DOCUMENT RESUME

ED 443 401 IR 020 311

| | |
|---|---|
| AUTHOR | Steinkuehler, Constance A.; Derry, Sharon J.; Levin, Joel R.; Kim, Jong-Baeg |
| TITLE | Argumentative Reasoning in Online Discussion. |
| SPONS AGENCY | National Inst. for Science Education, Madison, WI.; National Science Foundation, Arlington, VA. |
| PUB DATE | 2000-04-00 |
| NOTE | 57p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 24-28, 2000). |
| CONTRACT | REC-9725528; RED-9452971 |
| PUB TYPE | Reports - Research (143) -- Speeches/Meeting Papers (150) |
| EDRS PRICE | MF01/PC03 Plus Postage. |
| DESCRIPTORS | *Beliefs; Comparative Analysis; *Computer Assisted Instruction; Higher Education; Instructional Effectiveness; Intermode Differences; Opinions; *Persuasive Discourse; Pretests Posttests; Qualitative Research; *Recall (Psychology) |
| IDENTIFIERS | Self Explanation Strategies; Summarization |

ABSTRACT

This study compared the effects of three forms of online instruction on memory, belief change, and argumentation skill. Reading of a pro/con text was followed by: (1) online discussion in pairs compared to reading of the same text followed by two forms of individualized study techniques derived from the cognitive memory literature; (2) self-explanation; and (3) repeated summarization and study. Results were analyzed in terms of argument change from pretest to posttest, transfer of argument skills, text recall, reported and directly assessed opinion change, and perceptions of productivity and participation. Qualitative analysis of the transcripts from the online activities examined time on task, effects of pair agreement or disagreement, and unequal participation within the pairs. The pro/con texts, coding schemes, and data tables are appended. (Contains 13 references.) (MES)

# Argumentative Reasoning in Online Discussion

Constance A. Steinkuehler, Sharon J. Derry, Joel R. Levin, Jong–Baeg Kim

University of Wisconsin-Madison

Abstract

This study compared the effects of three forms of online instruction on memory, belief change, and argumentation skill. Reading of a pro/con text followed by (a) online discussion in pairs was compared to reading of the same text followed by two forms of individualized study techniques derived from the cognitive memory literature: (b) self–explanation and (c) repeated summarization and study. Across conditions, participant's argumentative reasoning improved from pretest to posttest essay in terms of both a decrease in the proportion of nonjustificatory statements and an increase in the number of metacognitive statements made; however, there were no statistical differences among conditions. Summarization produced greater levels of text recall than did self–explanation, which in turn was more effective than online discussion. Summarization also produced greater use of evidence from the text than did self-explanation and discussion. Use of evidence within the posttest essay significantly correlated with use of evidence in a second, transfer essay, indicating that participants in the summarization condition may have gained a greater propensity for using evidence to support an argument. Though participants in the discussion condition reported greater opinion change, participants in the self–explanation and summarization conditions exhibited greater actual opinion change. However, discussion produced greater accuracy in reported opinion change than did self–explanation, which, in turn, was more effective than summarization. Individuals who engaged in discussion also reported greater personal contribution to the learning activity, although perceived importance of the topic and perceived ease of the technology were equal across conditions. Qualitative analysis of the transcripts from the online activities revealed that participants in the discussion condition spent more time making housekeeping statements, talking socially, and discussing tangent topics than participants in the other two conditions.

The Secondary Teacher Education Project (STEP) at UW-Madison is incorporating more web-based communication and collaboration into its program. In particular, we are making greater use of online learning activities coordinated with readings in an effort to improve pre-service teachers' argumentative reasoning about important educational issues. This goal is important, since research indicates that the ability to formulate and evaluate reasoned arguments within a framework of competing alternatives is more accurately characterized as the exception rather than the norm (e.g., Kuhn, 1991). Understanding what types of experiences foster reasoned argumentation is critical from an instructional design standpoint. What kinds of activities should be incorporated into curricula to promote argumentation skills? How might such instruction capitalize on available online forums designed for communication and collaborative work? Does any form of engagement on a topic improve argumentative reasoning, or are some activities more productive catalysts than others? The purpose of the experiment described here is to provide the first step toward answering such questions.

In a recent study, Kuhn, Shaw, and Felton (1997) used dyadic discussion as a vehicle for examining whether increased time spent engaged in thinking about a topic improves reasoning about it—an assumption underlying much of the literature on education (Dewey, 1910; National Educational Association, 1961, as cited in Kuhn, Shaw, & Felton, 1997). Participants were assigned to one of two conditions—an experimental condition in which individuals met in pairs and discussed their views on capital punishment for an average of 10 to 15 minutes per week for five weeks, and a control condition in which no discussion occurred. In both conditions, pretest and posttest measures included an essay written by each individual stating and justifying their opinion on capital punishment and an opinion scale (Kuhn & Lao, 1996).

The results of Kuhn et al's (1997) study were encouraging. Although participants' opinions were fairly stable throughout both conditions, the arguments constructed to support those opinions improved significantly after discussion. In the experimental condition, participants shifted from one-sided, non-comparative arguments toward two-sided, comparative arguments (weighing both pros and cons). Analysis of transcripts of the discussions revealed two unexpected trends. First, in roughly half of the cases in which arguments improved, new elements incorporated into the posttest argument first appeared during discussion between individuals who held congruent rather than discrepant opinions. Second, in roughly half the cases, these new elements were first introduced into the discussion by the individual who later incorporated them into his or her posttest essay rather than by his or her discussion partner.

Although Kuhn et al.'s findings (1997) suggest that, over time, engagement in thinking about a topic does improve argumentative reasoning, the use of dyadic interaction as the vehicle for engagement bars us from knowing whether such improvement is attributable to time engaged as hypothesized or to the pair-interaction format. Is discussion the actual impetus for improvement in argumentation, or does any form of extended engagement in thinking about a topic affect such improvement?

The notion that social interaction is the primary catalyst for cognitive development is supported by numerous theoretical traditions, including the lines of research originated by both Piaget (1976) and Vygotsky (1978). By "[requiring] that students take positions or stances with respect to the claims and observations made by others" (O'Connor & Michaels, 1996, p. 64), discussion seems particularly well suited for improving argumentation skills. Numerous studies of group interaction in educational settings have highlighted this effect (e.g., Brown & Palincsar, 1989; Pontecorvo & Girardet, 1993; Warren & Rosebery, 1986). In light of such work, it seems reasonable to assume that discussion will more likely improve argumentation compared to engagement of any form.

If, as the research suggests, discussion is the underlying catalyst for the improvement in argumentation that Kuhn et al (1997) found, is it equally as productive in forums other than a face-to-face one? Innovative curricular designs such as that currently under development in the STEP project capitalize on the social function that technologies such as the Internet serve. Extension of Kuhn et al's (1997) findings to "electronic gatherings" (Sproull & Faraj, 1997) as well as traditional ones can inform the future design of innovative learning environments that take advantage of technology's capacity for connecting people across time and place.

The research presented in this paper is a first attempt toward answering the questions raised above. This experiment was designed to compare the effects of different online learning activities on argumentative reasoning. This study extends Kuhn et al.'s (1997) work in five ways. First, three experimental conditions rather than two were included: (a) reading of a given online text followed by discussion in pairs; (b) reading coupled with guided self-explanation (using the method previously developed by Chi, deLeeuw, Chiu, & LaVancher, 1994); and (c) repeated reading, summarization and study. Second, all three types of learning activities occurred in an online environment designed to support discussion as well as other forms of synchronous communication. Third, in order to probe the robustness of effects under typical college classroom conditions, all activities were limited to one class period rather than extended over a period of several weeks. Fourth, assuming that this limited time frame might constrain participants' ability to consider a topic from a variety of perspectives and that students typically receive reading assignments in connection with class discussions, participants read a text containing both pro and con arguments as part of the learning activity. This text served to "seed" the activity—in effect, giving the participants something to argue about. Finally, to begin to understand the processes underlying the impact of varying forms of instructional engagement on argumentative reasoning, participants' perceptions of productivity and participation, as well as their recall for the text, were assessed so that the relationships of these variables to argumentative reasoning could be explored.

Several hypotheses were advanced, including: (1) Online discussion was expected to produce greater improvement in students' ability to reason argumentatively. (2) Individualized forms of online instruction derived from the cognitive memory literature were expected to produce relatively better recall of the text. (3) Perceptions of participation and importance of the topic for consideration would be greater for online discussion relative to the two individual conditions and would be correlated with reported and actual opinion change, which would also be greater for online discussion.

