric and then a total score was calculated. The extent of demonstration in the case of each dimension was scored on a 3-point scale: “0” for “inaccurate,” “1” for “partially accurate,” and “2” for “accurate.” Participants were scored on each scenario for the six dimensions resulting in a possible range of 0–12 points per scenario, and an overall score range of 0–48 points. The scored reasoning scenarios had an internal consistency or Cronbach alpha of 0.82. Illustrative examples and cases of how judgments on the quality of reasoning were made using the rubric are presented in the following section.

Results

Prior Knowledge and Epistemological Commitments

Participants’ scores on the physics pretest and epistemology scenarios showed considerable variability. Table 3 presents participants’ scores on the physics pretest (possible scores range from 0 to 18 points), and the EC scenarios (scores range from −3 to 12 points). Overall, participants’ performance on both the physics pretest and the epistemological scenarios was normally distributed. For the physics pretest, 68% of participants had scores between 8.85 and 14.59 points (1 SD off a mean of 11.72), which is a relatively “good” result. In other words, participants’ prior conceptions regarding buoyancy in liquids could be described as moderately accurate. For the scenarios, 68% of scores lied within −0.28 and 5.78 points (1 SD off a mean of 2.75). Since the scenarios were developed for the purposes of this study, interpretation of scores was based on their relative distance about the mean. The higher a score above the mean, the higher the EC of consistency of theory with evidence was, and vice versa.

Independent samples t-tests were used to compare the means for the above two measures between participants enrolled in the two physics courses, PHYS 102 and 212. Table 4 presents the results of the comparisons between participants’ physics pretest and epistemology assessment scores. These results showed no statistically significant differences in participants’ PK and epistemology measures across the two courses. Thus, participants in the two courses were equivalent in their understandings of buoyancy and their ECs.

Table 5 presents the means and standard deviations in terms of physics PK and commitment to the consistency of theory with evidence for the subsample of 40 participants who were selected to sit for individual interviews. A comparison of the mean scores presented in Tables 3 and 5 indicates that on average, the performance of each group in the 2(PK) × 2(EC) matrix was substantially higher (or lower) on each of the two measures of physics pretest and EC scenarios than the average performance of the 139 participants. For instance, participants belonging to the High PK–High EC and High PK–Low EC groups had demonstrated higher PK (mean physics pretest scores of 15.35 and 13.95 respectively, see Table 5) relative to the average performance of all 139 participants on the physics pretest (mean score of 11.72, see Table 3). These results elucidate the variability and level of each group in the matrix relative to the two measures of PK and EC.
Table 2

Rubric used for scoring participants’ quality of reasoning

<table>
<thead>
<tr>
<th>Dimension</th>
<th>Criteria and Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Accurate conceptualization of the task</td>
<td>Shows partially accurate conceptualization of the task; either misinterprets the “given” or what is required</td>
</tr>
<tr>
<td>Consideration of all relevant variables</td>
<td>Considers some but not all of the relevant variables and/or does not accurately distinguish and connect the variables</td>
</tr>
<tr>
<td>Accurate interpretation and/or application of relevant theoretical ideas</td>
<td>Attempts to link or apply a relevant theory (such as but not limited to Archimedes’) but shows inaccurate conceptualization of theory, or fails to use it to build a valid argument or justification</td>
</tr>
<tr>
<td>Consideration of alternative or competing explanations/ideas</td>
<td>Mentions, in passing, one alternative or competing explanation but does not pursue the implications of this alternative for tackling the task</td>
</tr>
<tr>
<td>Depth of conceptual “processing”</td>
<td>Provides an explanation(s) that reveals partial conceptual processing: Explanations or justifications are partially valid; there are still gaps in the way ideas and variables are linked together to provide a supported and valid justification</td>
</tr>
<tr>
<td>Reaching accurate or supported conclusions</td>
<td>Reaches partially accurate or partially justified conclusion</td>
</tr>
</tbody>
</table>

Quality of Scientific Reasoning

Table 6 summarizes participants’ performance on the reasoning scenarios. In general, scores were relatively higher on the Floating Wood and Five Blocks scenarios. The Three Cubical Blocks and the Metal

Table 3

Participants’ scores on the online questionnaire (N = 139)

<table>
<thead>
<tr>
<th></th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Physics pretest</td>
<td>11.72</td>
<td>2.87</td>
<td>3.5</td>
<td>17.5</td>
</tr>
<tr>
<td>EC scenarios</td>
<td>2.75</td>
<td>3.03</td>
<td>-3</td>
<td>12</td>
</tr>
</tbody>
</table>
Illustrating the Use of Dimensions to Judge Quality of Reasoning

**Dimension I: Accurate Conceptualization of the Task.** Almost all participants showed accurate conceptualization of the reasoning scenarios. Two participants showed partially accurate conceptualization of the Three Suspended Cubes scenario. These latter two did not correctly interpret the “given” that the blocks are suspended from strings of varying lengths. They, thus, thought that the positions of the blocks in the container were not related to the varying lengths of the strings. One participant showed inaccurate interpretation of the Metal and Wooden Blocks scenario. He thought that the blocks when together and separate were not identical and, hence, believed that the density of each block could change.

**Dimension II: Consideration of All Relevant Variables.** Participants did not always consider the relevant variables in a given scenario. More importantly, in some cases, participants did not accurately differentiate or connect between variables. For example, for the Floating Wood scenario, one participant, A3, thought that the force of gravity was different from the weight of the block. When asked whether other forces act on the block in addition to the force of gravity and the buoyant force, A3 answered:

> There could be the force of just the block, the weight of the block . . . I mean, mg. Yeah, weight of the block. But I don’t know which way that works. It might work upwards. But other than that, I don’t think there’s any other forces besides those three.

Moreover, another participant, J20, gave the following answer regarding the addition of sugar to water for the Floating Wood scenario: “Sugar has weight and so I said that if you add sugar to water it would increase the weight of the water, maybe the force of the water will increase and so the block will float higher up.” In her justification, J20 did not accurately link the weight of the solution to the force acting upward on the floating block. In other words, she did not explain how the “weight” of the water relates or affects the force exerted by the water on the block.

In contrast, several participants considered all relevant variables and revealed accurate differentiation among variables. For example, A5 gave the following justification for his answer on the Five Blocks scenario:

<table>
<thead>
<tr>
<th>Group (n = 10)</th>
<th>Physics Pretest M</th>
<th>SD</th>
<th>EC Scenarios M</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>High PK–High EC</td>
<td>15.35</td>
<td>0.75</td>
<td>6.5</td>
<td>1.35</td>
</tr>
<tr>
<td>High PK–Low EC</td>
<td>13.95</td>
<td>1.07</td>
<td>0.1</td>
<td>0.99</td>
</tr>
<tr>
<td>Low PK–High EC</td>
<td>10.05</td>
<td>1.26</td>
<td>4.0</td>
<td>1.15</td>
</tr>
<tr>
<td>Low PK–Low EC</td>
<td>7.45</td>
<td>2.19</td>
<td>−1.2</td>
<td>1.55</td>
</tr>
</tbody>
</table>
Since block 1 was less than the mass of block 2 it had to be higher up in the water because the force of the mass of the block pushing down was less than the force of the water pushing up, the buoyant force.

