Effects of Argumentation Scaffolds and Problem Representation on Students’ Solutions and Argumentation Quality in Physics

Carina M. Rebello, University of Missouri, Science Education Center, cp5xc@mail.missouri.edu
Eleanor Sayre, Kansas State University, Department of Physics, esayre@phys.ksu.edu
N. Sanjay Rebello, Kansas State University, Department of Physics, srebello@phys.ksu.edu

Abstract: Prior studies have revealed students’ difficulties in problem solving with multiple representations due to poor reasoning skills. Research has shown that inclusion of argumentation tasks can improve these skills. We investigate alternative forms of argumentation scaffolds (to construct or evaluate an argument) integrated within physics problems, and their impact on students’ solutions and argumentation quality. Results suggest that the scaffolds, not mere inclusion of argumentation, can improve students’ argumentation skills. Additionally, in verbal and graphical representation problems, evaluation of an argument with prompts improves students’ conceptual quality of the solutions.

Introduction
Since the early 1980s, there have been numerous studies regarding college students’ difficulties with conceptual understanding and problem solving with multiple representations in physics due to poor reasoning skills (Hsu, Brewe, Foster, & Harper, 2004). Innovative pedagogical approaches such as Physics by Inquiry (PbI) (McDermott, 1996) have emerged within the physics community to help enhance students’ scientific reasoning skills. PbI employs both open-ended question tasks and hypothetical student debate tasks requiring application of conceptual knowledge and reasoning skills. Such debate tasks provide a useful context for emphasizing reasoning skills and argumentation in the curriculum. Prior studies have shown that embedding argumentation activities in problem solving can improve problem solving abilities, ways of thinking, and conceptual change (Jonassen & Kim, 2010). Argumentation activities can enhance scientific reasoning, facilitate construction and enhancement of knowledge (Driver, Newton, & Osborne, 2000), improve proficiency in advancing, critiquing and justifying claims, and facilitate formative assessment (Osborne, Erduran, & Simon, 2004). An emphasis on argumentation is consistent with the goal of improving students’ reasoning in problem solving and proficiency in advancing, critiquing, and justifying claims (Kuhn and Udell, 2003). Research also suggests that the development of argumentation skills may lead to deeper conceptual understanding. However, the relationship between argumentation and conceptual understanding is not straightforward (Dawson & Venville, 2009).

In this study we examined the effects of alternative forms of argumentation on students’ physics solutions. The PbI problem format was modified to scaffold students’ construction or evaluation of arguments and compared to typical PbI question formats. Additionally, the goal of this study was to examine the effects of argumentation and quality of solutions for problems consisting of different forms of representations.

Literature Review
Argumentation is a critical thinking used in formal and informal contexts to resolve questions, make decisions, formulate ideas, and solve well- and ill-structured problems (Cho & Jonassen, 2002; Jonassen, 1997). It has been used to assess problem solving abilities (Jonassen & Kim, 2010). Kuhn (1991) identifies five key argumentation skills: generating causal theories offering supporting evidence, envisioning conditions that undermine one’s theory, generating alternative theories, and rebutting alternative theories. Successful argumentation requires a problem solver to develop and articulate a reasonable solution, support the solution with data and evidence, and identify alternative solutions (Cho & Jonassen, 2002). The strength of an argument depends upon the context and nature of the task (Newton, Driver, & Osborne, 1999).

Studies have found that most U.S. students lack argumentation skills and have difficulty comprehending arguments, writing persuasive essays, differentiating theory and evidence, and producing evidence, alternatives, counterarguments, or rebuttals (Reznitskya et al., 2001). Zeidler (1997) identified several students’ problems with argumentation, such as: selecting only evidence that supports their claim or “my-side bias,” reliance on personal beliefs rather than counter evidence, overgeneralization, and making unsupported assertions. Additional studies found that students tend to rely on data to support their claims but frequently do not include warrants or backings (Bell & Linn, 2000). These findings have implications for the design of classroom activities to improve students’ argumentation.

Research on students’ problem solving reveals that most students use a “means-end approach” and have difficulty solving problems with multiple representations such as graphs (Meltzer, 2005; Rosengrant, Etkina, & van Heuvelen, 2006). To become better problem solvers students must improve their conceptual understanding, learn to solve problems using multiple representations (Rosengrant et al., 2006), and must be engaged in tasks that require meaningful justifications requiring application of argumentation during problem solving.
solving (Jonassen et al., 2009). Argumentation consists of two skills – constructing and evaluating arguments. Effective argumentation requires evaluating, weighting evidence and counterevidence, and considering arguments, counter-arguments, and rebuttals to arrive at a solution (Jonassen et al., 2009). Here we explore how introductory physics students can be supported to achieve superior argumentation skills in problem solving.