## Method

### Participants

One hundred and five undergraduates enrolled in an educational psychology course volunteered for the study in exchange for extra credit. The majority of the participants were enrolled in the School of Education; other majors included service fields such as communicative disorders, occupational therapy, and nutrition. Of the 105 volunteers who participated, 15 were excluded from analysis due to problems that occurred during their session—either the online environment used for the activity was inaccessible or an assigned partner was absent. The final data set thus represented 90 participants, 30 participants in each of three conditions. Participants were scheduled in pairs, randomly assigned to one of three conditions, and placed in chat rooms within the online environment to participate in the activities designated by the condition to which they were assigned. Because treatment was administered to pairs, the scores of each individual in a given pair were averaged together, resulting in a final data set consisting of 15 data points in each condition with the pair as the unit of analysis.

### Apparatus

This study employed use of an online environment, Tapped In (www.tapped.org), which is a virtual community center designed to facilitate teacher professional development. As used in this study, this environment appeared as a Java applet window similar to many basic chat rooms, consisting of two areas—a large upper panel in which text is viewed and a horizontal bar below in which users enter text. A scroll bar along the right side of the window enabled participants to review all the statements entered into the system at any given time. Figure 1 shows the computer window as it appeared to participants in the study. The environment automatically generated transcripts of all online activities.

---

Insert Figure 1 about here.

---

Design and Procedure

In overview, the experiment consisted of four parts—pretest, reading, an activity, and posttest. Activity was the only variable that differed across conditions. Treatments were administered to pairs of students in a small computer laboratory. Pairs of students received experimental instructions, tests, and training on the environment together, but worked individually during the instructional activity at separate terminals that were partitioned to eliminate visual contact between students. The topic for consideration was whether creationism should be placed on an equal footing with evolution in high school science classrooms.

Pretest. Participants were first given an explanation of the experiment followed by two pretest measures: (a) they prepared a brief written essay stating and justifying their opinion on whether creationism should be placed on an equal footing with evolution in high school science classrooms, and (b) they completed an opinion scale designed to directly assess opinion and, by comparison to posttest, opinion change. This scale, modeled after the one used by Kuhn and Lao

(1996), consisted of 13 points with the midpoint of the scale labeled "I have mixed or undecided feelings about placing creationism on an equal footing with evolution in high school science classrooms." Adjacent points (+1, −1) were also labeled with statements indicating uncertainty (e.g. "I am somewhat in favor of...but I'm not sure."). Points with increasing magnitude were labeled with statements indicating increasing conviction; for example +2 was labeled with the statement "I am somewhat in favor of…" and +3 was labeled with the statement "I am in favor of…" The scale included elongated ends (e.g., +4, +5, +6) in order to minimize extreme pretest scores. Each of these extreme points was labeled with the same statement indicating polarity in opinion; for example, +4, +5, and +6 were each labeled with the statement "I am totally in favor of placing creationism on an equal footing with evolution in high school science classrooms." However, +5 and +6 were also labeled with statements indicating increasing certainty: for example, both +5 and +6 also included the statement "I will never change my mind, no matter what new information or arguments I hear," and +6 also included the statement, "I can't imagine how anyone else could believe differently." (For a full discussion of the rationale behind this design, see Kuhn & Lao, 1996).

Reading. Following the pretest period, both participants in the pair were instructed to read an online text appearing on their respective computer screens. The text, presented in Appendix A, contained two arguments—one for and one against including creationism in secondary science curriculum—that synthesized the position statements made by various organizations currently involved in the debate (e.g., the Creation Research Society, the National Education Association, and the Wisconsin Department of Public Instruction). This text appeared as a web page on each participant's computer screen. Participants closed the window containing the text once they completed the reading assignment.

Training. After reading the text, both participants were briefly shown how to use the study environment for either discussion, self-explanation or summarization. Because training was brief (lasting an average of 3–4 minutes), required use of only one basic command (an open quotation mark at the beginning of each statement entered), and required reference to the reading, training was administered after the initial reading and immediately before the activity began.

Activities. Participants then engaged in one of three activities. The activities lasted 45 minutes in all three conditions. In the discussion condition, participants "talked" to one another on line. The discussion topic was posted in the text window, and participants were instructed use the entire 45 minute period to develop a joint statement on the topic. Figure 2 illustrates the discussion activity from the perspective of one participant.

---

Insert Figure 2 about here.

---

In the self-explanation condition, participants were guided through a self-explanation activity modeled after the method previously developed by Chi, deLeeuw, Chiu, & LaVancher (1994). During this activity, portions of the text (chunked into conceptually meaningful segments of a few sentences each) appeared in boxes in the online environment text window one segment at a time. Participants received 15 separate portions of text during the 45–minute activity period, with one portion of text appearing roughly every three minutes. Participants were instructed to type in their self–explanation of each segment after it appeared, following the directions previously used by Chi et al (1994). Figure 3 illustrates the self-explanation activity from the perspective of one participant.

---

Insert Figure 3 about here.

In the summarization condition, participants read the online text as in the other two conditions, removed the text and made notes (using the online environment) on its contents from memory, then studied their notes. After 15 minutes, the text reappeared on the screen and the activity was repeated—participants read the text again, made additional notes from memory, and studied their notes until the text reappeared a third and final time. Figure 4 illustrates the summarization activity from the perspective of one participant.

Insert Figure 4 about here.

Posttest. Following the activity, participants took a short break and then completed six posttest measures: (a) Participants first completed a short–answer questionnaire designed to assess recall for the given pro/con text. The recall test consisted of 11 short–answer questions. Two questions required recall of six arguments made within the text (three pro arguments, three con arguments); the remaining five questions assessed recall for selected details of the text. (b) Participants then completed a scale designed to measure reported opinion change, also modeled after one used by Kuhn & Lao (1996). This 13–point scale asked participants to indicate the extent to which their opinion had changed as a result of their participation. The midpoint on the scale was labeled "My opinion on placing creationism on an equal footing with evolution in high school science classrooms has not changed." The remaining points were labeled with statements indicating increasing change, ranging from "I am slightly more in favor..." (+1) to "I am much more in favor...." (+6). (c) Participants then wrote a second essay stating and justifying their opinion on whether creationism should be placed on an equal footing with evolution in high school science classrooms. (d) Following the essay, participants completed a scale designed to directly assess opinions, identical to the scale used at pretest. (e) Participants then wrote a second essay stating and justifying their opinion on a different topic, capital punishment. This essay was included to assess transfer of improvements in argumentative reasoning to a topic beyond the activity. (f) Last, participants completed a scaled productivity and participation measure modeled after a measure previously used by Kim, Derry, Steinkuehler, Street, and Watson (2000). Using this measure, participants indicated on five–point Likert scales their perception of the usefulness of the activity, the importance of the activity topic, and the facility of the technology. Additional five–point scales assessed the degree of effort each participant felt they had contributed, the extent to which they believed in evolution and in creationism, and whether they perceived their position on the issue of creationism versus evolution as representing a minority or majority opinion.

Assessment of Argumentative Reasoning

Segmentation. Two coders, "blind" to the pretest/posttest status of each essay, segmented both the topic essays on whether creationism should be placed on an equal footing with evolution in the high school classroom (180 essays total) and the transfer essays on capital punishment (90 total). Each essay was segmented into simple sentences or independent clauses. Percentage agreement for segmentation on the topic essays was 88%. Percentage agreement for segmentation on the transfer essays was 90%. Disagreements were resolved by taking the larger number of segments for the given essay.

Analytic Scheme. In order to assess changes in the quality of argument from pre– to posttest essay as well as the transfer of argument ability, we developed two coding schemes, both modeled after one used by Kuhn, Shaw and Felton (1997). Each coding scheme was topic–dependent. The first coding scheme was designed to assess arguments regarding whether the topic of whether creationism should be placed on an equal footing with evolution in high school science classroom; the second coding scheme was designed to assess arguments regarding the use of capital punishment. The two schemes, however, were structurally and functionally quite similar.

Both schemes enabled us to organize statements made within the essays into three broad categories:

(a) Functional arguments (Type I) address the purpose or function of the action under consideration (placing creationism within the high school science curricula or using capital punishment). For example, one participant made the following functional argument, stating that the purpose of teaching creationism equally with evolution in the high school science classroom is to promote critical thinking: "Teaching creationism alongside of evolution in science classrooms fosters critical and analytical thinking in students."

(b) Nonfunctional arguments (Type II) address the conditions under which the action is justified or circumstantial, administrative problems that could be remedied with no consideration of the purpose or function of the action under consideration. For example, one participant made the following nonfunctional argument addressing an administrative problem with teaching creationism with no consideration of the purpose of doing so: "Our schools possess neither the resources nor the teachers to offer the Judeo-Christian viewpoint of creationism."