A5 considered and compared the downward and upward forces acting on a given block. He related the magnitude of the downward force (or the weight) to the mass of the block. He decided on the position of the block based on comparing the masses of blocks 1 and 2 and the buoyant forces acting upon each. He concluded that the upward buoyant force was higher on block 1 in comparison to its mass before it reached equilibrium. Other participants, such as A6, expressed their reasoning for this scenario only in terms of the mass of the blocks:

Since the mass of each block is different, with 5 greater than the mass of the lowest, I feel that the lower it will sink depends proportionally to the mass. The higher the mass, the lower it will sink, so since 5 has the most mass, it is at the bottom, so that means 1, 2, 3, 4 align diagonally.

Focusing only on the masses of the blocks in relation to each other, rather than in relation to the density of water, was regarded as a partially accurate consideration of variables in the Five Blocks scenario. A more accurate response would consider both the masses and volumes, or the densities of each block relative to that of water and/or the weights and buoyant forces acting on the blocks.

Dimension III: Accurate Interpretation and/or Application of Relevant Theoretical Ideas. Participants applied a range of theoretical ideas across scenarios including Archimedes’ Principle. While these theoretical ideas did seem plausible in some cases, the way they were applied was not always accurate or justified. For example, D13 revealed the following “theory” in his answer for the Five Blocks scenario:

If the block weighed more than the water that’s displaced, it would keep on sinking. And if it weighed less than the water displaced, it would keep floating up, so there wouldn’t be like a direct relation, it would more so—whether or not it would be greater or less than the water that was displaced, so I had the two smallest masses [blocks 1 and 2] floating all the way up to the top, and being at the same location—or the same height.

D13 placed blocks 1 and 2 at the same floating level in water. While comparing the weight of a block to the weight of the water it displaces is plausible, D13 neglected to link other factors to the amount of water displaced. Block 1 has a smaller density and, therefore, would displace less water and float higher.

For the Three Cubical Blocks scenario, D12 had an interesting “theory” but also did not use it successfully to build a valid justification. When asked to rank the buoyant forces acting on the three cubes, D12 responded:

I think they are all equal, because I think water can only produce so much buoyant force, and for the force of gravity to be greater than that, is what causes an object to sink. And since all of the objects are freely suspended, they clearly would sink to the bottom, so they are all greater than the buoyant force that water is able to produce, and water cannot produce any more than that, so I think that they all must be equal . . . So, in this case, them all being suspended, and not—none of them floating to the top, just the strings being a different length, I think they would all sink to the bottom if the strings were not there.
While the conclusion is correct (buoyant forces are equal), the “theory” and the justification lack clarity and substance. It is not clear what D12 meant by the idea that water “is able to produce so much buoyant force.” And, when asked to elaborate, his answer remained vague. Perhaps he believed that if an object were to sink in water, then water would exert a maximum buoyant force on it which would, as a result, be weaker than its weight. His reasoning was that because none of the cubes could float, then they were bound to sink, and the strings were the only factor preventing their sinking. Hence, the water would be exerting all the maximum upward force it “could” apply on the cubical blocks. He regarded buoyant force as a property of the liquid independent of the submerged object. As such, his “theory” is distorted and not completely valid.

Moreover, few participants did link one or more scenarios to Archimedes’ Principle, but revealed inaccurate conceptualizations and applications of the principle. This is evident in the following response for the Three Cubical Blocks by J19:

Block A would have the largest buoyant force, block B would have the second most buoyant force and block C would have the least buoyant force based upon the fact that I believe it’s Archimedes' Principle that states the mass of the object in the water is equal to the volume of the water displaced, so block C has less mass, it would most likely to displace less water, so it would have a less buoyant force coming from underneath it.

Some reasoners demonstrated more accurate application and interpretation of Archimedes’ Principle. For instance, A2 gave the following response for the Floating Wood scenario:

Since there was gravity acting on it, and the block isn’t sinking, there must be some force acting in the opposite direction which is equal to it… The buoyant force would be because of the water. Because like the amount of mass of water that is displaced by the wood would result in the buoyant force, because then water would push in all directions at the block, every point, and that would be pointing up. The net resultant force would be pointing upwards.

A2 explicitly mentioned Archimedes’ Principle and revealed an accurate understanding of it. More importantly, he used it successfully in all scenarios to build valid inferences. Other participants, like A5, did not mention Archimedes’ Principle but expressed their answers in terms of the amount of water displaced to build valid justifications:

[Five Blocks] Initially I put that block 3 was just below block 2 and that block 4 was just below block 3 but still above block 5 but then I thought that instead of being suspended in water they would just sink all the way to the bottom because they weren’t displacing any more water.

Dimension IV: Consideration of Alternative Explanations. This dimension reflects the disposition to think of alternative hypothesis, which is regarded by some researchers as a strong indication of “high” reasoning (e.g., Kuhn, 2004; Stanovich & West, 1997). Participants revealed consideration of alternative explanations mostly when working on the Five Blocks scenario. This is because the nature of this scenario is such that the outcome is not readily apparent and more than one result is possible (block 3 could either sink or stay in the middle depending on its density relative to water). It is important to note, however, that consideration of competing ideas was distinguished from cases in which participants were totally confused between two options or answers. To demonstrate accurate consideration of alternative ideas, participants had to consider at least two competing explanations, weigh the pros and cons associated with each and compare the implications of each on the task. They had to provide a valid justification in case they favored one alternative against the other. For example, J20 revealed partial consideration of alternative explanations for the Five Blocks scenario:

First I thought that the heavier the blocks get the more they would sink down, but then I thought about it being in a swimming pool a person of different weight would either float or sink; nobody can actually stay in the middle of the water, so that’s why I changed my answer and that’s why 3 and 4 would go down.