Theoretical Framework
Toulmin’s argumentation pattern (TAP) (1958) which is commonly used to assess and/or instruct learners about argument structure describes argument features as (i) claims – conclusions or assertions, (ii) data – facts that provide the foundations of the claim, (iii) warrants – reasons, rules, or principles that provide connection between data, claim, or conclusion, (iv) backing – assumptions used to provide justification of the warrant, (v) qualifiers – conditions or limitations under which the claim is true, and (vi) rebuttals – conditions when the claim is not true (Driver et al., 2000). In this study a rubric adapting TAP is used to assess the structures of arguments offered by students.

Research Questions
We addressed the following research questions in the context of solving conceptual physics problems.

- What level of argumentation and conceptual quality do introductory college physics students, without training in argumentation, demonstrate in their solutions to physics problems?
- How do the argumentation and conceptual quality of students’ solutions for these problems change based on…
  - the representation – verbal or graphical – in which information is provided in the problem?
  - whether the problem requires learners to construct or evaluate arguments?
  - the use of specific prompts designed to scaffold the construction and evaluation of arguments?

Research Context
This study was conducted at a large, US public university, in a semester-long (15 week) course offered in the physics department. Participants were enrolled in an algebra-based physics course. The course lecture met three times weekly for 50 minutes each and one laboratory (110 minutes) and one recitation (50 minutes) a week. Topics covered include kinematics, dynamics, work and energy, rotational motion, oscillatory motion, fluids, wave and sound, and thermodynamics. The course is taken primarily by life/health science majors with an enrollment of about 400 students and employs a textbook and online homework system. Homework, recitation problems, and exams tend to emphasize quantitative problem solving rather than argumentation.

Data Sources
We created two problems – problem 1 (in verbal representation) and problem 2 (in graphical representation) – both targeting a key kinematics misconception (Trowbridge & McDermott, 1980; Beichner, 1994). Each problem was presented in four conditions – construct, evaluate, control construct, and control evaluate – with argumentation prompts shown in Table 1. Evaluate and construct prompts were adapted from Jonassen et al. (2009) and Mason and Scirica (2006). Control prompts were similar to those used in Pbl.

<table>
<thead>
<tr>
<th>‘Construct’ Prompts</th>
<th>‘Evaluate’ Prompts</th>
</tr>
</thead>
</table>
| ✓ Construct an argument to justify your answer.  
 ✓ Explain your position clearly and completely by providing all reasons that support your conclusion. Remember to consider:  
 ✓ What evidence supports your reasons?  
 ✓ One of your classmates may disagree with your conclusion. What might they think is the alternative conclusion?  
 ✓ What reasons would your classmate provide to support their conclusion?  
 ✓ What would you reply to your classmate to explain that your position is right? | ✓ Which statement (of the ones provided above) best describes the physical phenomenon? Or do you have another argument?  
 ✓ Explain, elaborate, and justify your preferred solution. Remember to consider:  
 ✓ What evidence and reasons support your selection?  
 ✓ Explain your reasoning for not choosing the alternative solution(s). What are the weaknesses in the alternative argument(s)?  
 ✓ How might a classmate supporting the other solution disagree with your preferred solution?  
 ✓ What would you reply to your classmate to explain that your position is right? |

<table>
<thead>
<tr>
<th>‘Control Construct’ Prompts</th>
<th>‘Control Evaluate’ Prompts</th>
</tr>
</thead>
<tbody>
<tr>
<td>✓ What is your answer? Explain your reasoning.</td>
<td>✓ Which statement do you agree with? Or do you have another answer? Explain your reasoning.</td>
</tr>
</tbody>
</table>
The study was completed four weeks into the semester. The problem versions were administered randomly to students in 10 out of 19 laboratory sections, each enrolling approximately 20 students. Participants were offered extra credit as incentive and assured confidentiality. Response rate was at least 95% for each of the participating laboratory sections yielding a total of 198 respondents. Figure 1 shows a flow chart of the design.

| 198 respondents in 10 out of 19 lab sections | Construct (N = 65) | Prob1 (verbal) [Construct] → Prob2 (graphical) [Construct] |
| Evaluate (N = 68) | Prob1 (verbal) [Evaluate] → Prob2 (graphical) [Evaluate] |
| Control (N = 65) | Prob1&2 (verbal & graphical) [Control Const] → [Control Evaluate] |

Figure 1. Flow chart showing the design of the study

Data Analysis
After reading through students’ written responses, we created three scoring rubrics – two for conceptual quality and one for argumentation quality. The first two rubrics were designed to analyze students’ physics conceptual quality of their problem solutions for each problem. The holistic conceptual quality rubric for Problem 1, which had two possible answers, was on a six-point scale: incorrect answer, does not recognize either of the possibilities, with little or no reasoning (0); incorrect answer – does not recognize either of the possibilities but has some good scientific reasoning (1); recognized one of the possibilities with incomplete or no reasoning (2); recognizes one of the possibilities with fully correct reasoning (3); recognizes both possibilities with reasoning, but does not fully justify both possibilities (4); and recognizes both possibilities with correct reasoning (5). The holistic conceptual quality rubric for Problem 2 was an adaptation of the first rubric, but because there was only one scientifically acceptable solution, the rubric had only four levels: incorrect answer with little or no scientifically correct reasoning (0); incorrect answer with some good scientific reasoning (1); correct answer with some incorrect scientific reasoning (2), and correct answer with correct scientific reasoning (3).