(c) Nonjustificatory arguments (Type III) are statements that have little or no argumentative force. For example, one participant made the following nonjustificatory argument, which is an appeal to precedent: "Evolution is always discussed in science classes."

Within the category of functional arguments, a second broad distinction was made between statements that address alternatives (Type IA) to the action under consideration and statements that do not (Type IB–IJ). For example, one participant made the following statement addressing an alternative to teaching both creationism and evolution, namely, teaching only evolution: "By not including creationism, it leaves children believing that modern science is the only answer to how life was created, and science is only one of several beliefs of how life on Earth and this universe began."

Insert Table 1 about here.

Table 1 presents a summary of the codes applied to the creationism/evolution essays (for a complete description of both coding schemes, see Table B1 and B2 in the Appendix). Both schemes were hierarchical: nonfunctional and functional arguments (Type I and II) were considered more advanced than nonjustificatory ones (Type III); functional arguments (Type I) were considered more advanced than nonfunctional ones (Type II); and arguments addressing alternatives (Type IA) were considered more advanced than basic functional ones (Type IB–IJ). (For a complete explanation of the rationale behind this design, see Kuhn, Shaw and Felton, 1997.)

In addition to assessing the argument elements participants included using the hierarchy of categories described above, we also assessed the frequency with which each participant made (a) statements containing evidence or recognizing its relevance and (b) metacognitive statements indicating a greater self-awareness or concern with one's own thinking processes. The following is an example of a statement that was coded as both (a) recognizing the relevance of evidence and (b) metacognitive: "Perhaps, I would have been swayed if I was presented with some of the evidence that they [i.e., the creationists] claimed to have collected." Statements that did not serve any argumentative function such as repetitions or elaborations of statements previously made were coded "null."

Coding Procedure. Kuhn, Shaw and Felton's (1997) coding scheme, which we used as a model for our own two coding schemes, was designed to assess arguments about capital punishment. Because their coding scheme could be readily adapted to our transfer essays addressing the same topic, we began our coding process with the transfer essays. Two coders, "blind" to the pretest/posttest status of each essay, analyzed the capital punishment data corpus. First, a subset of ten essays was coded jointly in order to insure that both coders were interpreting the analytic scheme in the same manner. Both coders then coded the remaining 80 essays independently using the capital punishment coding scheme in Table B2 of the Appendix. Adaptations to the coding scheme were negotiated throughout the coding process.

Interrater reliability was calculated per specific major coding category since statements within each major coding category are treated as equivalent. For example, statements categorized under "appropriate punishment" (Type I.C1—Type I.C3 in Table B2 of the Appendix) are treated as equivalent; therefore, for each essay, we compared the number of statements each coder assigned to the given category. Interrater reliability, calculated over the entire remaining set of 80 essays, was 0.90 (Pearson correlation). Disagreements were resolved through discussion.

By coding the transfer essays first, we were able to establish the expertise necessary for developing a second, similar scheme for creationism/evolution topic. Once coding of the transfer essays was completed, one coder, the primary researcher, developed a coding scheme for the creationism/evolution essays parallel to the coding scheme used for the capital punishment essays. First, one half of the creationism/evolution data corpus was randomly selected as the basis on which the coding scheme was developed. Through several iterations, working inductively from the essays to the a priori categorical structure discussed above (i.e., the categories functional, nonfunctional, nonjustificatory, and alternative statements), a coding scheme was developed, refined, and negotiated through discussion with the other authors. Once a stable and sufficient coding scheme was reached, the codes were then applied to the remaining 90 essays.

Interrater reliability was calculated per specific major coding category as described above. For example, statements categorized under "promotes critical thinking" (Type C—Type C5 in Table B1 of the Appendix) are treated as equivalent; therefore, for each essay, we compared the number of statements each coder assigned to the given category. Interrater reliability, calculated over a subset of 30 essays randomly selected from the data corpus, was 0.82 (Pearson correlation). The primary researcher's codes were used when discrepancies occurred.

## Results

Argument change from pretest to posttest. Comparison of each participant's pre- and posttest essays allowed us to identify changes that occurred in the quality of each argument and to compare the effects of each condition on argument skill. Argument improvement was defined as: (a) a decrease in the proportion of nonjustificatory statements made, (b) a decrease in the proportion of nonfunctional statements made, (c) an increase in the number of statements addressing alternatives to the action under consideration; (d) an increase in the use (or acknowledgment of the relevance) of evidence, (e) an increase in the number of metacognitive statements, and finally (f) an increase in the range of argument (the number of distinct warranting claims made).

---

Insert Table 2 about here.

Table 2 presents the mean change in the quality of argument from pretest to posttest for each condition. Across conditions, there was a statistical decrease from pretest to posttest essay in the proportion of nonjustificatory statements made, with no statistical differences among the three conditions. There was no decrease in the proportion of nonfunctional statements made, either across or among conditions.

There was, however, an across–conditions statistical increase in the number of statements made that contained evidence, with Fisher LSD comparisons revealing that essays from the summarization condition contained more evidence than did those in the other two conditions. Further analysis revealed that this increase in evidence statements made by participants in the summarization condition was limited to statements containing evidence originally presented within the pro/con text that participants read.

Across conditions, there was a statistical increase in the number of metacognitive statements made, with no differences among conditions. No increases in either the range of argument or in the number of statements addressing alternatives were detected, either across or among conditions.

Transfer of argument skills. The quality of arguments regarding capital punishment was assessed using the same variables described above. We then assessed the relationship between the quality of arguments contained within the posttest essays on capital punishment and the quality of arguments contained within the creationism/evolution posttest essays in order to determine whether abilities in argumentative reasoning demonstrated at posttest were limited to the topic discussed or whether such abilities transferred to a second topic.

Both the range of argument presented and the use of evidence in the creationism/evolution essay and the capital punishment essay were statistically correlated to a modest degree (pooled within–conditions correlations of .41 and .32, respectively).

Recall test. The short–answer questionnaire designed to assess recall for the given pro/con text consisted of 11 short–answer questions totaling 18 points possible. The first two questions required recall of six arguments made within the text (three pro arguments, three con arguments). The remaining five questions assessed recall for selected details of the text. Scoring was weighted such that the first two questions, requiring recall of more substantial material than only detail, comprised two-thirds of the total points possible. Two raters, "blind" to the condition status of each recall test, scored all recall tests. Percentage agreement, calculated by comparing the number of points each coder awarded each response, was 92%. Disagreements in assessment were resolved through discussion.

---

Insert Figure 5 about here.

---

Figure 5 presents the mean recall of the text for each condition. Means (and standard deviations) for the discussion, self-explanation, and summarization conditions were 6.00 (2.31), 8.60 (1.58) and 11.30 (1.01), respectively. The difference among the three conditions was statistically significant, $p < .001$, with Fisher LSD comparisons revealing that all three condition means differed statistically from one another.

Reported and Directly Assessed Opinion Change. Table 3 presents the mean absolute value of both directly assessed and reported opinion change for each condition.

---

Insert Table 3 about here.

---

The three conditions differed statistically in the magnitude of both students' directly assessed and their self-reported absolute opinion change, $F(2, 42) = 13.84$, $p < .001$, and $F(2, 42) = 4.46$, $p < .025$, respectively. Fisher LSD comparisons revealed that: (1) on the former measure, students in the self-explanation and summarization conditions exhibited more opinion change than did students in the discussion condition; and yet (2) ironically, on the latter measure, students in the discussion condition reported more opinion change than did those in the self-explanation and summarization conditions. Further individual–level analyses revealed a statistically significant pooled–within conditions negative relationship between actual and reported opinion change. The relationship was strong in the summarization condition, moderate in the self–explanation condition, and negligible in the discussion condition.

Perceptions of Productivity and Participation Principal components analysis of the questionnaire measuring perceptions of productivity and participation revealed three main components. These components were named "perceived importance of topic," "perceived ease of technology," and "perceived personal contribution." Cronbach's alpha coefficient for each component was .77, .79, and .60, respectively.

---

Insert Table 4 about here.

---

Table 4 shows the means and standard deviations of each condition for the three main components. There were no significant differences in either perceived importance of topic, $F < 1$, or perceived ease of technology, $F(2, 42) = 1.06$, $p > .05$. However, there was a significant difference among conditions in perceived personal contribution, $F(2, 42) = 4.38$, $p < .05$, with Fisher LSD comparisons revealing differences between the discussion condition and each of the other two conditions, self–explanation and summarization. In addition, individual–level analyses revealed a small but statistically significant pooled within–conditions relationship between perceived personal contribution and actual opinion change (but not between perceived personal contribution and reported opinion change).