Journal of Research in Science Teaching
J20 considered the possibility that the blocks would be at increasing depths depending on their increasing masses against the possibility that the heavier blocks would sink. She used an example from everyday life to decide between the two alternatives. However, she did not fully explain why she thinks that objects cannot be suspended underneath the surface of water and, hence, her justification remained deficient. Similarly, K20 mentioned an alternative explanation for the Three Cubical Blocks scenario but did not provide a complete justification:

So they [blocks A & B] would both have to have the same buoyant force but C is less massive which would create a smaller buoyant force—then, halfway through, my mind started like picture things like when I’m in the pool or something and it always seems like it’s harder to hold things down with my feet than hold it at the surface with my hands. So I started thinking that possibly the buoyant force has something to do with depth, but I’m not really sure about that, and if that were the case then B would be the greatest buoyant force and then A and C would have equal ones.

It should be noted that participants received a full score for having considered and justified two or more alternative explanations irrespective of the scientific accuracy of these explanations and/or justifications.

**Dimension V: Depth of Conceptual Processing.** Participants provided explanations and justifications that revealed differences in depth of conceptual processing. In some cases, the explanations either did not make any sense or were mere repetition of the observations mentioned in the scenario. For example, J20 justified her answer for the Metal and Wooden Blocks scenario by simply describing or listing the objects in the task: “Since nothing has changed, in the first beaker there were two blocks and in the second beaker there are still 2 blocks, the water level will stay the same.” Moreover, another participant, A3, when asked to predict what would happen to the Three Cubical Blocks if the strings were cut, gave a response that made no sense: “Can any float? [Pause] if it has to float, the one that should float is C. That’s what I think. C should float. Maybe they can all float. That’s what I think.”

In other cases, participants provided explanations that revealed partial conceptual processing in that there were gaps in the way ideas were linked together. This is shown in the following response to the Three Cubical Blocks by A4:

I felt like if B has a greater mass then it would sink further down, but C is less of mass than B, so I think it would be like, floating higher than B, but A like the same thing, C it got like, I am thinking buoyancy force helped it not to sink or something like that. I thought A and C were the same.

A4 believed that the buoyant force acting on block C “helped it not to sink” and, thus, is greater than that acting on block B which is at a lower level in water. She based her inference regarding the strength of the buoyant forces on the depth of the blocks. However, she also was trying to consider the effects of the masses of the blocks but was not successful in clearly linking the factors of mass and depth together to build a valid argument.

**Dimension VI: Reaching Accurate Conclusions.** Inferences or conclusions arrived at were not always accurate. For instance, many participants ranked the buoyant forces in the Three Cubical Blocks scenario based on the masses and/or positions of the blocks and, hence, did not provide correct conclusions. “The object that is more massive will have a greater buoyant force” (C9). Furthermore, for the Metal and Wooden Blocks scenario, most conclusions regarding the level of water were inaccurate:

I am pretty sure they would be equal [water level] . . . because nothing is changing, other than the position of the blocks, so while the blocks are in different positions and the wooden block is now floating higher, the densities, the masses, and the volumes balance out. (D12)

Another inaccurate response for this scenario was given by T39:

I think it would stay the same because the water is being used like to measure the volume of the two blocks, and I don’t think it matters if the blocks are seen as one body or two bodies. [Asked to clarify
her idea] If the volumes of the blocks don’t change, it doesn’t matter if they’re like connected or separate.

It should be noted that the above dimensions are not mutually exclusive. Judgments, in some cases, were challenging, especially in terms of choosing and applying a criterion or a set criteria. Additionally, judgments were “relative” in nature as compared to being absolute. That is, in some cases, the range of participants’ responses dictated the assignment of scores, such as, “partially accurate” as compared to “accurate.” However, the variance in participants’ responses lent validity to the dimensions under consideration and the criteria in use. Additionally, it should be emphasized that reaching “accurate” or canonical responses was only one of the dimensions under consideration. In other words, many participants achieved high scores in terms of the quality of their reasoning even though they did not get the right answer.

Illustrative Case Studies of Participants’ Reasoning

This section presents the cases of three participants who demonstrated different levels of reasoning and further elucidates participants’ quality of reasoning.

The Case of Randy. This participant performed “poorly” on all reasoning scenarios. Overall, his answers revealed confusion among variables and weak conceptual processing. There was no indication of an accurate interpretation and application of a relevant theory. There was also no indication of accurate consideration and systematic differentiation between alternative explanations. For the Floating Wood scenario, Randy (a pseudonym) listed several variables without relating and distinguishing between them:

Randy: Well, the forces, I thought of, gravity acting on the wood pushing downward, but then you have to look at the buoyancy and density of the wood, and its surface area, the density of the water compared to the wood which would equal how much it would float for how high, that pushes it upward allowing it to float not sink.
Researcher: Okay and how do these forces compare?
Randy: Well, gravity is constant, and it would balance out to the point where the buoyancy and surface area and density of the wood would allow it to float at whatever level they would find equilibrium of some sort.
Researcher: Can you repeat?
Randy: The downward force of gravity against the force of displacement of water and the surface area and density of the wood and the water allowing the float would create equilibrium and would determine the level of which the wood would float, how high it would float.

Randy could infer that the block is in equilibrium because it is floating, but he conflated a number of variables and did not differentiate between “force,” “density,” and “surface area.” He also did not differentiate between force of gravity and gravitational acceleration, which he referred to as constant. Moreover, for the Three Cubical Blocks scenario, he based his ranking of the buoyant force on the difference in mass of the blocks but failed to provide a logical and valid justification:

Researcher: How did you rank the buoyant force acting on these blocks and what is your explanation and thoughts here?
Randy: Well, since all the blocks had equal volume and blocks A and B had equal mass and block C had less mass than A and B, I ranked the amount of buoyant force increasing on block C because of its less mass. And with everything being equal except for the mass of the three blocks—that’s why I ranked block C is having the most force pushing it up more and block A and B to be equal, the reason for why the positions are different that I am not sure of but I said block A and B should have equal buoyancy.

This segment also shows that Randy did not encode the scenario accurately because he could not account for the different positions of the cubical blocks. He misinterpreted or did not attend to the scenario’s text that the strings which hold the cubes have different lengths.
Moreover, Randy’s explanations did not reveal logical flow of ideas. There were “gaps” and contradictory statements in his responses:

(Five Blocks scenario)
Researcher: How do you think the densities compare?
Randy: I did not factor into how the densities are but knowing that they have the same size and shape and the variable is their masses, I would assume that 5 is much more dense than 1 causing it to sink more, but again then depending on their densities it could be different, but assuming that they are all equally dense compared to their mass, I think that’s how they would float, that’s all.

It is worth mentioning that Randy had performed weakly on the online questionnaire in both PK and epistemological beliefs.

