Using Argument Diagramming Software to Teach Critical Thinking Skills

Maralee HARRELL
Department of Philosophy, Carnegie Mellon University
Pittsburgh, PA 15213 USA

ABSTRACT

There is substantial evidence from many domains that visual representations aid various forms of cognition. We aimed to determine whether visual representations of argument structure enhanced the acquisition and development of critical thinking skills within the context of an introductory philosophy course. We found a significant effect of the use of argument diagrams, and this effect was stable even when multiple plausible correlates were controlled for. These results suggest that natural—and relatively minor—modifications to standard critical thinking courses could provide substantial increases in student learning and performance.

Keywords
Argument Diagram, Argument Analysis, Critical Thinking, Educational Software, Philosophy.

INTRODUCTION

In the introductory philosophy class at Carnegie Mellon University (*80-100 What Philosophy Is*), as at any school, one of the major learning goals is for the students the students to develop general critical thinking skills. Although there is no generally accepted, well-defined list of skills that constitutes the set we call "critical thinking skills," there seems to be fair agreement on the types of skills to which educators are referring when they speak about teaching critical thinking to their students. Many of these skills have been identified broadly as a global package of knowledge and behaviors [3], [7], [21], [25], or more specifically as a deliberative activity [4], [15].

There is, of course, a long history of interest in teaching students to "think critically" but it's not always clear in what this ability consists. In addition, even though there are a few generally accepted measures (e.g. the California Critical Thinking Skills Test, and the Watson Glaser Critical Thinking Appraisal, but see also [8] and [23]), there is surprisingly little research on the sophistication of students' critical thinking skills, or on the most effective methods for improving students' critical thinking skills. The research that has been done shows that the population of US college students in general has very poor skills [14], [18], [24], and that very few college courses that advertise that they improve students' skills actually do [1], [22], [27].

Most philosophers can agree that one aspect of critical thinking is the ability to analyze, understand, and evaluate an argument—tasks we may call, for the sake of brevity, "argument analysis" This view is expressed by Moore and Parker [19] and by Lee [17]. Kuhn [14] says that "argumentative reasoning skills are in fact fundamental to what educators call 'critical' thinking." (p. 5), and Ennis [7] says that "analyzing arguments" is one of the critical thinking abilities. This covers identifying the stated and unstated premises and the conclusion, and "seeing the structure of an argument." (p. 12).

We have evidence that are students are gaining these argument analysis skills [10], even though typically, even in philosophy classes at Carnegie Mellon (other than logic courses), students are not taught explicitly how to analyze arguments. Instead, if they are taught this skill at all, they are taught implicitly by demonstration from an instructor—e.g., when an instructor writes out the premises of an argument on the chalkboard and leads students through a discussion of the truth of the premises and how well they support the conclusion. Students are also often asked to reconstruct an author's argument—in, e.g., a critical essay—but are not given any general guidelines or methods for completing this kind of task. Thus, we are particularly interested in the efficacy of various alternative teaching methods to increase argument analysis performance.

We are interested, then, in what sorts of methods might be useful to aid students in the tasks of analyzing and evaluating arguments. Analyzing an argument from written text consists in identifying the main conclusion, the premises (including sub-conclusions), and the structure of the argument (i.e., how the premises work together to support the main conclusion), while evaluation consists in determining whether the premises actually do support the conclusion (validity or strength of the argument), and whether the premises are true.

What sorts of methods might be useful to aid students in these tasks? Larkin and Simon [16] argue that diagrammatic representations of information can make recognition of important features and drawing inferences easier than a sentential representation of the same information. In addition, they argue that there are significant differences in the explicitness of the information as well as in the efficacy of search for information between diagrammatic and sentential representations of the same information. Winn [30] makes a similar argument that maps and diagrams contain much more information that is easier to access than plain text just in virtue of the spatial relationships between the parts and between the parts and the frame. Indeed, research on student learning has consistently shown the efficacy of using diagrams to aid text comprehension [2], [5], [20], [26], as well as vocabulary development, post-reading activities and writing preparation [12].