## Discussion

Participants demonstrated improvements in argumentative reasoning in terms of a shift toward arguments containing more justificatory statements than nonjustificatory ones, clearly one of the most basic requirements for constructing an argument. Participants also demonstrated an increasing awareness of, and concern about, the consistency and quality of their thinking, evidenced by participants' increase in metacognitive statements. These improvements indicate that mere engagement in thinking about a topic may well be the underlying causal mechanism of such improvements, as Kuhn, Shaw, and Felton (1997) suggest.

There is no evidence, however, that participants demonstrated improvements in argumentative reasoning in terms of a shift toward functional rather than nonfunctional statements. Although nonfunctional statements do contribute to an argument to some extent, arguments focused on conditions rather than the function or purpose of the action itself are inherently fairly weak. The use of such arguments indicates a failure to differentiate between the conditions under which an action may be justified and the purpose or function that might warrant the action in the first place. In addition, across conditions, there were no increases in the number of statements addressing alternatives to the action under consideration, indicating that participants did not improve in their ability to weigh the value of the action under consideration against its alternatives. Finally, across conditions, there were no increases in range of arguments made.

Of the six variables used to investigate improvement in argumentative reasoning, only one variable differentiated among conditions—the number of statements made that contained or acknowledged the relevance of evidence. In particular, summarization produced greater use of evidence than did self–explanation and discussion. Further analyses revealed that this increase was limited to statements containing evidence originally presented within the pro/con text. Summarization also produced greater levels of text recall than did self–explanation, which in turn was more effective than online discussion. In other words, individuals who could recall the text then leveraged the evidence it contained as support within their own argument.

The question that arises, then, is whether what participants in the summarization condition gained was merely information for the pro/con text. Analysis of data from the transfer task, however, indicates that this may not be the case. Use of evidence within the creationism/evolution essay correlated with use of evidence in the second, transfer essay on capital punishment. Although recall for the creationism/evolution pro/con text may account for participants' use of evidence in an essay on that topic, it cannot account for participants' use of evidence in an essay on a second, separate topic. Recall of information in the pro/con text on the issue of whether creationism should be placed on an equal footing with evolution in the high school science classroom is of little assistance in constructing an argument on the issue of capital punishment. What participants in the summarization condition gained may be more than simply recall of given information; rather, those participants appear to have gained a greater propensity for using evidence to support an argument more generally.

These results are interesting given that the conditions did not differ in other ways that were expected to favor online discussion. Participants in all three conditions changed their opinions on the given issue from pretest to posttest. Though participants in the discussion condition reported greater opinion change, participants in the self–explanation or summarization conditions exhibited greater actual opinion change. Individuals who engaged in discussion did, however, more accurately assess the degree to which their position had shifted than individuals who engaged in self–explanation; in turn, individuals who engaged in self–explanation demonstrated greater accuracy than individuals who engaged in summarization.

Such findings have import for our results regarding argument improvement: Though participants in all three conditions demonstrated an increasing concern about the consistency and quality of their thinking, as evidenced by the increase in metacognitive statements from pretest to posttest essay in all three conditions, only participants in the discussion condition demonstrate any accuracy on the definitively "metacognitive" task of assessing their opinion change. In other words, though participants in the other two conditions may show a similar level of concern about their reasoning, they do not demonstrate a similar level of accuracy. Research investigating opinion change as a result of exposure to mixed evidence (e.g., Kuhn & Lao, 1996; Miller, McHoskey, Bane, & Dowd, 1993) indicates that individuals are, in general, rarely accurate in their assessment of the degree to which their position has moved. Given our results, it is plausible that discussion is one intervention that might mediate such inaccuracy in report. Such findings warrant further research.

Individuals who engaged in discussion also reported a greater personal contribution to the activity in general than individuals engaged in self-explanation or summarization. The difference in perceived contribution intuitively makes sense—in general, individuals engaged in discussion feel that they contribute more personal effort and ideas to the activity than individuals engaged in self–explanation or summarizing and studying a text. This difference, however, is relatively small. The small but positive relationship between perceived personal contribution and actual opinion change indicates that, to some extent, individuals who exhibit belief change also perceive a greater personal investment of effort and ideas to the activity. Perhaps belief change exacts a cognitive price, resulting in greater perceptions of personal expenditure in individuals who revise their opinions than individuals who do not. Substantiating this claim, however, requires further research.

## Qualitative analysis of the transcripts

Numerous lines of research support the notion that discussion may be a particularly well–suited activity for improving argument skills. Our results, however, do not show that discussion is more productive than self–explanation or summarization, two individualized study techniques derived from the cognitive memory literature, in terms of improving argumentative reasoning, increasing recall for the text, or fostering opinion change. An important question to ask, then, is what particular kind of discussion activity is actually being compared. In order to investigate what activities actually occurred in the discussion condition, we conducted a qualitative analysis of the transcripts generated from the online environment.

Time on task. Were there notable differences in the way individuals in the discussion condition used their time during the 45–minute activity compared to individuals in the self-explanation or summarization conditions? In other words, could one explanation of our results be that participants in discussion spent less time on topic, resulting in fewer learning gains? In order to investigate this possible explanation, we compared the average number of (a) data–entry turns, defined as the number of distinct times an individual entered a comment into the system regardless of the comment's length, (b) words total, (c) words per data–entry turn, and (d) system errors, resulting from omission or misuse of the command required to enter text, that participants in each condition made. Table 5 presents the means and standard deviations of each variable for each condition.

Insert Table 5 about here.

Although participants in the discussion condition made more data–entry turns than participants in the other two conditions, they contributed roughly the same number of words as participants in the self–explanation condition. Although participants in the summarization condition made the greatest learning gains, they actually contributed only half (roughly 55%) the total number of words compared to the other two conditions. In addition, participants in the discussion condition made more system errors on average than participants in the self–explanation condition, who in turn made slightly more errors on average than participants in the summarization condition. Average number of turns made, number of words contributed, and length of comment (average number of words per turn) do not clearly correlate with the greater learning gains in the summarization condition. Participants in the summarization condition took fewer turns and contributed fewer words overall than the other two conditions. The average length of comments made by participants in the summarization condition is roughly the same as the average length of comments made by participants in the discussion condition. Only the number of system errors follows a pattern we might expect to explain the greater learning gains achieved by participants in the summarization condition; however, the average number of errors made in each condition is fairly small and the difference between conditions is substantially less than one. Further research is required to investigate whether such small differences in the average number of system errors made effects the overall learning gains in each condition.

In each condition, we also coded all data–entry turns that were recognizably off–task using the following coding categories: (a) housekeeping statements, including statements about what should be done next, remarks about how to use the online system, and clarifications of spelling or typing errors; (b) social talk, including salutations, talk about the weather, academic majors, etc. (c) tangent topic, including statements on related but peripheral topics such as the pope; and (d) "null" statements containing no discernable content. Tables 6 presents the percentage of the total number of words in each condition that were classified in the four categories listed above.

---

Insert Table 6 about here.

---

Participants in the discussion condition spent more time making housekeeping statements, talking socially, and discussing tangent topics than participants in the other two conditions. The majority of off–topic content appears to consist of social talk; one pair of participants in particular, however, accounts for 5.5% of the total 8.7% of social talk, leaving roughly 3.2% of the rest of the discussion transcripts as social talk. Participants in the discussion condition, therefore, generally spent the same about of time making housekeeping statements, discussing tangent topics, and interacting socially. To some extent, both housekeeping statements and social interaction are necessary to keeping the general conversation on task, to insure that participants understand one another, and to maintain an amiable conversation. Discussion of tangent topics, however, does not seem to serve a similar functional role. To some extent, then, participants in discussion did spend less time "on–task." The extent to which this may contribute to the lack of learning gains in the discussion condition warrants further consideration.

Effects of pair agreement or disagreement. Was discussion between pairs that disagreed more productive than discussion between pairs who agreed? In order to investigate whether pair agreement or disagreement effected improvement in argumentative reasoning, we classified each individual who participated in the discussion activity as "pro" creationism in the high school science curricula or "con" creationism in the high school science curricula based on their response to the pretest opinion scale. Of those fifteen pairs, eight held congruent opinions (three pairs consisted of two "pro" individuals, five pairs consisted of two "con" individuals) and seven held discrepant opinions (seven pairs consisted on one "pro" individual and one "con" individual). There were no significant effects of whether the pair agreed or disagreed on argument improvement. In addition, of the eight discussion pairs who failed to successfully complete a joint statement on the topic, only four consisted of pairs who disagreed. In other words, agreeing pairs were as likely to fail to conclude the discussion with a formal joint statement as disagreeing pairs. These results indicate that pair agreement may not be the most important variable to consider in the design of future learning environments that incorporate online discussion.