The Case of Roger. This participant revealed a relatively “better” quality of reasoning compared to Randy. He provided justifications that revealed partial accuracy in linking ideas and variables together. Roger consistently thought of competing explanations and provided examples from everyday life experiences. However, he was not successful in providing complete justifications and did not always arrive at accurate conclusions:

(Five Blocks scenario)
Researcher: Okay, so you’re relying on the increasing masses here. And actually you have written densities?
Roger: Yeah.
Researcher: Okay, did you think of any other alternatives? Did you have any other ideas?
Roger: Well, I wasn’t sure if 3 or 4 was going to just sink. I don’t know if objects can be suspended in the water, but I think it can be. So, at first I was thinking that maybe 1 and 2 were kind of above the water, while 3, 4, and 5 were just right at the bottom of the water. But I’ve seen in movies, I’ve seen things to be suspended in water, so…
Researcher: Oh, really?
Roger: Yeah. I don’t know how accurate that is, but I think this is a more scientific drawing [diagonal line].

While Roger did not use Archimedes’ Principle, he applied his own theoretical ideas in a valid and logical manner. For the Three Cubical Blocks scenario, he was especially attentive to the strings holding the blocks and he inquired whether they were straight as shown in the figure or could be “crooked.” He based his inferences on observations of the strings and revealed deep conceptual processing as shown in the following excerpt:

Roger: Well, first, based on the information given, we know that A and B have the same density while C has the lower density. But I think because these strings are suspended and the strings are not being pulled up, and the fact that the strings are all straight, tells me that the force of the cube acting downward is all greater than the force acting upwards. So, I would have to say that all the buoyant forces are equal.
Researcher: And that’s because of the strings?
Roger: Yes. If the string was not as straight, not straight, I would have to say that there is a greater buoyant force. But seeing that all the strings are straight, I guess it’s kind of like suspending an object in air, you see that the string is going to be straight, so—and that tells us that the force going down is obviously greater than whatever force that’s pushing the object up. And so, all the buoyant forces are equal.
Researcher: But the masses for A and B are more than C. So, doesn’t this affect the buoyant force in your opinion?
Roger: I think that’s a possibility, but to be honest, I’m not sure what the mass does to the buoyant force. If I had equations this would be much easier to explain.
Researcher: So, you think that if the buoyant force was strong enough, it would have kind of tilted the string?
Roger: Yes, tilted the string.
Researcher: And since the string is not tilted, then this buoyant force is weak, and—but, I mean, couldn’t it be weaker on one than the other? That’s my question.
Roger: I see. Obviously for A and B, the buoyant forces are the same because they have the same density and the same mass. For C, it is possible, I think it is, that the buoyant force is actually less because the weight that the downward force of the block C is actually going to be less than A or B, because it has a smaller mass. Okay, can I change my mind?
Researcher: Of course.
Roger: Okay, so C does have a less buoyant force because if you, let’s say, have two metal pieces of different mass that’s floating—or, no, let’s take away those strings here, A would be at a lower place in the water than C. So, I would have to say [pause]—no, I think the buoyant forces are all the same. Because if there were no strings and A sank more, that just means that A is heavier. That doesn’t change the buoyant force.

Roger was classified according to the results of the online questionnaire as having Low PK and High EC.

The Case of John. This participant revealed “high” reasoning across all scenarios. The excerpt below shows accuracy in linking variables, applying theoretical ideas (water displacement and buoyant force), and building valid justifications. The excerpt also demonstrates rich flow of ideas and skillful coordination between observations and inferences:

(Five Blocks scenario)
John: I said that block 1 would float higher in the water than block 2, and blocks 3, 4, and 5 would all be sunk to the bottom, because block 2 seems like it’s almost entirely submerged in water, meaning that—if the density of the fluid times the volume displaced equals the mass of the entire block, then the force of gravity would be matched by the buoyant force and the block would not move, because block 1 does not need to displace as much fluid as block 2, it is allowed to float higher in the water. Block 2 requires all of its volume to be in the water to displace as much water as it can in order for its buoyant force to just barely match the force of gravity. Anything that’s more massive than block 2 could not displace enough fluid to keep it to float and because its sinking, it’ll all sink to the bottom, so blocks 3, 4 and 5 will sink all the way to the bottom.
Researcher: Do you see any other possible result?
John: Well, there’s a hair of block 2 that’s above the water so if block 3 was minuscule more dense than block 2 then it could float an inch deeper in the water, but I went under the assumption that the top of block 2 was at the top of the water.
Researcher: Okay
John: But I think the question might be implying that someone might think or draw block 1 highest go down dangling to block 5 but that’s impossible because anything that sinks at all will sink at the bottom unless it’s neutrally buoyant.
Researcher: What do you mean by neutrally buoyant?
John: I think block 2 is neutrally buoyant because its mass is equal to the mass of water displaced so you can put it anywhere in the tank in the water and it would stay put, it wouldn’t sink or rise.

John belonged to the High PK and low epistemology group.

Reactions to Archimedes’ Principle

Overall, 10 of the 40 interviewees used Archimedes’ Principle to tackle the physics tasks during the interviews. However, providing the 30 who did not with Archimedes’ Principle at the conclusion of the interview did not result in substantial and favorable changes in their reasoning. It could thus be argued that successful reasoning about the physics tasks used in the present study demanded more than knowledge and straightforward application of the principle. Participants had to coordinate their theoretical ideas (whether these were derived from Archimedes’ Principle or not) with the “evidence” provided in the tasks and reason about the relevant ideas and variables in order to come up with a justifiable conclusion.

Journal of Research in Science Teaching
Interview Epistemology Scenario

Regarding the experimental epistemology scenario that participants re-completed in the interviews, responses showed that the overwhelming majority of participants (88%) chose the same, or nearly the same degree of EC as they did when they responded to the online questionnaire (either high or low commitment). Only five participants (12%) ended-up with opposite levels in terms of EC. In general, participants who demonstrated strong commitment noted that their ratings decreased because the additional information had “shaken” their belief in the theory. These participants could identify the contradiction presented and react to it by successively decreasing their ratings. Some participants, while acknowledging the contradiction, did not change their ratings. They either could not provide valid and clear justifications or thought that more evidence is needed and one or two counter examples would not “hurt” the theory. Others believed that in any theory, there are hidden factors that are yet to be “discovered” and conflicting examples could be accounted for by those unknown factors. Moreover, some participants simply misinterpreted the presented evidence or “twisted” it to fit the theory.

These results lend credence to the validity of the epistemology scenarios in assessing respondents’ commitment to the consistency of theory with evidence. To start with, the scenarios generated the same response from the overwhelming majority (88%) of participants even though their responses to the scenarios on the online questionnaire and during the interview were separated by, at least, a 2-month time period. Additionally, interview responses showed that participants’ reasoning for choosing to decrease their “belief” in the theory or not was consistent with the assumptions underlying the development of the scenarios. The above results increased confidence in the use of these scenarios to categorize participants as having “high” or “low” commitments to the consistency of theory with evidence.