One candidate alternative teaching method, then, is instruction in the use of argument diagrams as an aid to argument comprehension (see Figure 1).

We believe that the ability to construct argument diagrams significantly aids in understanding, analyzing, and evaluating arguments, both one's own and those of others. If we think of an argument the way that philosophers and logicians do—as a series of statements in which one is the conclusion, and the others are premises supporting this conclusion—then an argument diagram is a visual representation of these statements and the inferential connections between them.

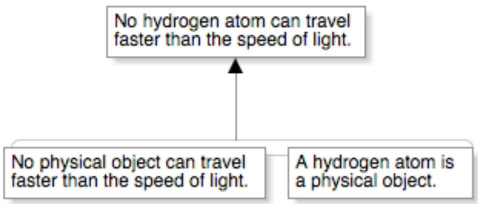

Figure 1: Example of a diagram for a simple argument.

Recent research on argument visualization (particularly computer-supported argument visualization) has shown that the use of software programs specifically designed to help students construct argument diagrams can significantly improve students' critical thinking abilities over the course of a semester-long college-level course [13], [28], [29]. But, of course, one need not have computer software to construct an argument diagram; one needs only a pencil and paper. To our knowledge there has been no research to determine whether the crucial factor is the mere ability to construct argument diagrams, or the aid of a computer platform and tutor, or possibly both. We believe that these two factors—the basic ability to construct argument diagrams and the tools with which they are constructed—need to be tested separately in studying the improvement of students' argument analysis skills.

Hypothesis: Students who are able to construct argument diagrams and use them during argument analysis tasks will improve in performance on critical thinking tasks over the course of a semester long introductory philosophy class significantly more than students in the same class who do not have this ability.

Our introductory philosophy course was a natural place to study the skills acquisition of our students. We typically teach 4 or 5 lectures of this course each semester, with a different instructor for each lecture. While the general curriculum of the course is set, each instructor is given a great deal of freedom in executing this curriculum. For example, it is always a topics based course in which epistemology, metaphysics, and ethics are introduced with both historical and contemporary primary-source readings. It is up to the instructor however, to choose the order of the topics and the assignments. The students who take this course are a mix of all classes and all majors from each of the seven colleges across the University. This study tests the hypothesis by comparing the pretest and posttest scores of students in the introductory philosophy class in the Spring and Fall of 2004 who were able to construct argument diagrams to the scores of those students in the introductory philosophy class who did not have this skill.

METHOD

Participants

One hundred thirty-nine students (46 women, 93 men) in each of four lectures of introductory philosophy in the Spring of 2004, and 130 students (36 women, 94 men) in each of five lectures of introductory philosophy in the Fall of 2004 of 80-100 were studied. In each semester, each lecture of the course had a different instructor and teaching assistant, and the students chose their section. Over the two semesters there were 6 instructors, and 3 of those 6 (Lecturer 1, Lecturer 2 and Lecturer 4) taught one lecture in both semesters studied. During each semester, the students taught by Lecturer 1 were taught the use of argument diagrams to analyze the arguments in the course reading, while the students in the other lectures were taught more traditional methods of analyzing arguments.

Materials and Procedure

Prior to the first semester of the study, the four instructors of the introductory philosophy class in the Spring of 2004 met to determine the learning goals of this course, and design a pair of exams to test the students on relevant skills. In particular, the identified skills were to be able to, when reading an argument, (i) identify the conclusion and the premises; (ii) determine how the premises are supposed to support the conclusion; and (iii) evaluate the argument based on the truth of the premises and how well they support the conclusion.

These exams were used as the pretest and the posttest for the Spring of 2004. For each question on the pretest, there was a structurally (nearly) identical question with different content on the posttest. The tests each consisted of 6 questions, each of which asked the student to analyze a short argument. In questions 1 and 2, the student was only asked to state the conclusion (thesis) of the argument. Questions 3-6 each had five parts: (a) state the conclusion (thesis) of the argument; (b) state the premises (reasons) of the argument; (c) indicate (via multiple choice) how the premises are related; (d) the student was asked to provide a visual, graphical, schematic, or outlined representation of the argument; and (e) decide whether the argument is good or bad, and explain this decision.