Unequal participation within the pairs. Did individuals within in each pair contribute equally, or did one individual contribute more the conversation, in effect "dominating" the interaction? What were the effects of unequal participation on change in argumentative reasoning? We compared the number of data–entry turns and number of words contributed by each member of the discussion pairs. In three of the 15 discussion pairs, one individual contributed at least 1.5 times the number of both turns and words than their partner. We then compared the change in argumentative reasoning demonstrated by the two individuals within each of the three pairs. In each pair, the individual who "dominated" the conversation demonstrated an improvement in argumentative reasoning in at least one more of the six variables total examined than their partner. In addition, the individual who "dominated" also demonstrated a decline in argumentative reasoning in at least one less of the six variables than their partner did. These results indicate that, in conversations in which one person contributes more than the other, the individual who "dominates" gains more from the conversation than the individual who does not "dominate." Though these results are suggestive, further research with more subjects in a controlled experiment is necessary in order to substantiate these results.

The results of this analysis and future studies should inform the design of innovative learning environments that incorporate new forms of communication and collaboration made available through computer technologies. Exploring what types of experiences foster reasoned argumentation is critical for instruction that aims to improve reasoning on both the group and individual level. However, it is important to know what tradeoffs may occur. Results from our study suggest that relatively unguided online discussion could be no more effective, perhaps even less effective, in terms of improving argumentative reasoning, increasing recall for a given text, or fostering belief–change than other cognitively–based activities. Participants engaged in discussion appear to spend less time actually engaged on the topic; however, further analysis is necessary if we are to gain insight into what other particular characteristics of discussion might effect argumentative reasoning. The analysis described above is an early step toward understanding the particular characteristics of the discussion process more in–depth and comparing the effects of discussion as a learning activity with other, individualized and cognitively–based ones.

## Acknowledgments

This research was conducted through SRI International with funding from the National Science Foundation (Grant No. REC-9725528) and the National Institute for Science Education, which is a partnership of the University of Wisconsin-Madison and the National Center for Improving Science Education, Washington, DC, with funding from the National Science Foundation (Cooperative Agreement No. RED-9452971). However, the ideas expressed herein are not endorsed by and may not be representative of positions endorsed by the sponsoring agencies. The authors are grateful to Leonard Abbeduto for his helpful comments on the design and implementation of this experiment, and to Kimberly Bauer for her assistance in collecting the data.

References

Brown, A. L., & Palincsar, A. S. (1989). Guided cooperative learning and individual knowledge acquisition. In L. B. Resnick (Ed.), Knowing, learning and instruction. Essays in honor of Robert Glaser. (pp. 361-392). Hillsdale, NJ: Erlbaum.

Chi, M. T. H., deLeeuw, N., Chiu, M.-H., & LaVancher, C. (1994). Eliciting self-explanations improves understanding. Cognitive Science, 18, 439-477.

Kim, J. B., Derry, S. J., Steinkuehler, C. A., Street, J. P., & Watson, J. G. (2000). Online based collaborative learning. Paper presented at the Annual Meeting of American Educational Research Association, New Orleans, LA.

Kuhn, D. (1991). The skills of argument. Cambridge: Cambridge University press.

Kuhn, D., & Lao, J. (1996). Effects of evidence on attitude: Is polarization the norm? Psychological Science, 7, 115-120.

Kuhn, D., Shaw, V., & Felton, M. (1997). Effects of dyadic interaction on argumentative reasoning. Cognition and Instruction, 15, 287-315.

Miller, A. G., McHoskey, J. W., Bane, C. M., & Dowd, T. G. (1993). The attitude polarization phenomenon: Role of response measure, attitude extremity, and behavioral consequences of reported attitude change. Journal of Personality and Social Psychology, 64, 561-574.

O'Connor, M C. & Michaels, S. (1996). Shifting participant frameworks: Orchestrating thinking practices in group discussion. In D. Hicks (Ed.), Discourse, learning, and schooling. Cambridge: Cambridge University Press.

Piaget, J. (1976). Postscript. Archives de Psychologie, 44, 223-228.

Pontecorvo, C. & Girardet, H. (1993). Arguing and reasoning in understanding historical topics. Cognition and Instruction, 11, 365-395.

Sproull, L., & Faraj, S. (1997). Atheism, sex, and databases: The net as social technology. In s. Kiesler (Ed.), Culture of the internet. (pp. 35-52). Mahwah, NJ: Erlbaum.

Vygotsky, L. S. (1978). Mind in society: The development of higher psychological processes (V. John-Steiner, E. Souberman, M. Cole, & S. Scribner, Eds. and Trans.), Cambridge, MA: Harvard University Press.

Warren, B. & Rosebery, A. S. (1986). "This question is just too, too easy!" Students' perspectives on accountability in science. In L. Schauble & R. Glaser (Eds.), Innovations in learning: New environments for education. (pp. 97-126). Mahwah, NJ: Erlbaum.

## APPENDIX A

### Pro Argument

Creationism should be given an equal footing with evolution in high school science instruction. Evolution is only one theory among many that attempts to explain the origins of man; however, current textbooks and instruction censor scientific information that runs counter to evolution and, as such, misrepresents the theory of evolution as accepted fact. Since creationism is the most widely accepted alternative theory to evolution and the majority of Americans are Christian, equal instruction in creationism is appropriate and fitting. Moreover, creationism is a viable scientific hypothesis, since evidence is used in support of the claim. For example, members of the Institute for Creation Research have found scientific evidence to support creationist theories, though their work has remained unpublished in established scientific journals. Our constitutional right to freedom of religion insists that students be allowed the opportunity to learn about the scientific views of their own beliefs and not forced to take classes that do not represent these views. Creation by intelligent design is an equally reasonable explanation for the origin of the complexity we see in living things. Arguments can and have been made against the theory of evolution; for example, it has been argued that random change over time could not possibly transform simple systems into more complex ones since the second law of thermodynamics states that, with time, everything in the universe tends to undergo progressive degradation. By teaching creationism as well as evolution, students learn the dominant contrasting arguments about man's origins and are able to make their own critical evaluations rather than being restricted into consideration of only that theory which public schools have deemed worthy of representation.

### Con Argument

Efforts to place creationism on an equal footing with evolution in high school science instruction should be opposed. Including creationism in the curriculum of public schools crosses the line of separation between church and state. The US Supreme Court acknowledged this when it ruled against Louisiana in 1987 in the Edwards v. Aguillard case, stating that creationism is inherently religious and cannot constitutionally be presented as a scientific explanation of origins in public schools. Furthermore, science and religion address different questions in fundamentally different ways; therefore, the presentation of creationism, a religious concept, within the science curriculum is inappropriate. Science yields testable hypotheses that must be confirmed by evidence, and its theories are always subject to revision or replacement. In contrast, religion is ultimately based upon faith and provides explanations that cannot be challenged by observable evidence. Again, the court system has acknowledged this—when the state of Arkansas attempted to place creationism in the science classroom, the court reviewed the definition of science and determined that creationism is not a science (McLean v. Arkansas Board of Education, 1982). The exclusion of religious explanations from the science class does not amount to telling students that they should not maintain those beliefs—only that those beliefs are not acceptable as science. Years of scientific studies have provided the most acceptable explanations of the origin and development of life on the earth. The theory of evolution has the general consensus of the scientific community because it integrates and clarifies many otherwise isolated scientific facts, principles and concepts in a manner which is consistent with known evidence. Furthermore, teaching Judeo-Christian views on man's origins fails to represent the diversity of religious beliefs currently held in America, and, given the limited time and resources endemic to education, such exclusion of minority religions would remain unable to be rectified.

APPENDIX B

TABLE B1

Coding scheme for arguments regarding whether creationism should be placed on an equal footing with evolution in high school science classrooms

Pro Arguments

I. Functional arguments

A. Alternatives to C=E are ineffective or less effective than C=E.

A1. ONLY–E misrepresents E as fact not theory.

A2. ONLY–E censors information against E (or for C).

A3. ONLY–E does not promote critical thinking.

A4. ONLY–E does not allow students to decide.

A5. ONLY–E fails to teach students about an important part of cultural history.

A6. ONLY–E does not foster tolerance.

A7. ONLY–E is against freedom of religion.

A8. ONLY–E is anti-religious.

A9. ONLY–E lowers morality.

A10. ONLY–C is not adequate science instruction.

B. C=E is appropriate science instruction.

B1. C is scientific.

B2. C is a theory or explanation of origins.

B3. There is evidence for C.

B4. C has been proven.

B5. Many believe C.

B6. E is not scientific.

B7. E is only a viewpoint or belief

B8. There is no evidence for E.

B9. There is evidence against E.

B10. E may be revised or replaced E.

B11. E is only a theory, not a proven fact.