Prior Knowledge, Epistemological Commitments, and Reasoning

The four groups in the 2(PK) × 2(EC) matrix had substantially different mean scores for quality of reasoning. Table 7 presents descriptive statistics for the reasoning performance of each group. Figure 1 shows histogram graphs of the frequency distribution of reasoning scores with the normal curve fit for each group. The High PK–High EC group is skewed to the right while the other groups show approximately normal distributions about their mean scores.

The mean reasoning scores of the four groups and the main effects of each factor are summarized in Table 8. It could be seen that, given the same level of PK, participants with higher EC demonstrated better reasoning than those with lower EC. On the other hand, given the same level of EC, participants with higher PK demonstrated better reasoning than those with weaker understandings of the addressed content. Moreover, there is no interaction in the effects of the two factors (PK and EC) on reasoning quality. The difference between mean reasoning scores for those with High EC and High PK (1.85) is equal to the difference between mean reasoning scores for those with Low EC and Low PK. Likewise, the difference between mean reasoning for those with High PK and Low EC (6.05) is the same as that for Low PK and High EC. Lack of interaction is also shown by the parallel lines in Figure 2. It could, therefore, be concluded that the higher the EC, the higher the quality of reasoning after controlling for PK.

A two-way ANOVA was used to compare the means of the four independent groups and to compare the means across categories of one factor while controlling for another. The ANOVA results are presented in Table 9. Results indicated that the means of the four groups (the overall model) were significantly different ($F = 6.46, p < 0.001$). It should be noted that the $F$-test statistic is robust to violations of assumptions

<table>
<thead>
<tr>
<th>Group</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low PK–Low EC</td>
<td>10</td>
<td>19.1</td>
<td>3.47</td>
<td>12</td>
<td>24</td>
</tr>
<tr>
<td>Low PK–High EC</td>
<td>10</td>
<td>23.6</td>
<td>6.88</td>
<td>14</td>
<td>36</td>
</tr>
<tr>
<td>High PK–Low EC</td>
<td>10</td>
<td>27.3</td>
<td>7.95</td>
<td>17</td>
<td>45</td>
</tr>
<tr>
<td>High PK–High EC</td>
<td>10</td>
<td>31.2</td>
<td>6.52</td>
<td>25</td>
<td>44</td>
</tr>
<tr>
<td>Total</td>
<td>40</td>
<td>25.3</td>
<td>7.66</td>
<td>12</td>
<td>45</td>
</tr>
</tbody>
</table>

Journal of Research in Science Teaching
regarding normality and equality of population variances for balanced or equally sized samples, as is the case in the present study (Agresti & Finlay, 1999). In addition, Levene’s test of homogeneity of variances gave a significance value of 0.221, which strengthens the assumption that the error variance is equal across groups.

Table 9 indicates that the between-groups estimate of the population variance is 6.5 times the within-groups estimate of variance at a probability value of 0.001. In other words, the variability between groups is significantly more than the variability within groups. This provided very strong evidence against the null hypothesis that the population means of the four groups are equal: A difference exists among the true mean reasoning values of the four groups.

Moreover, results of the ANOVA test substantiated the absence of interaction between the two predictors, PK and EC ($F = 0.02, p < 0.884$). And, controlling for EC, there was a very strong evidence that a difference exits in mean reasoning between the two levels (high and low) of PK ($F = 15.09, p < 0.001$).

Table 8
Reasoning mean scores and main effects across the four groups

<table>
<thead>
<tr>
<th></th>
<th>Prior Knowledge</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>High</td>
<td>Low</td>
<td>Mean</td>
</tr>
<tr>
<td>Epistemological commitment</td>
<td>31.20</td>
<td>23.60</td>
<td>27.40</td>
</tr>
<tr>
<td>High</td>
<td>27.30</td>
<td>19.10</td>
<td>23.20</td>
</tr>
<tr>
<td>Mean</td>
<td>29.25</td>
<td>21.35</td>
<td></td>
</tr>
</tbody>
</table>
Also, controlling for PK, the mean reasoning scores between low and high levels of ECs were significant ($F = 4.27, p < 0.046$).

Next, multiple pairwise comparisons using the conservative Tukey test were used to identify significant differences between the means of the four groups. The results of the multiple comparisons appear in Table 10. Statistically significant differences were found between the (a) Low PK–Low EC and High PK–High EC groups, and (b) Low PK–Low EC and High PK–Low EC groups. Table 10 indicates that the 95% confidence interval for the population mean difference $\mu_{H-L} - \mu_{L-L}$ is $(4.35, 19.85)$ and for $\mu_{H-H} - \mu_{L-L}$ is $(0.45, 15.95)$. Therefore, the population mean value of reasoning was between 4 and around 20 units higher for the High PK–High EC than for the Low PK–Low EC group. Also, the population mean was between 0.4 and 16 units higher for the High PK–Low EC than the Low PK–Low EC group.

Finally, a multiple regression analysis was conducted. The categories or levels of each predictor were given artificial quantitative values (low = 1; high = 2). Table 11 shows the parameter estimates in the regression model having no interaction. The relationship between quality of reasoning and both predictors of PK and ECs considered together is given by the multiple regression model: $E(Y) = 7.2 + 7.9 \text{PK} + 4.2 \text{EC}$. Both PK and EC have a positive effect on reasoning. The coefficient of PK ($\beta_1 = 7.9$) is the estimated difference in mean reasoning scores between individuals with High and Low PK, having the same level of EC. And, the coefficient of EC ($\beta_2 = 4.2$) is the estimated difference in mean reasoning between High and Low ECs for each level of PK. Moreover, the tests of the main effect of each predictor indicated that this difference is statistically significant. Overall, PK is a stronger predictor of reasoning than ECs ($\beta_1 > \beta_2$).