After a cursory analysis of the data from this first semester, we decided against including questions for the Fall of 2004 in which the student only had to state the conclusion (i.e. questions 1 and 2 from the Spring 2004 tests). Thus, we designed a new pretest and posttest, each of which consisted of five questions in which the student had again to analyze a short argument. Each question in the Fall 2004 tests had the same five parts as questions 3-6 of the Spring 2004 tests. The Fall 2004 tests thus had 5 questions for directly testing critical thinking skills rather than 4.

In the Spring of 2004, the pretest was given to all students during the second day of class. The students in Lectures 1 and 4 were given the posttest as one part of their final exam (three days after the last day of classes). The students in Lectures 2 and 3 were given the posttest on the last day of classes. In the Fall of 2004, the pretest was given to all students during the third day of class, and the posttest on the last day of classes.

RESULTS AND DISCUSSION

Pretests and posttests were paired by student; there were 139 pairs of tests for the Spring of 2004 and 130 pairs for the Fall of 2004. There were 13 tests in the Spring of 2004 and 12 tests in the Fall of 2004 which did not have pairs, due to students dropping the course, or adding the course late. Each pre-/post-test pair was assigned a unique ID, and the original tests were photocopied (twice, one for each coder) with the identifying information replaced by the ID. The paired tests were coded by two different sets of coders: one session and set of coders for the Spring 2004 tests, and one for the Fall 2004. Each coder independently coded all pairs of tests in his or her group (278 total tests in Spring 2004, and 260 total tests in Fall 2004). After an initial coder-calibration session, each coder was given the tests to be coded in a unique random order.

The codes assigned to each question (or part of a question, except for part (d)) were binary: a code of 1 for a correct answer, and 0 for an incorrect answer. For part (d) of each question, answers were coded according to the type of representation used: Correct argument diagram, Incorrect or incomplete argument diagram, List, Translated into logical symbols like a proof, Venn diagram, Concept map, Schematic like: P1 + P2/Conclusion (C), Other or blank.

Since we were interested in how the use of argument diagramming aided the student in answering each part of each question correctly, the code a student received for part (d) of each multi-part question were preliminarily set aside, while the addition of the codes received on each of the other question-parts determined the raw score a student received on the test.

The primary variables of interest were the fractional pretest and posttest scores for the 18 question-parts for the Spring of 2004, and the 20 question-parts for Fall 2004. In addition, the following data was recorded for each student: whether the student constructed the correct argument diagram in part (d) of each question, which lecture the student was enrolled in, the student's final grade in the course, the student's year in school, the student's home college within the university, the student's sex, and whether the student had taken the concurrent honors course associated with the introductory course.

Comparison of Scores and Gains by Diagram Use

Recall that for the Spring 2004 pretests and posttests, part (d) of questions 3-6 was coded based on the *type* of answer given. From this data, a new variable was defined that indicates how many correct argument diagrams a student had constructed on the posttest. This variable is PostCAD (value = 0, 1, 2, 3, 4). Similarly, for the Fall 2004 pretests and posttests, the type of answer given on part (d) of questions 1-5 was the data recorded. We again defined the variable PostCAD (value = 0, 1, 2, 3, 4, 5), indicating how many correct argument diagrams a student had constructed on the posttest.

The second hypothesis implies that the number of correct argument diagrams a student constructed on the posttest was correlated with the student's gain and standardized gain. For Spring 2004 there were very few students who constructed exactly 2 correct argument diagrams on the posttest, and still fewer who constructed exactly 4. Similar data obtained for Fall 2004. Thus, we grouped the students by whether they had constructed No Correct argument diagrams (PostCAD = 0), Few Correct argument diagrams (PostCAD = 1 or 2), or Many Correct argument diagrams (PostCAD = 3 or more). The results for both Spring 2004 and Fall 2004 are shown in Figure 2.