B12. Many do not believe E.

B13. C & E can be combined or synthesized into one theory.

B14. No theory of human origins can be proven.

B15. No theories should ever be ignored.

B16. C=E presents E as one theory among many.

B17. C=E teaches students that scientific theories are based on evidence.

B18. C=E teaches students that theories can be revised.

B19. C=E allows students to engage in scientific practice.

B20. C=E integrates science with other disciplines.

C. C=E promotes critical thinking.

C1. C=E allows students to consider alternative or contrasting explanations.

C2. C=E challenges students to understand different views.

C3. C=E teaches students to challenge beliefs or theories.

C4. C=E allows students to compare and contrast alternative theories.

C5. C=E allows students to construct their own theories..

D. C=E allows students to decide.

D1. C=E presents both sides of the debate

E. C=E increases students' knowledge.

F. C=E prepares students for society (culture).

F1. C=E teaches students about our cultural history.

F2. C=E informs students about the diversity of beliefs within our culture.

F3. C=E fosters tolerance.

F4. C=E exposes students to an important social debate.

G. C=E creates a positive classroom atmosphere.

G1. C=E makes students feel that their views are considered important.

G2. C=E fosters classroom discussion.

G3. C=E interests or engages students.

H. C=E allows freedom or religion or beliefs.

H1. C=E allows students to study their own beliefs.

H2. C=E forces all students to learn contrary beliefs (rather than just one group).

H3. Many people believe C.

H4. Not everyone believes E.

I. C=E does not violate the separation of church and state.

I1. C=E does not force religious beliefs on students.

I2. The teacher can present both sides objectively (only the facts).

I3. C can be presented as a scientific theory, not as a religious belief.

I4. E already brings religion into the classroom.

I5. Students are not forced to take the class.

J. C=E is necessary.

J1.C is not taught in other classes.

J2. Non–C students are not exposed to C elsewhere.

II. Nonfunctional arguments (focused on conditions that make C=E justified without consideration of its functions)

A. C=E is justified if multiple theories of origin are taught.

B. C=E is justified only if C is presented as a religious belief, not as a scientific theory.

C. C=E is justified only if C is presented as a scientific theory, not a religious belief.

D. C=E is justified only if the teacher emphasizes that C and E both are theories, not facts.

E. C=E is justified only if the teacher justifies why C & E are the only theories being taught.

F. C=E is justified only if the teacher emphasizes the difference between science and religion.

G. C=E is justified only if it is taught in an unbiased manner.

H.C=E is justified only if taught in non–science classes.

I. C=E is justified only if it is offered as an elective course or lesson.

J. C=E is justified only if it is taught in settings outside public schools.

K. C=E is justified only if the parents give their permission.

L. C=E is justified only if the students request it.

III. Nonjustificatory arguments

A. Justification based on sentiment.

B. Appeal to precedent (C=E has been done for a long time).

C. Appeal to majority (many or most think C=E is a good idea).

D. Appeal to authority (without intervening argument).

E. I believe in C.

F. I believe in both C & E.

G. I do not believe in E.

Con Arguments

I. Functional arguments

A. Alternative exist that are preferable to C=E.

A1. ONLY–E presents the best scientific evidence we have to students.

A2. ONLY–E teaches students what is currently accepted by the science community.

A3. ONLY–E teaches students that religious beliefs are not science.

A4. ONLY–E allows students to decide.

A5. ONLY–E prepares students for modern society.

A6. ONLY E allows freedom of religion.

A7. ONLY E is not anti–religious.

A8. ONLY–E keeps church and state separate.

A9. C<E informs students that E is a theory, not a fact.

A10. C<E exposes students to both sides of the debate.

A11. C<E teaches students that scientific theories are based on evidence.

A12. C<E teaches students that religious beliefs are not science.

A13. C<E promotes critical thinking.

A14. C<E allows students to decide.

A15. C<E informs students about cultural diversity.

A16. C<E fosters tolerance.

A17. C<E allows freedom of religion.

A18. C<E keeps church and state separate.

B. C=E is not appropriate science instruction.

B1. C is not scientific.

B2. C is a religious.

B3. C is based on faith.

B4. C is only a viewpoint or belief.

B5. There is no evidence for C.

B6. There is evidence against C.

B7. The evidence for C has not been published in any scientific journals.

B8. C cannot be revised or replaced.

B9..C is only a theory, not a proven fact.

B10. C does not have the consensus of the scientific community.

B11. Many do not believe C.

B12. E is scientific.

B13. E integrates other scientific facts, principles and concepts.

B14. E is not religious.

B15. E is a theory or explanation of origins.

B16. E is not just a viewpoint or belief

B17. There is evidence for E.

B18. E can be revised or replaced.

B19. E has been proven.

B20. E has the consensus of the science community.

B21. Many believe E.

B22. C=E confounds science and religion.

B23. C=E misrepresents C as fact not theory.

B24. C=E hinders the advancement of science.

C. C=E deters critical thinking.

D. C=E is against freedom of religion.

D1. C=E does not allow non–C students to study their own beliefs.

D2. C=E forces non–C students to learn contrary beliefs.

D3. C=E fails to equally represent all religions in instruction.

D4. C=E discriminates against non–C students.

D5. C=E portrays C as unsound.

D6. Not everyone believes C or is religious.

D7. C is religious.

E. C=E violates the separation of church and state.

E1. C=E forces religious beliefs on students.

E2. C is religious.

E3. Public schools are state institutions .

F. C=E is unnecessary.

F1. C is taught outside public schools.

F2. C & E are not opposite sides of a debate.

II. Nonfunctional arguments

A. C=E Nonfunctional arguments (focused on possibly remediable defects in administration of C=E without consideration of its function)

A1. People may object to C=E.

A2. There may be lawsuits against the schools.

A3. Teacher may not present C=E objectively.

A4. Teachers may lack expertise in C.

A5. Teachers may be uncomfortable with C.

A6. C=E may promote more religion in schools.

A7. C=E is too burdensome.

B. C<E Nonfunctional arguments (focused on conditions that make C<E justified without consideration of its functions.)

B1. C<E is justified if multiple theories of origin are mentioned.

B2. C<E is justified only if C presented as a religious belief, not a scientific theory.

B3. C<E is justified only if C is discussed generally (not tied to one specific religion).

B4. C<E is justified only if the teacher emphasizes that E is a theory, not a proven fact.

B5. C<E is justified if the teacher explains or justifies why E is being emphasized.

B6. C<E is justified only if both C and E are taught in an unbiased manner.

B7. C<E is justified only if the topics are presented in an open discussion.

B8. C<E is justified only if the students request it.

C. ONLY–E Nonfunctional (focused on conditions that make ONLY–E justified without consideration of its function)

C1. ONLY–E is justified only if the arguments or evidence both for and against E is presented.

C2. ONLY–E is justified only if the teacher emphasizes that E is a theory, not a fact.

C3. ONLY–E is justified only if the lessons on E are optional.

III. Nonjustificatory arguments

A. Justification based on sentiment.

B. Appeal to precedent (C=E has not been done very widely or as widely as it once was).

C. Appeal to majority (many or most are against C=E).

D. Appeal to authority (without intervening argument).

E. I do not believe in C.

F. I believe in E.

---

Note. C = creationism, E = evolution, 'C=E' = teaching creationism on an equal footing with evolution in high school science classrooms, ONLY–E = teaching only evolution in high school science classrooms, ONLY–C = teaching only creationism in high school science classrooms, and C<E = mentioning creationism but focusing on evolution in high school science classrooms.

TABLE B2

Coding scheme for arguments regarding capital punishment

Pro Arguments

I. Functional arguments

A. Alternatives to CP are ineffective or less effective than CP.

A1. Alternatives to CP are not effective as deterrents.

A2. Alternatives to CP are not effective in protecting society from criminals.

A3. Alternatives to CP are not sufficient punishment.

A4. Alternatives to CP fail to rehabilitate criminals.

A5. Alternatives to CP are too burdensome or costly a way to serve their purpose.

A6. Alternatives to CP waste lives.

A7. Alternatives to CP are less humane than CP.

B. CP reduces crime.

B1. CP deters people from crime.

B2. CP protects society from the acts of criminals.

C. CP is an appropriate punishment.

C1. An eye for an eye.

C2. Criminals have forfeited the right to life and privileges associated with it.

C3. Compensates victim or victim's family.

