Undergraduate Cognitive Psychology Students' Evaluations of Scientific Arguments in a Contrasting-Essays Assignment

Jordan P Lippman, Frances K. Amurao, James W. Pellegrino,
University of Illinois at Chicago, LSRI (M/C 057), Chicago IL, 60607
Email: jlippman@uic.edu, famura2@uic.edu, pellegjw@uic.edu,
Trina C. Kershaw, University of Massachusetts Dartmouth, Psychology,
North Dartmouth, MA 02747, tkershaw@umassd.edu

Abstract: This study investigated upper-level college students' understanding of evidence use in quality scientific arguments. Responses to a required online 'contrasting-essays' assignment provided instructors with quick access to formative information about students' capacity to evaluate the quality of argument and the use of evidence therein. Students inconsistently applied criteria for strong and weak evidence, focusing on superficial aspects of writing quality and ease of comprehension instead of the need for relevant, empirical, and disconfirming evidence.

Argumentation has a central role in scientific thinking because it is the means by which scientists generate and communicate their empirical findings and causal explanations (e.g., Kuhn, 1993). Our prior work (Lippman et al., 2006) and the work of other researchers (e.g., Chinn, 2006; Sandoval & Millwood, 2005), suggests that novices, such as advanced high school students and undergraduates, have great difficulty learning to create scientific arguments (e.g., Chinn, 2006). They have particular difficulty with explaining how empirical evidence provides support for their main claims (Sandoval & Millwood, 2005).

We constructed a "contrasting-essays" activity, based on the contrasting-cases activity designed by Schwartz and Bransford (1998), to introduce students to the type of scientific arguments we expected them to produce for course assignments and to simultaneously inform us about their proficiency with evaluating arguments Completed shortly before their first argumentation essay assignment at the beginning of the semester, the contrasting-essays activity asked students to read, compare, and analyze the quality of the argument and the use of evidence in two sample essays. Because we assumed students would need at least a basic understanding of the content material to evaluate written arguments, we had them work with sample essays about content covered the prior week in class and in their textbook readings (i.e., results from mental imagery studies and the theory of functional equivalence). We took great care to design both essays to have flaws so that at first glance it would not be obvious which was the best essay or why, and we assumed that an analysis of student responses would provide insight into their understanding of argument and their criteria for evaluating evidence.

Method

Ninety-eight students from an upper-level *Introduction to Cognition* course at a large, urban, midwestern university provided consent so the researchers could analyze their responses to various course assignments (117 students were enrolled in the course).

The Comparison and Analysis questions from the online assignment required students to make a forced answer selection and then explain their answers. We scored student responses to forced-choice Comparison and Analysis items as Correct if the answer selections matched the instructor key developed prior to constructing the sample essays.

The first two authors worked together to develop and apply coding schemes for the explanation items. For each question, we discussed its purpose and predicted the types of student responses we would see. Then, we independently attempted to fit this hypothesized scheme to a random sample of responses – deleting and adding codes as necessary to capture the diversity of responses and minimize the number of low-frequency codes and uncodeable responses. We then discussed the meaning of our revised schemes and negotiated a final coding scheme for each question. Subsequently, we independently applied the coding scheme to all the responses for a question and resolved disagreements through discussion. Our mean reliability on all questions was 89.10% simple agreement (calculated as total number of agreements divided by the total number of codes applied per question); our lowest reliability was on the *best argument* question (84.10% simple agreement). We developed schemes that shared a set of code types (e.g., argument, evidence, writing quality and comprehension) common to the entire group of Comparison and Analysis questions and that also contained unique sub-codes applicable to individual questions. (the sub-codes for each question are not listed but are available from the first author).

Results and Discussion

On Comparison questions students had to choose the essay with the best (1) argument, (2) use of evidence, and (3) explanation of evidence. On Analysis questions they had to indicate the strongest and weakest evidence for the main claim in each essay. Students answered approximately 2 of 3 Comparison questions, and approximately 1 of 4 Analysis questions correctly. Table 1 presents the percentage of students who were correct on each evidence analysis question for the weak and strong essays.

Table 1: Percent Correct Choices on the Evidence Analysis Questions for the Weak and Strong Sample Essays.

Question	Weak Essay	Strong Essay
Strongest Evidence	28%	18%
Weakest Evidence	11%	45%

To understand this pattern of (relatively poor) performance, we first examined the most frequently selected answers on the Analysis questions, and, second, the types of codes assigned to explanations for Comparison and Analysis questions. The most frequent selection for the Strongest Evidence in the weak essay was an in-class demonstration (incorrect) and in the strong essay the most frequent choice was a study that did not rule out the alternative position (also incorrect). This pattern of findings suggests students think strong evidence is well-described, easy to understand, relevant, and consistent with the main claim but they fail to appreciate the need for empirical studies that rule-out alternative positions through disconfirmatory logic.

The relatively high accuracy on the Weakest Evidence question for the strong essay was probably a result of our inclusion of fewer but higher quality pieces of evidence in this essay compared to the weak essay. This may have lead to artificially high performance because students could have used a process of elimination to select the demonstration as the Weakest Evidence in the strong essay instead of knowing the properties of weak evidence. Interestingly, this demonstration was also the most frequent selection for the Strongest Evidence.

In the weak essay, the students' most frequently selected answers for Weakest Evidence included an inaccurately-described study, a study unrelated to the main claim, and a class-demonstration. Interestingly, we presented a theoretical counterargument against the main claim as a 'study' designed to support the main claim, and very few identified this as the weakest evidence. The inconsistency in how they judged the strength of class-demonstrations as evidence across Analysis questions suggests they do not appreciate or consistently apply the epistemological criteria of empirical data serving as evidence in scientific arguments. Furthermore, their insensitivity to blatantly inaccurate descriptions of evidence suggests students lack the deep understanding needed to assess the quality of evidence. Students made reasonable but not always ideal selections for the strongest or weakest evidence, but they seemed inconsistent in the criteria used to make these decisions.

Examining the percentage of each type of code applied reveals that approximately 70% of the total codes on Comparison questions focused on the quality of the argument or evidence. In contrast, only 40% of codes on the Analysis questions came from these categories. The relatively high proportion of codes (~50%) focusing on the writing quality and ease of comprehension over more important properties of evidence such as relevance or validity on the Analysis questions agrees with other findings about novices' difficulty creating arguments that effectively use empirical evidence (Sandoval & Millwood, 2005)

References

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

Lippman, J. Pellegrino, J, Koziol, R., & Whitehair, E. (2006) Lessons learned from using an asynchronous online discussion board to facilitate scientific thinking in a large cognitive psychology lecture class. In S. Barab, K. Hay & D. Hickey (Eds.), *Proceedings of the seventh annual international conference of the learning sciences: Making a difference.* (pp. 956 - 957). Mahwah, NJ: Erlbaum.

Sampson, V. D., & Clark, D. B. (2006). Assessment of argument in science education: A critical review of the literature. In S. Barab, K. Hay & D. Hickey (Eds.), *Proceedings of the seventh annual international conference of the learning sciences: Making a difference.* (pp. 655 - 661). Mahwah, NJ: Erlbaum.

Sandoval, W. A., & Millwood, K. A. (2005). The quality of students' use of evidence in written scientific explanations. *Cognition and Instruction*, 23, 23-55.

Schwartz, D.L., and Bransford, J.D. (1998) A time for telling, *Cognition and Instruction, 16,* 475-522.