Since the differences between No Correct and Few Correct is insignificant for both semesters, we did a planned comparison of the variables Gain and StGain for the group of Many Correct with the other two groups combined, again using the variable for the pretest as a covariate. This analysis again indicates that the differences in the pretest scores was significant for predicting the gain (Spring 2004: $df = 1$, $F = 132.00$, $p < .001$; Fall 2004: $df = 1$, $F = 133.00$, $p < .001$), and the standardized gain (Spring 2004: $df = 1$, $F = 31.29$, $p < .001$; Fall 2004: $df = 1$, $F = 28.66$, $p < .001$).

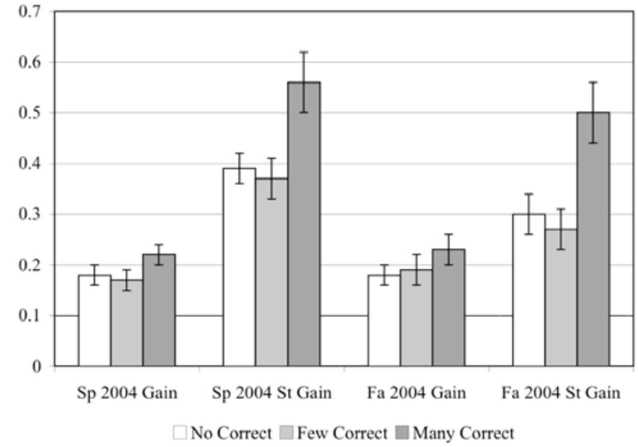

Figure 2: Comparison of gains and standardized gains in both Spring and Fall 2004 for students who constructed No, Few or Many Correct argument diagrams on the posttest.

In addition, this analysis indicates that in each semester, even accounting for differences in pretest score, the differences in gains between students who constructed many correct argument diagram and the other groups were significant (Spring 2004: $df = 1$, $F = 28.13$, $p < .001$; Fall 2004: $df = 1$, $F = 37.78$, $p < .001$), as were the differences in the standardized gains (Spring 2004: $df = 1$, $F = 22.27$, $p < .001$; Fall 2004: $df = 1$, $F = 34.14$, $p < .001$), with the average gain and standardized gain being higher for those who constructed many correct argument diagrams than for those who did not.

These results show that the students who mastered the use of argument diagrams—those who constructed 3 or 4 correct argument diagrams for Spring 2004, and 3, 4 or 5 correct argument diagrams for Fall 2004—gained the most from pretest to posttest, and gained the most as a fraction of the gain that was possible. Our second hypothesis is thus highly confirmed.

Interestingly, those students who constructed few correct argument diagrams were roughly equal on all measures to those who constructed no correct argument diagrams. This may be explained by the fact that nearly all (85%) of the students who constructed few correct argument diagrams and all (100%) of the students who constructed no correct argument diagrams were enrolled in the lectures in which constructing argument diagrams was not explicitly taught; thus the majority of the students who constructed few correct argument diagrams may have done so by accident. This suggests some future work to determine how much the mere ability to construct argument diagrams aids in critical thinking skills compared to the ability to construct argument diagrams in addition to instruction on how to read, interpret, and use argument diagrams.

Prediction of Gain and Standardized Gain

While the results of the above sections seem to confirm our hypothesis that students who constructed correct argument diagrams improved their critical thinking skills more than those who did not, it is possible that there are many causes besides gaining diagramming skills that contributed to the students' improvement. In particular, since during both semesters the students of Lecturer 1 were the only ones explicitly taught the use of argument diagrams, and all of the students were able to chose their lecturer, it is possible that the use of argument diagrams was correlated with instructor's teaching ability, the student's year in school, etc.

To test the hypothesis that constructing correct argument diagrams was the only factor in improving students' critical thinking skills, we first considered how well we could predict the improvement based on the variables we had collected. We defined new variables for each lecturer that each had value 1 if the student was in the class with that lecturer, and 0 if the student was not (Lecturer 1, Lecturer 2, Lecturer 3, and Lecturer 4 for Spring 2004; and Lecturer 1, Lecturer 2, Lecturer 4, Lecturer 5, and Lecturer 6 for Fall 2004).