To assess argumentation quality of students’ written responses we used a rubric adapted from Sadler and Fowler’s (2006) Argumentation Quality Rubric, which in turn was adopted from Toulmin’s (1958) TAP framework. The rubric allows for analysis of how claims are justified on a five-point scale: no justification provided (0), to one in which claims are justified with elaborate grounds and a counter position, in which contradictory evidence is also evaluated (4). Grounds refer to possible supports for a justification including: data, backing, and warrants. This rubric was adapted into a six-point rubric to allow not only the inclusion of a counter position(s) in the justification statement but also the inclusion of a rebuttal(s) in their solutions to address counterclaim concerns (5). Both conceptual and argument quality were coded by two independent raters (including the author) who were familiar with the design and purpose of the study. After independent coding, the raters discussed all codes to reach 100% agreement.

In order to compare the quality of arguments offered by participants’ responses in each of the four conditions (construct, evaluate, control-construct, and control-evaluate), a multivariate analysis of variance (MANOVA) was performed. The four conditions (construct, evaluate, control construct, and control evaluate) served as the independent variable and the quality of conceptual quality scores and quality of argumentation scores as the dependent variables. Univariate analysis of variance (ANOVAs) was conducted to determine if the conditions have a significant effect on each dependent variable (conceptual quality and argumentation quality).

Findings
Participants from all conditions demonstrated variability in conceptual and argumentation quality in response to the verbal and graphical problem statements. The distribution of scores appears to vary across both problems. Means and standard deviations of conceptual and argumentation quality for each problem are shown in Table 2.

<table>
<thead>
<tr>
<th>Version</th>
<th>Conceptual Quality</th>
<th>Argumentation Quality</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Problem 1 (verbal)</td>
<td>Problem 2 (graphical)</td>
</tr>
<tr>
<td></td>
<td>Mean ± SD</td>
<td>Mean ± SD</td>
</tr>
<tr>
<td>Construct</td>
<td>2.062 ± 1.01</td>
<td>0.862 ± 0.63</td>
</tr>
<tr>
<td>Evaluate</td>
<td>2.368 ± 1.53</td>
<td>1.691 ± 1.11</td>
</tr>
<tr>
<td>Control Construct</td>
<td>1.569 ± 1.03</td>
<td>1.200 ± 1.16</td>
</tr>
<tr>
<td>Control Evaluate</td>
<td>1.508 ± 1.40</td>
<td>1.477 ± 1.23</td>
</tr>
</tbody>
</table>

The MANOVA analysis using conditions (construct, evaluate, control construct, and control evaluate) as independent variables and conceptual quality scores and argumentation quality scores as the dependent variables revealed a statistically significant difference among the four conditions for the two provided problems [Wilks’ Λ = 0.347, F(12.0, 677.6) = 27.8, p < .001, η² = 0.297]. Univariate ANOVAs were explored to test if
the independent variable has a significant effect on each dependent variable. Results (see Table 3) show significant differences between conceptual and argumentation scores for the two problems.

Table 3: Univariate ANOVAs.

<table>
<thead>
<tr>
<th>Problem</th>
<th>F</th>
<th>df</th>
<th>p</th>
<th>Effect Size</th>
</tr>
</thead>
<tbody>
<tr>
<td>Problem 1 (Verbal) – Argumentation Quality</td>
<td>104.97</td>
<td>3, 259</td>
<td>.000</td>
<td>.549</td>
</tr>
<tr>
<td>Problem 1 (Verbal) – Conceptual Quality</td>
<td>6.94</td>
<td>3, 259</td>
<td>.000</td>
<td>.074</td>
</tr>
<tr>
<td>Problem 2 (Graphical) – Argumentation Quality</td>
<td>84.05</td>
<td>3, 259</td>
<td>.000</td>
<td>.493</td>
</tr>
<tr>
<td>Problem 2 (Graphical) – Conceptual Quality</td>
<td>7.74</td>
<td>3, 259</td>
<td>.000</td>
<td>.082</td>
</tr>
</tbody>
</table>