![Figure 2. Reasoning scores showing no interaction.](image)

Table 9

<table>
<thead>
<tr>
<th>Source</th>
<th>Sum of Squares</th>
<th>df</th>
<th>Mean Square</th>
<th>$F$</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Corrected model</td>
<td>801.40*</td>
<td>3</td>
<td>267.13</td>
<td>6.46</td>
<td>0.001</td>
</tr>
<tr>
<td>PK</td>
<td>624.10</td>
<td>1</td>
<td>624.10</td>
<td>15.09</td>
<td>0.000</td>
</tr>
<tr>
<td>EC</td>
<td>176.40</td>
<td>1</td>
<td>176.40</td>
<td>4.27</td>
<td>0.046</td>
</tr>
<tr>
<td>PK $\times$ EC</td>
<td>0.90</td>
<td>1</td>
<td>0.90</td>
<td>0.022</td>
<td>0.884</td>
</tr>
<tr>
<td>Error</td>
<td>1489.00</td>
<td>36</td>
<td>41.36</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>27894.00</td>
<td>40</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*$R^2 = 0.35$ (adjusted $R^2 = 0.29$).
Discussion and Conclusions

Evidence furnished by prior research studies, which was mostly correlational in nature, has been indicative of a relationship between learners’ ECs or PK and their reasoning. As far as PK is concerned, Sadler and Zeidler (2005) reported a positive influence of content knowledge on informal reasoning and Schauble (1996) highlighted the relationship between learners’ preconceptions and experimentation strategies, while Greenhoot et al. (2004) concluded that prior beliefs negatively affect or bias reasoning especially when understandings of methodological concepts (including the function of evidence) are weak. On the other hand, only two studies examined the direct relationship between epistemological beliefs and reasoning: Weinstock and Cronin (2003) found that epistemological understandings are positively related to levels of informal reasoning, and Hogan and Maglienti (2001) reported strong epistemological standards for scientists or high reasoners as compared to students. Based on such findings, researchers inferred that both PK and epistemological beliefs impact the quality of reasoning among learners. However, beyond a generalized qualitative sense, little is known about the extent to which, and the nature of the ways in which, PK and epistemological beliefs impact and/or interact in influencing the quality of reasoning. To start with, little is known about the nature of ECs and their impact on reasoning processes. What is more, to the best of our knowledge, no studies have examined the relationship between ECs and scientific reasoning while accounting for the role of PK as an intervening factor. To date, the possible conflating effects or differential impacts of PK in the context of investigating the role of epistemological beliefs in reasoning have, in a sense, mostly been sidestepped through the use of content-free or content-lean tasks in which the role of learners’ PK was minimized or totally ignored (e.g., Stanovich & West, 1997).

From this perspective, by coupling an integrative perspective on reasoning (i.e., one that incorporated both domain-specific and domain-general elements) with some methodological improvements over prior investigations, the present study adds to our understanding of the nature of the sought relationship. First, far from minimizing or sidestepping the potentially important role of PK, we adopted a framework for reasoning as the coordination of theory and evidence (Kuhn, 2004) in which reasoners consciously employ their PK to examine and evaluate content-rich situations. In the present study, reasoning tasks were contextualized in the specific scientific domain of hydrostatics, which provided a means to account for the role of participants’ PK in reasoning in addition to their ECs. To achieve this latter goal, nonetheless, we deployed three

Table 10

<table>
<thead>
<tr>
<th>Group (I)</th>
<th>Group (J)</th>
<th>Mean Difference (I – J)</th>
<th>Sig.</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>H-PK/H-EC</td>
<td>L-PK/L-EC</td>
<td>12.10**</td>
<td>0.001</td>
<td>4.35 19.85</td>
</tr>
<tr>
<td></td>
<td>L-PK/H-EC</td>
<td>7.60</td>
<td>0.056</td>
<td>–0.15 15.35</td>
</tr>
<tr>
<td></td>
<td>H-PK/L-EC</td>
<td>3.90</td>
<td>0.534</td>
<td>–3.85 11.65</td>
</tr>
<tr>
<td>H-PK/L-EC</td>
<td>L-PK/L-EC</td>
<td>8.20*</td>
<td>0.035</td>
<td>0.45 15.95</td>
</tr>
<tr>
<td></td>
<td>L-PK/H-EC</td>
<td>3.70</td>
<td>0.577</td>
<td>–4.05 11.45</td>
</tr>
</tbody>
</table>

*The mean difference is significant at the 0.05 level.
**The mean difference is significant at the 0.01 level.

Table 11

<table>
<thead>
<tr>
<th>Parameter</th>
<th>B</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>7.2</td>
<td>0.111</td>
</tr>
<tr>
<td>PK</td>
<td>7.9**</td>
<td>0.000</td>
</tr>
<tr>
<td>EC</td>
<td>4.2*</td>
<td>0.043</td>
</tr>
</tbody>
</table>

*Coefficient is significant at the 0.05 level.
**Coefficient is significant at the 0.01 level.
methodological improvements. First, unlike prior investigations, which were mostly correlational in nature, this study utilized a quasi-experimental factorial design. Second, to the best of our knowledge, this is the first study in which both participants’ PK and ECs were independently assessed using probes that differed from those aimed at assessing their reasoning. In other words, different instruments were used to independently assess the three variables of interest, namely, PK, ECs, and quality of reasoning. Third, unlike the case of most (if not all) prior studies, we provided, and illustrated the use, of an explicit scoring rubric to judge the quality of participants’ reasoning, thus, demystifying such judgments. This latter approach was used to address a gap in the research literature on student reasoning, namely, the scarcity of information on what dimensions and/or criteria were used by researchers to judge reasoning performance among their participants.

The combination of the above-mentioned conceptual and methodological elements allowed us to examine the ways in which PK and ECs impact, as well as their interaction or potentially differential influence on, the quality of reasoning among undergraduate science students. What is more, the study’s design enhances our confidence in making inferences that are causal in nature even though randomized experimental studies are still needed to establish cause–effect relationships.

Findings showed significant differences in the quality of reasoning among groups with varying levels of PK and ECs. Both PK and ECs were found to be key factors in determining the quality of reasoning for participant college science students. Overall, there was a significant and positive association between each of PK and ECs and the quality of reasoning. Controlling for PK, the quality of reasoning was positively related to ECs. The higher the EC to the consistency of theory with evidence, the higher the quality of reasoning. And, in cases of weak appreciation for the role of evidence and resistance to contradictory evidence, the process of coordinating theory with evidence was inadequate. Moreover, the reasoning of individuals with High PK–High EC was significantly higher than that for students in the Low PK–Low EC group. Significant differences also existed between individuals with High PK and Low EC, and those with Low PK and Low EC. Therefore, high levels in both PK and epistemological beliefs were strongly associated with high levels in reasoning and vice versa. Also, a particularly important finding was that PK significantly impacts the quality of reasoning especially when ECs are weak. In other words, as ECs weaken, the role of PK in scientific reasoning is strengthened. Alternatively, it is possible that learners’ ECs would contribute significantly more to the quality of their reasoning in domains where their relevant PK is relatively weak. We further illustrate the importance of this latter finding in the following sections.

Epistemology, Reasoning, and Instrumentation

The epistemological scenarios designed for the purpose of the present study aimed to measure the degree of commitment to the consistency of theory with evidence. These scenarios focused on one of the core dimensions of epistemology (Hofer & Pintrich, 1997), namely, beliefs regarding justification of knowledge. Other, more traditional, assessments of epistemological beliefs, such as Likert-scale questionnaires (e.g., Hofer, 1997; Schommer, 1990), usually identify global epistemological beliefs related to elements of epistemology.