D. CP controls prison population.

II. Nonfunctional arguments (focused on conditions that make CP justified without consideration of its functions)

A. CP is justified only if guilt is established beyond reasonable doubt.

B. CP is justified only if a criminal is judged competent to be responsible for his or her own actions.

C. CP is justified only if it is applied consistently.

D. CP is justified only if the crime is sufficiently grave.

E. CP is justified only in the case of repeated crime.

F. CP is justified only if the crime was committed intentionally.

G. CP is justified only if the criminal is beyond rehabilitation.

H. CP is justified only if the criminal shows no remorse.

III. Nonjustificatory arguments

A. Justification based on sentiment.

B. Appeal to precedent (CP has been in use for a long time).

C. Appeal to majority (many or most think it is a good idea).

D. Appeal to authority (without intervening argument).

E. Crime exists and needs a remedy.

Con Arguments

I. Functional arguments

A. Alternatives exist that are preferable to CP.

A1. Alternatives to CP are better as deterrents.

A2. Alternatives to CP are better in protecting society from criminals.

A3. Alternatives to CP are better punishment.

A4. Alternatives to CP allow rehabilitation of criminals.

A5. Alternatives to CP are more efficient than CP.

B. CP does not reduce crime or reduce it sufficiently.

B1. CP is not effective in deterring people from crime.

B2. CP is not effective in protecting society from the acts of criminals.

C. CP is not an appropriate punishment.

C1. CP commits the same crime it is meant to punish.

C2. CP does not right the wrong (does not restore loss to victim of crime).

C3. We lack the right to take life.

C4. We lack the right to make judgments of who should live or die.

C5. We lack the right to make judgments of other people's actions.

C6. CP violates the principle of forgiveness.

C7. Any killing is wrong.

C8. CP is violent and barbaric.

C9. CP serves no purpose.

C10. CP is a release or escape from punishment.

C11. Two wrongs don't make a right.

C12. CP does not give criminals the chance to reform.

II. Nonfunctional arguments (focused on possibly remediable defects in administration of CP without consideration of its functions)

A. CP may punish innocent people.

B. CP may punish people who are not responsible for their actions.

C. CP is not administered uniformly (may be discriminatory against certain groups).

D. CP is not administered efficiently (e.g., may be drawn out and costly).

III. Nonjustificatory arguments

A. Justification based on sentiment.

B. Appeal to precedent (CP has not been widely used or is not as widely used as it once was).

C. Appeal to majority (many or most are against CP).

D. Appeal to authority (without intervening argument).

Table 1 Summary of the coding scheme used to assess the quality of arguments regarding whether creationism should be placed on an equal footing with evolution in high school science classrooms.

| Pro Arguments | Con Arguments |
|---|---|
| I. Functional arguments | I. Functional arguments |
| A. Alternatives to C=E are ineffective. | A. Alternatives exist that are preferable. |
| B. C=E is appropriate science instruction. | B. C=E not appropriate science instruction. |
| C. C=E promotes critical thinking. | C. C=E deters critical thinking. |
| D. C=E allows students to decide. | D. C=E is against freedom of religion. |
| E. C=E increases students' knowledge. | E. C=E violates the separation of church and state. |
| F. C=E prepares students for society. | F. C=E is unnecessary. |
| G. C=E creates a positive classroom atmosphere. | II. Nonfunctional arguments |
| H. C=E allows freedom or religion. | A. C=E Nonfunctional arguments |
| I. C=E does not violate the separation of church & state. | A1. People may object to C=E. |
| J. C=E is necessary. | B. C<E Nonfunctional arguments |
| II. Nonfunctional arguments | B1. C<E is justified if multiple theories of origin are mentioned. |
| A. C=E is justified only if multiple theories of origin are taught. | C. ONLY–E Nonfunctional |
| III. Nonjustificatory arguments | C1. ONLY–E is justified only if the arguments or evidence both for and against E is presented. |
| A. Justification based on sentiment. | III. Nonjustificatory arguments |
| | Justification based on sentiment. |

(a) Note. C = creationism, E = evolution, 'C=E' = teaching creationism on an equal footing with evolution in high school science classrooms, ONLY–E = teaching only evolution in high school science classrooms, ONLY–C = teaching only creationism in high school science classrooms, and C<E = mentioning creationism but focusing on evolution in high school science classrooms.

Table 2 Mean change (post – pre) in the quality of argument, by experimental condition (N = 15 pairs per condition)

| Measure | Condition | | |
|---|---|---|---|
| | Discussion | Self–Explanation | Summarization |
| Nonjustificatory statements (proportion)[a] | – 0.10 | – 0.01 | – 0.15 |
| Nonfunctional statements (proportion)[b] | 0.00 | 0.01 | 0.01 |
| Alternative statements[c] | – 0.03 | – 0.20 | – 0.53 |
| Evidence statements[d] | 0.03 | 0.03 | 0.47 |
| Metacognitive statements[e] | 0.47 | 0.77 | 0.43 |
| Range of Argument[f] | – 0.27 | 0.10 | 1.17 |

[a] Pooled within-conditions SD = 0.189
[b] Pooled within-conditions SD = 0.244
[c] Pooled within-conditions SD = 1.174
[d] Pooled within-conditions SD = 0.428
[e] Pooled within-conditions SD = 1.377
[f] Pooled within-conditions SD = 2.362

Table 3 Mean absolute amount of directly assessed or reported opinion change, by experimental condition (N = 15 pairs per condition)

| Measure | Condition: Discussion | Condition: Self–Explanation | Condition: Summarization |
|---|---|---|---|
| Direct Assessment[a] | 1.10 | 3.07 | 2.93 |
| Self–Report[b] | 1.70 | 0.67 | 0.87 |

[a] Pooled within-conditions SD = 1.144

[b] Pooled within-conditions SD = 1.005

Table 4 Means (and standard deviations) for perceived importance of topic, ease of technology, and personal contribution, by experimental condition (N = 15 pairs per condition).

| | Condition | | |
|---|---|---|---|
| | Discussion | Self-Explanation | Summarization |
| Perceived Importance of Topic (5 = very important) | 4.30 (.51) | 4.18 (.41) | 4.18 (.62) |
| Perceived Ease of Technology (5 = very easy) | 4.37 (.49) | 4.58 (.32) | 4.42 (.45) |
| Perceived Personal Contribution (5 = high contribution) | 4.45 (.41) | 4.17 (.39) | 4.08 (.24) |

Table 5 Mean (and standard deviation) data–entry turns, words total, words per turn, and system errors, by experimental condition.

| | Condition | | |
|---|---|---|---|
| Measure | Discussion | Self–Explanation | Summarization |
| Data–entry turns | 41 (18) | 29 (12) | 23 (6) |
| Words total | 757 (194) | 764 (237) | 423 (123) |
| Words/ Data–entry turns | 18 (6) | 26 (15) | 19 (7) |
| System errors | 1.6 (1.6) | 1.2 (1.4) | 1.0 (1.2) |

Table 6 The percentage of the total number words in each condition that were classified as housekeeping statements, social talk, tangent topics, and null statements.

| | Condition | | |
|---|---|---|---|
| | Discussion | Self–Explanation | Summarization |
| Housekeeping statements | 3.5 | Less than 0.1 | Less than 0.1 |
| Social talk | 8.7 | Less than 0.1 | 0.2 |
| Tangent topics | 3.3 | 0 | 0 |
| Null Statements | Less than 0.1 | Less than 0.1 | 0 |
| Total off–topic | 15.5 | 0.1 | 0.3 |

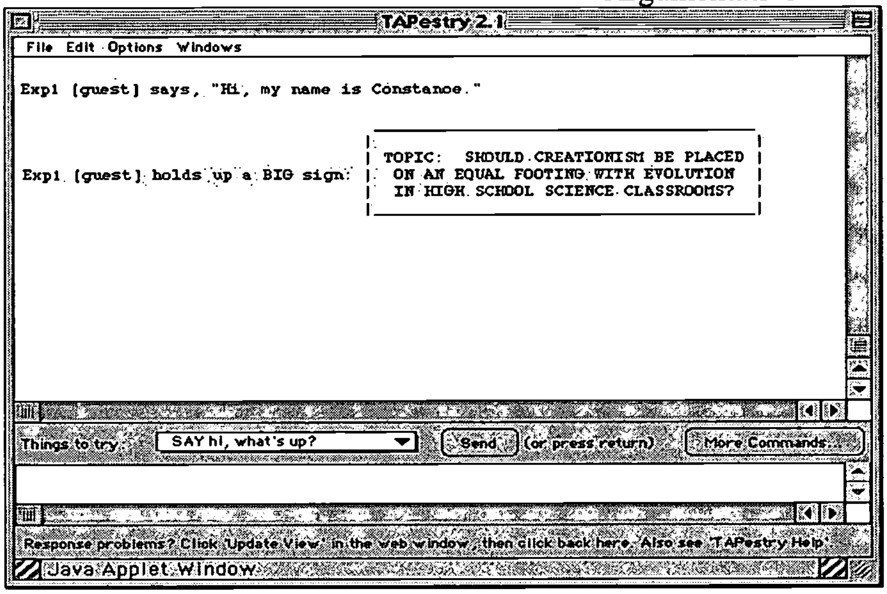


Figure 1 The computer window used by participants in the online environment Tapped In (www.tapped.org).