For each semester, we performed two linear regressions —one for the gain, and one for the standardized gain—using the pretest fractional score, the lecturer variables, and the variables Sex, Honors, Grade, Year and College as regressors. The results of these regressions showed that the variables Sex, Honors, Grade, Year and College are not significant as predictors in either semester of posttest score, gain or standardized gain. We then performed three more linear regressions on the data from each semester—again on the gain, and the standardized gain—this time using PostCAD as a regressor, in addition to the pretest fractional score, the lecturer variables, and the variables Sex, Honors, Grade, Year and College. Again, the results showed that the variables Sex, Honors, Grade, Year and College are not significant as predictors in either semester of posttest score, gain or standardized gain. Then, ignoring the variables that were not significant for either semester, we ran the regressions again. The results for the last set of regression analyses are given in Tables 2 and 3.

Table 1: Results of the regression analyses for each semester when the predicted variable is the gain.

	1st Regression		2nd Regression	
	Coeff.	SD	Coeff.	SD
Spring 2004				
Constant	0.534***	0.036	0.548***	0.035
Pretest	–0.694***	0.062	–0.756***	0.062
Lecturer 1	0.122***	0.025	0.052	0.031
Lecturer 2	0.071**	0.024	0.076**	0.023
Lecturer 3	0.080**	0.024	0.040	0.026
PostCAD			0.034**	0.010
Fall 2004				
Constant	0.505***	0.031	0.444***	0.030
Pretest	–0.657***	0.067	–0.788**	0.064
Lecturer 1	0.082*	0.039	0.074*	0.035
Lecturer 2	0.023	0.030	0.112***	0.031
Lecturer 5	–0.114***	0.032	–0.026	0.032
PostCAD			0.053***	0.009

Note: *$p < .05$, **$p < .01$, ***$p < .001$

We can also see that in the Spring of 2004 when including the variable PostCAD, the variables Lecturer 1 and Lecturer 3 are no longer significant as predictors of gain and standardized gain; that is, when controlling for how many correct argument diagrams a student constructed, the students of Lecturers 1 and 3 were not significantly different from the students of Lecturer 4.

The results show that in each set of regressions a student's pretest score was a highly significant predictor of the gain and standardized gain. A plausible explanation for the negative coefficient for the pretest variable when predicting both gain and standardized gain is that, since the grade a student received on the posttest counted as a part of his or her grade in the course, the students who scored the lowest on the pretest had more incentive to improve, and thus, as a percentage of what they could have gained, gained more than the students who scored highest on the pretest.

In the comparison of the two different regression analyses in each semester, we can see that before including the variable PostCAD the coefficients for Lecturer 1 and Lecturer 3 were significantly positive for predicting a student's gain and standardized gain, while the coefficients for Lecturer 5 are significantly negative for predicting a student's gain and standardized gain. Interestingly, though, the coefficient for Lecturer 2 was significantly positive in the Spring of 2004, but insignificant in the Fall of 2004, for predicting a student's gain and standardized gain.

Table 2: Results of the regression analyses for each semester when the predicted variable is the standardized gain.

	1st Regression		2nd Regression	
	Coeff.	SD	Coeff.	SD
Spring 2004				
Constant	0.818***	0.103	0.851***	0.101
Pretest	–0.948***	0.176	–1.096***	0.179
Lecturer 1	0.305***	0.069	0.136	0.090
Lecturer 2	0.199**	0.069	0.211**	0.067
Lecturer 3	0.209**	0.069	0.112	0.075
PostCAD			0.083**	0.029
Fall 2004				
Constant	0.623***	0.068	0.494***	0.065
Pretest	–0.659***	0.069	–0.951***	0.139
Lecturer 1	0.169*	0.084	0.150*	0.075
Lecturer 2	0.080	0.065	0.281***	0.067
Lecturer 5	–0.188**	0.069	–0.009	0.069
PostCAD			0.118***	0.020

Note: *$p < .05$, **$p < .01$, ***$p < .001$

In the Fall of 2004, however, the coefficient of Lecturer 1 remains significantly positive as a predictor for gain and standardized gain when including the variable PostCAD; that is, even when controlling for how many correct argument diagrams a student constructed, the students of Lecturer 1 did better than the students of Lecturers 4, 5 and 6. Also in the Fall of 2004, after the variable PostCAD is introduced, the variable Lecturer 5 is no longer significant as a predictor of gain and standardized gain; that is, when controlling for how many correct argument diagrams a student constructed, the students of Lecturer 5 were not significantly different from the students of Lecturers 4.