The choice of assessment tool to examine epistemological dimensions determines to a large extent the validity of results in research on scientific reasoning. The type of instrumentation used could either uncover or blur the relationship between epistemology and reasoning. What is more, it seems that focusing on specific dimensions of epistemology—such as appreciation for the consistency of theory and evidence, rather than attempting to place individuals in global epistemological categories (such as absolutists, multiplists, relativists, etc.) could be a more useful methodological approach. Instead of assessing the extent to which individuals agree or disagree with a given statement, designing scenarios that indirectly elicit their epistemological beliefs, by examining their reactions to contradictions between theory and evidence for instance, could be a more sensitive and accurate research methodology.

Prior Knowledge, Epistemological Commitments, and Reasoning

It was not surprising that the present results showed a strong impact of PK on the quality of scientific reasoning. Research studies have long emphasized the role of PK in reasoning (e.g., Klahr & Dunbar, 1988; Klaczynski et al., 1997; Kuhn et al., 1992). In the design of the present study, PK was considered an
The same content was used, that is, buoyancy in liquids and Archimedes’ Principle, in the physics pretest and the reasoning scenarios. The study examined students’ reasoning processes as they used their PK to solve problems relating to buoyancy. Therefore, it was expected that those with better understandings of the content would, in general, reveal better reasoning.

The conceptualization of reasoning employed in this study was that of the coordination of theory with evidence (Kuhn, 2004). Successful coordination is based on the ability to distinguish between theory and evidence, identify relations between them, and reflect back on one’s own thinking processes (i.e., have metacognitive awareness of the whole process). While participants were tackling the reasoning scenarios, they had to employ their prior understandings of the topic. However, to display successful coordination between their theoretical ideas and the evidence or observations presented in the tasks, they had to accurately build relations between their ideas and the instances given in these tasks. In other words, they needed to link ideas and provide explanations and justifications that reveal depth of conceptual processing. Participants also had to reflect back on their ideas and seek alternative or competing explanations. So, it is not merely what they knew that determined the quality of their reasoning, but rather what they knew and how they used and implemented their knowledge.

PK, while crucial, is surely not the sole determinant of successful reasoning. Being committed to the consistency of theory with evidence does indeed contribute to successful coordination of theory and evidence. ECs related to consistency of theory with evidence seem to spark the processes of careful evaluation of evidence and examination of prior ideas. The case of Roger serves as an exemplar: Roger belonged to the Low PK–High EC group. He had revealed weak or partially accurate PK (as indicated by his score on the physics pretest) but displayed a relatively high quality of reasoning (as indicated by his score on the reasoning scenarios). Roger had a strong commitment to checking the consistency of theory with evidence, as indicated by his performance on the epistemology scenarios. Thus, it appears that Roger’s strong EC regarding the role of evidence contributed more to the quality of his reasoning than did his PK. His weak PK of the addressed content did not impede his reasoning or coordination of theory and evidence. In the reasoning scenarios, he revealed persistence in considering alternative explanations and weighing them against each other. He used “simple” ideas but in a skillful manner. He attended to the details presented in each reasoning scenario and reflected back on his own thinking. He tried hard to use evidence to support inferences. The case of Roger shows that PK is not a necessary precondition for high reasoning.

On the other hand, the role of ECs in reasoning could be overshadowed by PK. John presented a case that was opposite to Roger’s. John revealed accurate prior understandings on the physics pretest and had the highest reasoning score. John, however, revealed a relatively weak commitment to checking consistency of theory with evidence in both the online questionnaire and the interview epistemology task. He, therefore, displayed strong attachments to his prior beliefs. For the relevant physics content of this study, John’s prior conceptions were highly accurate. His apparently weak EC regarding the role of evidence did not, in turn, impede his reasoning or coordination of theory and evidence. John’s case is a reminder of the study’s finding that the role of PK in reasoning is more vivid when ECs are weak. It would be interesting to examine John’s reasoning in novel tasks in which his PK would be deficient, or in contexts that provide data conflicting with his prior beliefs.

Overall, the findings support the notion that advanced scientific reasoning is most often accompanied by accurate PK as well as sophisticated ECs. Nevertheless, high reasoning could be manifested, albeit with a smaller probability, if either of PK or ECs is strong. The reasoning of individuals belonging to the High PK–Low EC group and Low PK–High EC group needs more examination to further establish the relative influence of PK and ECs on reasoning. Also, other factors, affective or motivational (e.g., Klaczynski & Narasimham, 1998), and how they interact with prior and epistemological beliefs is worth further exploration.

Implications

Scientific reasoning is important in science learning and everyday life problem-solving (Reif & Larkin, 1991; Williams et al., 2004). Understanding the nature of, and the factors that influence, scientific reasoning has significant curricular and instructional implications for the purpose of enhancing students’ reasoning skills. A particularly relevant finding in this regard was that for participant college science students who had
comparable levels of PK, skillful reasoning was associated with a strong EC to the consistency of theory with evidence. This finding highlights the importance of the need for instructional activities and sequences that intentionally and explicitly help learners develop sophisticated ECs (Abd-El-Khalick & Lederman, 2000b), particularly in relation to dimensions of personal epistemology focused on the nature of knowledge and the role of evidence in supporting knowledge claims. Instruction that is solely focused on promoting content knowledge and understandings, while necessary and surely helpful, might not suffice to promote higher levels of reasoning among learners. Therefore, devising curricula that focus on and promote students’ epistemological beliefs are likely to contribute to the goal of achieving more desired and enhanced quality of reasoning.

Additionally, the present findings have implications for the nature of instruction targeting the development of learners’ ECs for the purpose of enhancing their reasoning. Our findings indicate that the quality of reasoning can hardly be divorced from the content and context in which it is undertaken. Indeed, irrespective of the strength and sophistication of their EC, students in the High EC group still needed to draw on their content understandings to reach accurate conclusions through reasoning. Thus, it could be argued that ECs are better nurtured in content-specific domains before explicitly helping learners to perceive—again by application—their relevance to tackling problems in a different set of domains. Students’ epistemological understandings could be enhanced by designing domain-specific inquiry activities that, nonetheless, explicitly address the generalized role of evidence in the justification of knowledge claims. Previous studies have examined the role of inquiry teaching in enhancing students’ reasoning skills (e.g., Schauble et al., 1995). Findings indicated, for instance, that it is not enough to expose students to hands-on activities without giving them direct instruction on experimentation strategies. As such, if the goal is to promote the quality of reasoning by enhancing dimensions of epistemological beliefs, then there is a need to design instructional techniques that incorporate explicit and reflective elements into inquiry-oriented approaches (Abd-El-Khalick, 2005).