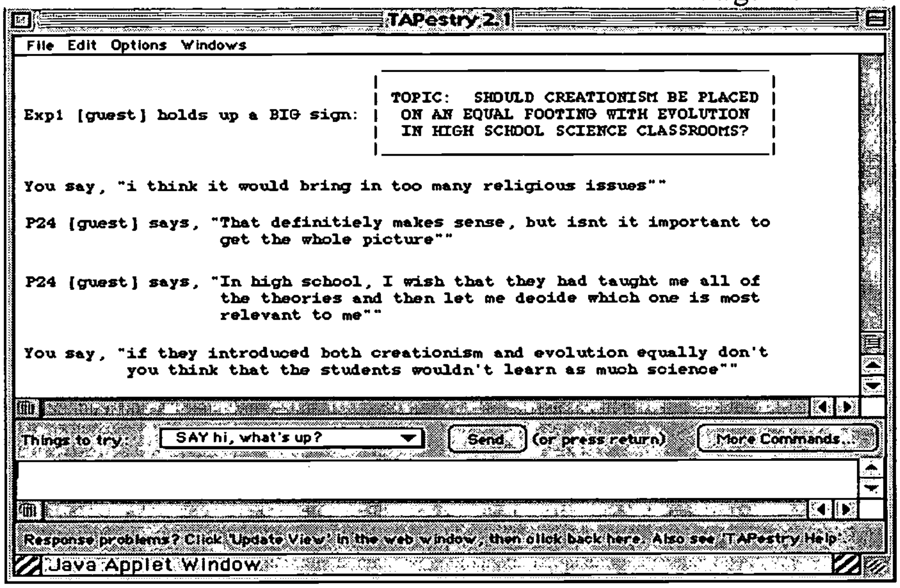


Figure 2 Excerpt from the transcripts of the paired discussion condition.

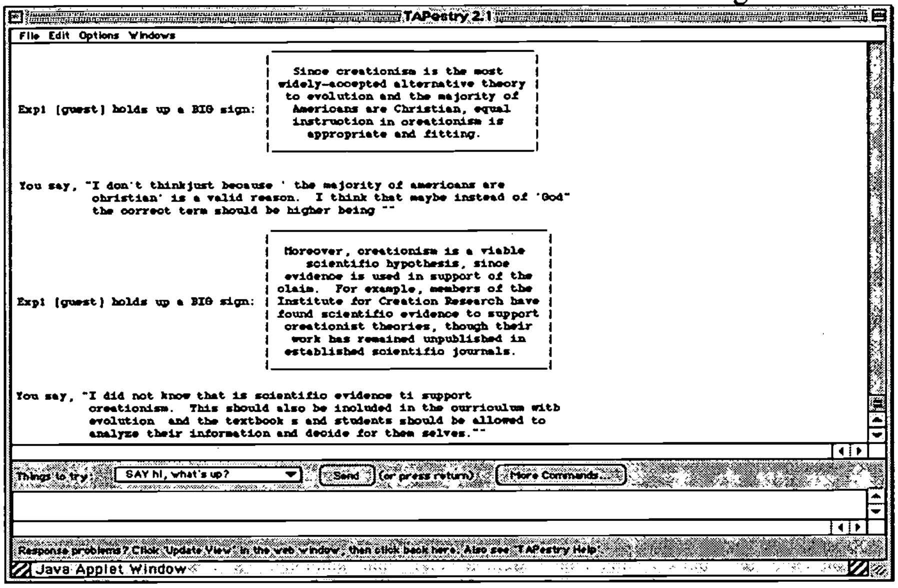


Figure 3 Excerpt from the transcripts of the self–explanation condition.

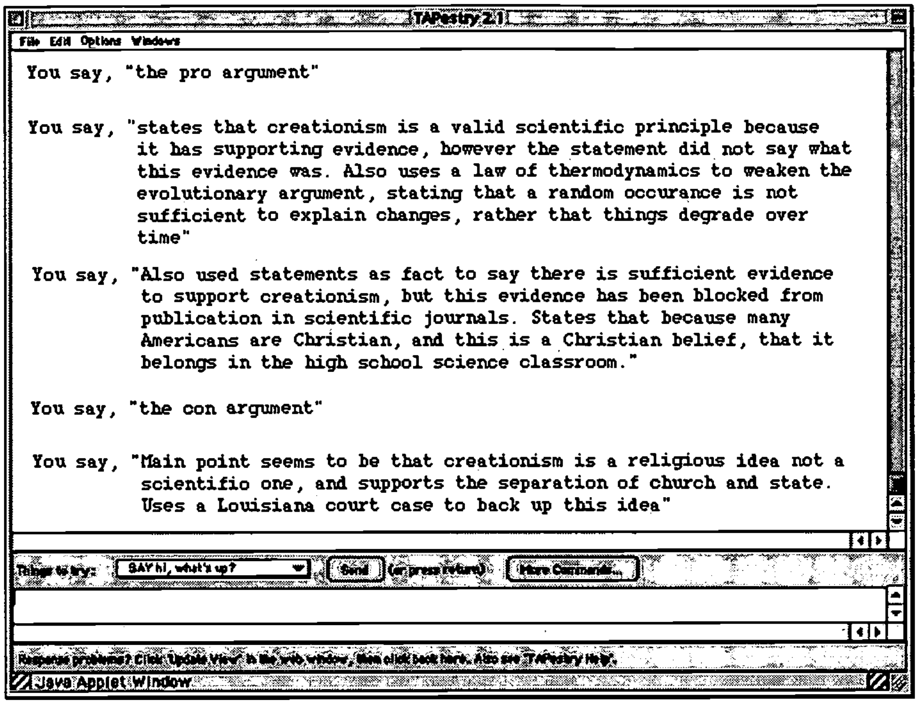


Figure 4 Excerpt from the transcripts of the summarization condition.

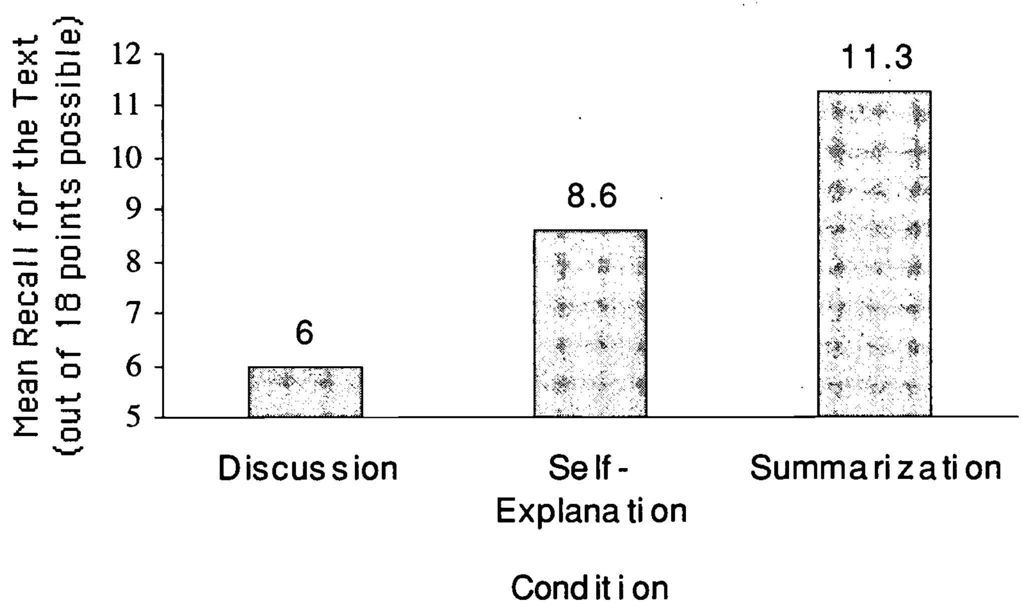


Figure 5. Mean recall for the given pro/con text for each condition.

# NOTICE

## REPRODUCTION BASIS

☐ This document is covered by a signed "Reproduction Release (Blanket)" form (on file within the ERIC system), encompassing all or classes of documents from its source organization and, therefore, does not require a "Specific Document" Release form.

☑ This document is Federally-funded, or carries its own permission to reproduce, or is otherwise in the public domain and, therefore, may be reproduced by ERIC without a signed Reproduction Release form (either "Specific Document" or "Blanket").