Interestingly, the situation for Lecturer 2 is reversed; after introducing the variable PostCAD into the model in the Spring of 2004, the coefficient for Lecturer 2 was still significantly positive for predicting a student's gain and standardized gain, implying that when controlling for how many correct argument diagrams a student constructed, the students of Lecturer 2 did better than the students of the other lecturers. However, although Lecturer 2 had not been a significant predictor before the variable PostCAD was introduced in the Fall of 2004, after this variable is introduced the coefficient for Lecturer 2 becomes significantly positive for predicting gain and standardized gain, implying that when controlling for how many correct argument diagrams a student constructed, the students of Lecturer 2 did significantly better then the students of Lecturers 4, 5 and 6.

Importantly for testing our second hypothesis, in both semesters when PostCAD is introduced into the model, the coefficient for PostCAD is significantly positive for predicting a student's gain and standardized gain. For the Spring of 2004, this implies that the only measured factors that contributed to a student's posttest score and gain from pretest to posttest was being taught by Lecturer 2 and his or her ability to construct correct argument diagrams on the posttest. For the Fall of 2004, the analysis implies that the only measured factors that contributed to a student's posttest score and gain from pretest to posttest was being taught by Lecturer 1 or Lecturer 2 and his or her ability to construct correct argument diagrams on the posttest.

Thus, in the Spring of 2004, Lecturer 1—the only lecturer who explicitly taught argument diagramming—was not a direct contributing factor to the posttest score, gain or standardized gain. Rather, the students of Lecturer 1 did better only because they were significantly more likely than the other students to construct correct argument diagrams. However, in the Fall of 2004, Lecturer 1 is a direct contributing factor to the posttest score, gain and standardized gain. So, the students of Lecturer 1 performed as they did because they were both significantly more likely than the other students to construct correct argument diagrams, and benefited from other aspects of Lecturer 1's course.

These data seem to support a simple causal picture, shown in Figure 3.

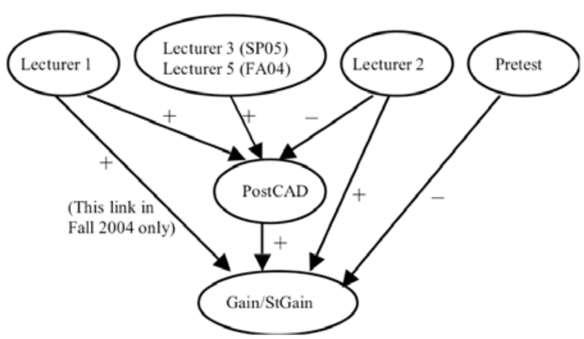

Figure 3: A diagram representing a plausible picture of the causal links between the variables that are significant predictors of gain and standardized gain.

ARGUMENT DIAGRAMMING SOFTWARE

During the past three years, argument diagramming has been used extensively in several Carnegie Mellon University undergraduate philosophy courses, both in demonstrations, and in student homework, to analyze and evaluate arguments from primary source readings. In demonstrations, the diagrams were originally presented fully-formed, constructed using the drawing tools in Word; for homework, the students either did the arguments by hand, or used the same sorts of drawing tools.

The drawbacks of this way of presenting argument diagrams is that it is very cumbersome and time consuming, not to mention static. With this sort of tool, the argument diagram must be constructed ahead of time, and then presented it to the students fully-formed. The advantage of a software program is the potential to construct the diagram "on the fly" both for the user, and for presentation to the students. It is especially important to be able to present a diagram dynamically to students so they can see, as well as participate in, its construction. As noted above, research has shown that the use of software programs specifically designed to help students construct argument diagrams can significantly improve students' critical thinking abilities [12], [28], [29].