As far as research is concerned, this study highlights some methodological issues that present challenges to investigating the relationship between scientific reasoning and personal epistemology. First, there is a need to develop valid instruments that could be easily administered and scored in the case of large samples. Conducting in-depth individual interviews to examine the quality of student reasoning is not feasible in the case of large sample. Needless to say, larger sample sizes make more feasible the process of setting up more experimental manipulations as well as result in higher statistical power, which consequently increase the potential to identify more nuanced relationships between and among the various factors that impact student reasoning. Research is needed to build feasible and valid approaches to assessing both reasoning and epistemology. In this regard, the role of PK should not be ignored in research on scientific reasoning. It would be fruitful to examine how PK and epistemological beliefs impact the quality of reasoning across tasks embedded in different science content areas.

Second, the examination of participants’ performance on the reasoning scenarios used in the present study indicates that students, in general, do not question their PK or existing ideas, which reflects a weak epistemological stance. Providing students with the relevant theory (Archimedes Principle in the present case) did not by itself induce them to revise and change their original responses. Therefore, in future research, students could be specifically asked to distinguish between their ideas and the theory and evidence with which they are provided and then asked to draw conclusions. For instance, students could be asked in what ways is a certain theory (e.g., a canonical theory that is relevant to reasoning about a certain situation) plausible and the ways in which this theory is similar or different from the ideas they already ascribe to. Such an approach might help learners to achieve some “distance” from their prior beliefs when evaluating evidence and hence demonstrate more successful theory—evidence coordination. The validity of this latter hypothesis remains to be established through further research.

References


Journal of Research in Science Teaching


Appendix A:
Illustrative Items From the Questionnaire Used to Assess Participants’ Prior Knowledge of Hydrostatics

(1) How can a steel ship weighing about 500 t float in water?
(2) A boulder is thrown into a deep lake. As it sinks deeper and deeper into the water, does the buoyant force on it increase, decrease, or stays the same? Justify.
(10) Use the letter corresponding to each object (i.e., A, B, C, or D) to answer the following questions:

The teacher walks by and hands Tony and Mary four objects made of the same material as the large and small balls:

The objects are:

A
B
C
D

Mary reminds Tony that the large ball sank earlier when they put it in the water.

What do you think will happen to each of the objects A, B, C, and D, if they were placed in the water tank below?
Which objects (if any) that will sink to the bottom?
Which objects (if any) will float?
Which objects (if any) will remain suspended under the surface?

Appendix B:
An Illustrative Example of the Epistemological Scenarios Used to Assess Participants’ Commitment to the Consistency of Theory With Evidence

Mass Extinction Theory

Several mass extinctions had occurred throughout Earth’s history. Probably the most famous of these is the extinction that occurred about 65 million years ago and brought an end to the reign of dinosaurs. This extinction happened toward the end of the Cretaceous period (abbreviated with the letter K). It was not just the dinosaurs that died out during this mass extinction. Indeed, the catastrophic event responsible for this extinction had caused the death of around 70% of all living species on Earth. Mammals somehow survived this extinction to usher in a new period, the Tertiary period (abbreviated with the letter T), and later became the dominant group of large animals on Earth. The geological boundary between these two periods is known as the Cretaceous–Tertiary (or K–T) boundary.

Scientists believe that a giant asteroid, 10 km in diameter, struck Earth around the time of the formation of the K–T boundary at a velocity of more than 10 km/second. When it hit the ground, its kinetic energy would be converted to heat in a nonnuclear explosion 10,000 times as strong as the total world arsenal of nuclear weapons. The impact liberated enormously huge amounts of energy and caused a chain of environmental disasters, such as global fires, storms, tsunamis, cold and darkness, greenhouse warming, and acid rains. These events led to the mass extinction of dinosaurs in no more than a thousand years.

The asteroid would clearly kill everything within sight of the fireball generated on impact. The impact also would have lifted huge amounts of rock, dust, and vapor into the atmosphere. Coupled with ash from global fires, the dust in the atmosphere would result in extended periods of darkness by blocking sun rays. The darkness would produce extremely cold temperatures; a condition termed “impact winter.” Also, the vapor released from the impact would stay in the atmosphere more than dust, and so the impact winter would have been followed by greenhouse warming. Many plants and animals that survived the extreme cold of impact winter would eventually be killed by a subsequent period of extreme heat. Additionally, without sunlight plant photosynthesis would have stopped, and food chains everywhere would have collapsed. Dinosaurs that survived the initial impact eventually would have no access to reliable and plentiful sources of food they needed for their survival.

The impact theory is supported by good evidence. Scientists studying the clay layer at the K–T boundary in Gubbio, Italy, came upon an unusually high spike in the amount of the element iridium. The levels of iridium contained in the K–T boundary clay were roughly 30 times the normal levels found on Earth. Iridium is an extremely rare element; most of the Earth’s allotment is alloyed with iron in the Earth’s core. The iridium spike found at the K–T boundary could not have come from Earth or cosmic dust. On the other hand, iridium is highly concentrated in some meteorites. It has been calculated that a chondritic asteroid approximately 10 km in diameter would contain enough iridium to account for the iridium spike contained in the K–T boundary throughout the globe. Moreover, analysis of the K–T clay layer has revealed the presence of soot (very fine ash), which is believed to have come from the enormous global fires caused by an impact. Shocked quartz crystals have also been found in the clay. This form of altered quartz only occurs under conditions of extreme temperature and pressure such as those caused from meteorite impacts or nuclear explosions.
Appendix C: Reasoning Scenarios

(1) A block of wood is floating on the surface of still water as shown in the figure below.

(a) Is there any force or forces acting on the wooden block? If yes, specify the kind of force and compare their magnitudes (strengths).

(b) Sugar is dissolved in the water. As a result, would the block float lower, higher, or at the same level? Explain why.

Journal of Research in Science Teaching
(2) Five blocks of the same size and shape but different masses are shown in the figure below. The blocks are numbered in order of increasing mass (i.e. \( m_1 < m_2 < m_3 < m_4 < m_5 \)). All the blocks are held approximately halfway down in an aquarium filled with water and then released. The final positions of blocks 2 and 5 are shown in the figure. On the diagram, sketch the final positions of blocks 1, 3, and 4. Explain your reasoning.

(3) Three cubical blocks of equal volume are suspended from strings of varying lengths. Blocks A and B have the same mass and Block C has less mass. The blocks are lowered into a tank of water and end up in the positions shown in the figure below.

Rank the buoyant force acting on each block from largest to smallest. If any buoyant forces are equal indicate that explicitly. Explain your reasoning.

(4) A metal block sits on top of a floating wooden block.

If the metal block is dropped into the bottom of the beaker, does the level of water in the beaker increase, decrease, or stay the same? Justify your answer.