Follow-up Tukey’s HSD analysis with an overall alpha level of .05 revealed that for problem 1 (verbal) argumentation quality, there is a significant difference between scores on evaluate and construct conditions, the latter having a greater argumentation quality. Scores on the evaluate condition were also statistically greater than scores on the control conditions. Additionally, scores on the construct condition were statistically greater for argumentation quality than scores on the control conditions. Thus, results for argumentation quality in problem 1 (verbal) indicate that the construct condition with prompts yields a higher argumentation quality than the evaluate condition with prompts or either of the control conditions. Also the evaluate condition with prompts tends to yield higher argumentation quality than either of the control conditions as well. Results suggest that the construct and evaluate prompts used in problem 1 do help produce higher argumentation quality than the control problem statements, which are similar to problem statements used in the PbI curriculum. Also for the verbal problem statement, a construct argument condition with prompts produced higher argumentation quality.

For problem 1 (verbal) conceptual quality in problem solutions, scores on the evaluate condition were statistically greater than scores on the control conditions. Thus in the evaluate condition, if prompts are provided students have a higher conceptual quality in their problem solutions, but problem format (construct or evaluate) does not make a statically significant difference on the conceptual quality of problem solutions.

For problem 2 (graphical) argumentation quality scores on construct were statistically higher than scores on control conditions. Also there were significant differences for scores on evaluate and control conditions. Scores on the construct condition were also statistically higher than evaluate conditions. Thus, results for argumentation quality in problem 2 (graphical) indicate that prompts are important, particularly for the construct condition rather than for the evaluate condition. Yet, problem format does not make a difference on students’ argumentation quality.

Finally, for problem 2 (graphical) the evaluate conditions yielded a higher conceptual quality in the problem solution than construct conditions. Also, there was a statistically significant difference between evaluate and control-construct conditions and between construct and control-evaluate conditions. Thus, results for conceptual quality in problem 2 (graphical) indicate that prompts do not matter for the graphical question to improve conceptual quality. If prompts are given, they are more effective in the evaluate condition than the construct condition but the use of prompts is not more effective than the corresponding control conditions.

Conclusions and Implications

Our results indicate there are statistical differences between the four conditions for each problem. Both problem 1 (verbal) and problem 2 (graphical) assess the same conceptual topic (kinematics) and were prone to the same misconception. Though their representations were different, the results of the argumentation qualities of the two problems were very similar across the different conditions. Results for argumentation quality in problem 1 indicate that if prompts are provided, the construct condition produces higher argumentation quality. Additionally, for argumentation quality in problem 2 the construct condition produces a higher argumentation quality than the evaluate or control conditions. Hence, results regarding argumentation quality suggest that the problem format (construct/evaluate) does not make a difference but whether or not prompts are used does make a difference. Similar to Jonassen et al.’s (2009) results, argumentation quality for each problem representation suggests that using construct and evaluate alternative argument tasks with prompts produce higher argumentation qualities. In fact, typical problem statements (“explain your reasoning”) used in the PbI and other curriculum tend not to produce higher argumentation quality if students are not appropriately guided to provide reasons. Hence, the open-ended and hypothetical student debate problems as used in the PbI curriculum may not produce higher argumentation level such as considering possible counter-arguments and rebuttals, without appropriate scaffolding to do so. However, unlike Jonassen et al.’s (2009) results, our results indicate that there is a significant difference between construct and evaluate conditions for stronger argument structures.

There were also notable similarities between the problems in conceptual quality. It seems that the evaluate condition with prompts for problems 1 (verbal) and 2 (graphical) can have a greater impact on conceptual quality. For problem 1 (verbal), the construct condition does not outperform the evaluate condition on conceptual quality. Specifically, the use of a hypothetical student debate problem used in PbI may not produce higher conceptual quality justification used in arguments without suitable scaffolding to appropriately
evaluate arguments and justify a position. However, results indicate that there was no difference in conceptual quality between construct and evaluate conditions indicating that argumentation formats do not make a difference. Also, there are no conceptual quality differences between construct and control-construct conditions. Thus, the use of prompts does not make a significant difference in conceptual quality in the construct condition. Finally, results for conceptual quality in problem 2 (graphical) indicate that the evaluation condition yields a better conceptual quality problem solution than the construct condition or the control-construct condition. This result could be attributed to the presentation of two alternative graphical solutions, one with a misconception and the other scientifically correct, for students to evaluate and choose from as opposed to constructing justifications entirely on their own. Interestingly for problem 2 (graphical), prompts do not improve conceptual quality. Yet if prompts are provided, they are more effective for the evaluate condition than the construct condition. The difference in the effectiveness of the prompts depending upon problem representations has not been probed previously. Thus in light of the results obtained here, more research in this area is warranted to study the issue further.

References