There are many of these sorts of software packages on the market; some of the more prominent are Araucaria, Argutect, Athena Standard, Inspriation, Reason!Able. Since argument diagramming is being taught in so many courses for the purposes of both understanding and evaluating texts, and creating novel arguments, we reviewed the advantages and drawbacks of each of these packages from this perspective. We conclude that they are each too powerful in some aspects and too deficient in others for our purposes [9], [11].

Because we believe that software specifically designed for argument diagramming will aid students in developing argument analysis skills, we have been developing our own software package, iLogos, the latest version of which is available for free here:
http://www.phil.cmu.edu/projects/argument_mapping/download

This particular software has not been tested yet to see if its use provides gains over and above argument diagramming without such a tool. Causal diagramming software that shares the same architecture has been tested, and shown to significantly increase a student's ability to perform causal reasoning tasks [6]. We are confident, however, that the use of this program will at least make argument diagramming instruction, demonstration and practice much easier than other tools.

CONCLUSIONS

Many, if not most, undergraduate students never take a critical thinking course in their time in college. There may be several reasons for this: the classes are too hard to get into, the classes are not required, the classes do not exist, etc. It is difficult to understand, though, why any of these would be the case since the development of critical thinking skills are a part of the educational objectives of most universities and colleges, and since the possession of these skills is one of the most sought-after qualities in a job candidate in many fields.

Perhaps, though, both the colleges and employers believe that the ability to reason well is the kind of skill that is taught not intensively in any one course, but rather across the curriculum, in a way that would ensure that students acquired these skills no matter what major they chose. The research seems to show, however, that this is not the case; on tests of general critical thinking skills, students average a gain of less than one standard deviation during their entire time in college, while most of this gain comes just in the first year.

What this study shows is that students do improve substantially their critical thinking skills if they are taught how to construct argument diagrams to aid in the understanding and evaluation of arguments. Although we studied only the effect of the use of argument diagrams in an introductory philosophy course, we see no reasons why this skill could not be used in courses in other disciplines. The creation of one's own arguments, as well as the analysis of others' arguments occurs in nearly every discipline, from Philosophy and Logic to English and History to Mathematics and Engineering. We believe that the use of argument diagrams would be helpful in any of these areas, both in developing general critical thinking skills, and developing discipline specific analytic abilities. In addition to testing the efficacy of our software, we hope to perform more studies in the future to test these conjectures.

REFERENCES

[1] D. Annis & L. Annis, "Does philosophy improve critical thinking?", **Teaching Philosophy**, Vol. 3, 1979, pp. 145-152.

[2] B.B. Armbruster & T.H. Anderson, "Mapping: Representing informative text graphically", in C.D. Holley & D.F. Dansereau (Eds.), **Spatial learning strategies**, New York: Academic Press, 1982.

[3] S.D. Brookfield, **Developing Critical Thinkers: Challenging Adults to Explore Alternative Ways of Thinking and Acting**, San Francisco: Jossey-Bass, 1987.

[4] S.S. Carey, **The Uses and Abuses of Argument**, Mountain View, CA: Mayfield, 2000.

[5] D.F. Dansereau, K.W. Collins, B.A. McDonald, C.D. Holley, J. Garland, G. Diekhoff, & S.H. Evans, "Development and evaluation of a learning strategy program", **Journal of Educational Psychology**, Vol. 71, 1979, pp. 64-73.

[6] M. Easterday, V. Alevan, & R. Scheines, "Tis better to construct than to receive? The effects of diagram tools on causal reasoning", **Artificial Intelligence in Education**, in press.

[7] R.H. Ennis, "A Taxonomy of Critical Thinking Dispositions and Abilities", in J.B. Baron & R.J. Sternberg (Eds.), **Teaching Thinking Skills: Theory and Practice**, New York: Freeman, 1987.

[8] D.F. Halpern, **Thought and knowledge: An introduction to critical thinking**, Hillsdale, NJ: L. Erlbaum Associates, 1989.

[9] Harrell, Maralee. 2005a. "Using Argument Diagramming Software in the Classroom." *Teaching Philosophy*, 28: 163-177.

[10] Harrell, Maralee. 2005b. "Using Argument Diagrams to Improve Critical Thinking Skills in 80-100 *What Philosophy Is.*" Technical Report CMU-PHIL-176. Sept. 7, 2005.

[11] Harrell, Maralee. 2005c. "Using Argument Diagramming Software in the Classroom." 2005 North American Computing and Philosophy Conference.

[12] D.D. Johnson, S.D. Pittleman, & J.E. Heimlich, "Semantic mapping", **Reading Teacher**, Vol. 39, 1986, pp. 778-783.

[13] P.A. Kirschner, S.J.B. Shum, & C.S. Carr (Eds.), **Visualizing argumentation: Software tools for collaborative and educational sense-making**. New York: Springer, 2003.

[14] D. Kuhn, **The skills of argument**. Cambridge: Cambridge University Press, 1991.

[15] J.G. Kurfiss, **Critical Thinking: Theory, Research, Practice, and Possibilities**, ASHE-ERIC Higher Education Report No. 2. Washington, D.C.: ERIC Clearing House on Higher Education and the Association for the Study of Higher Education.

[16] J.H. Larkin & H.A. Simon, "Why a Diagram is (Sometimes) Worth Ten Thousand Words", **Cognitive Science**, Vol. 11, 1987, pp. 65-99.

[17] S.P. Lee, **What Is the Argument? Critical Thinking in the Real World**, New York: McGraw Hill, 2002.

[18] M.L. Means & J.F. Voss, "Who reasons well? Two studies of informal reasoning among children of different grade, ability, and knowledge levels", **Cognition and Instruction**, Vol. 14, 1996, pp. 139-178.

[19] B.N. Moore & R. Parker, **Critical Thinking, 8th Ed.**, New York: McGraw Hill, 2007.

[20] J.D. Novak & D.B. Gowin, **Learning how to learn**. New York: Cambridge University Press, 1984.

[21] R.S. Nickerson, "Why Teach Thinking?", in J.B. Baron & R.J. Sternberg (Eds.), **Teaching Thinking Skills: Theory and Practice**, New York: Freeman, 1987.

[22] E. Pascarella, "The development of critical thinking: Does college make a difference?", **Journal of College Student Development**, Vol. 30, 1989, pp. 19-26.

[23] R. Paul, A. Binker, K. Jensen, & H. Kreklau, **Critical thinking handbook: A guide for remodeling lesson plans in language arts, social studies and science**. Rohnert Park, CA: Foundation for Critical Thinking, 1990.

[24] D.N. Perkins, R. Allen, & J. Hafner, "Difficulties in everyday reasoning", in W. Maxwell & J. Bruner (Eds.), **Thinking: The expanding frontier**. Philadelphia: The Franklin Institute Press, 1983.

[25] L.B. Resnick, **Education and Learning to Think**, Washington, D.C.: National Academy Press, 1987.

[26] R.M. Schwartz, & T.E. Raphael, "Concept of definition: A key to improving students' vocabulary", **Reading Teacher**, Vol. 39, 1985, pp. 198-205.

[27] K. Stenning, R. Cox, & J. Oberlander, "Contrasting the cognitive effects of graphical and sentential logic teaching: reasoning, representation and individual differences", **Language and Cognitive Processes**, Vol. 10, 1995, pp. 333-354.

[28] C.R. Twardy, "Argument Maps Improve Critical Thinking", **Teaching Philosophy**, Vol. 27, 2004, pp. 95-116.

[29] T. van Gelder, "Enhancing deliberation through computer supported visualization", in P.A. Kirschner, S.J.B. Shum, & C.S. Carr (Eds.), **Visualizing argumentation: Software tools for collaborative and educational sense-making**. New York: Springer, 2003.

[30] W. Winn, "Learning from maps and diagrams", **Educational Psychology Review**, Vol. 3, 1991, pp. 211-247.