Argument Diagramming and Critical Thinking in Introductory Philosophy

Maralee Harrell
Carnegie Mellon University
Argument Diagramming and Critical Thinking in Introductory Philosophy

Running Head: ARGUMENT DIAGRAMMING INSTRUCTION IN PHILOSOPHY

Maralee Harrell

Department of Philosophy, Carnegie Mellon University, Pittsburgh, USA

Maralee Harrell
Associate Teaching Professor
Director of Undergraduate Studies
Department of Philosophy
Carnegie Mellon University
135 Baker Hall
Pittsburgh, PA, USA 15213
412.468.8152
mharrell@cmu.edu
http://www.hss.cmu.edu/philosophy/faculty-harrell.php

(Received; final version received)
Abstract

In a multi-study naturalistic quasi-experiment involving 269 students in a semester-long introductory philosophy course, we investigated the effect of teaching argument diagramming (AD) on students’ scores on argument analysis tasks. An argument diagram is a visual representation of the content and structure of an argument. In each study, all of the students completed pre- and posttests containing argument analysis tasks. During the semester, the treatment group was taught AD, while the control group was not. The results were that among the different pretest achievement levels, the scores of low-achieving students who were taught AD increased significantly more than the scores of low-achieving students who were not taught AD, while the scores of the intermediate- and high-achieving students did not differ significantly between the treatment and control groups. The implication of these studies is that learning AD significantly improves low-achieving students’ ability to analyze arguments.

Keywords: critical thinking; philosophy; argument diagramming; argument mapping, graphic organizers, schema
Argument Diagramming and Critical Thinking in Introductory Philosophy

The past few decades have seen a tremendous amount of educational effort and research directed at the improvement of students’ general critical thinking (CT) skills. The search for pedagogical methods that efficiently develop these skills is part of a growing national concern that our high school students are under-prepared for the rigors of college and that our college students are being well trained for particular industries, but inadequately prepared to be participating members of our democratic society (Perkins, Allen, & Hafner, 1983; Kuhn, 1991; Means & Voss, 1996).

In response to these concerns, many universities have made the development of CT skills part of their mission statements, along with adding at least one CT course to their graduation requirements. Unfortunately, many studies have shown that very few college courses actually improve these skills (Annis & Annis, 1979; Pascarella, 1989; Resnick, 1987; Stenning, Cox, & Oberlander, 1995). In our introductory philosophy course, we aim at real improvement, and have recently developed a CT-focused curriculum to supplement the content-focused curriculum traditionally used in this course.

Although there is no generally accepted, comprehensive list of skills that constitutes “CT skills,” there seems to be fair agreement on many types of skills to which educators are referring when they speak about teaching CT. Specifically, most agree that one aspect involves the ability to reconstruct, understand, and evaluate an argument—cognitive tasks we may describe as, “argument analysis” (Ennis, 1987; Kuhn, 1991; Fisher & Scriven, 1997).
The first step in learning argument analysis is acquiring the skill of reading a text for the argument, as opposed to, e.g., reading for the plot (as in a novel) or the facts (as in a textbook). Mandler (1984) provides an overview of research supporting the claim that adults and children as young as 3-years-old possess “story schemata” that guide understanding when reading or listening to a story. Thus, learning the skill of reading for the argument requires students to develop a new schema, or set of schemata, with which they can interpret the text appropriately.

Schema theory, first introduced by Bartlett (1932, 1958) and further developed by Evans (1967), Mandler (1984) and Rumelhart & Ortony (1977), explains cognition as information processing mediated by schemata. A schema is a packet of knowledge containing both data and information about the interconnections among the data. Rumelhart (1980) refers to schemata as the representations of concepts stored in memory, and Sweller (1994) describes schemata as representations of either concepts or problem-solution procedures.

To facilitate the acquisition of new schema, Sweller (1994) recommends reducing the extraneous cognitive load during the learning process. One common way of reducing extraneous cognitive load is by using graphic organizers (GOs), such as diagrams, to supplement regular reading and instruction. Previous research has shown that students’ use of GOs is generally efficacious in producing improvements on a wide range of cognitive tasks—including those generally labelled CT tasks—that are significantly higher than improvements gained by students engaged in reading and regular instruction alone (Horton, et al., 1993; Moore & Readance, 1984).
In this paper, we investigate whether teaching our students argument analysis with a particular kind of GO in our introductory philosophy aids the development of new schema better than the traditional way argument analysis is taught. For our course and our students, argument analysis consists in identifying which statement is the main conclusion, which are sub-conclusions and which are the premises; identifying the structure of the argument by determining how the premises work together to support the conclusion; and evaluating the argument by determining whether the premises actually do support the conclusion, and whether the premises are true.

Our supplementary curriculum consists in teaching students how to read text for an argument and construct argument diagrams (ADs). An AD is a visual representation of content and structure of an argument (for an overview of the development of argument diagramming see Reed, Walton, & Macagno, 2007). For illustration, consider the following argument.

I think everyone would agree that life is worth protecting, and that the environment sustains all of us. It stands to reason, then, that we need to protect the environment. One particular threat to the environment is the emission of greenhouse gasses. This is because greenhouse gasses trap the energy of the sun, causing the warming of the planet, and the warming of the planet could have catastrophic effects on the environment. So, we just can’t avoid the conclusion that we need to reduce greenhouse gas emissions.

For the diagram, the claims are put into boxes, the inferential connections are represented by arrows, and all the excess verbiage is removed (see Figure 1).
Previous research indicates that learning argument diagramming does aid the development of students’ CT skills in a curriculum devoted to informal logic (van Gelder, Bissett, & Gumming, 2004). In order to test the efficacy of our merely supplementary curriculum, we designed a quasi-experimental study in which some of the students in introductory philosophy received AD instruction along with the regular content instruction for the course while the other students in the course received only the regular content instruction. The students completed pretests at the beginning of the semester and posttests at the end, each consisting of a series of argument analysis tasks. We repeated this study, with modifications, the following semester.

Ito summarize: in this context, schema theory would suggest that students with no existing appropriate schemata for reading text will gain significant understanding by acquiring a new schema and tuning it with regular practice. We conducted two studies
aimed at testing this prediction. Our first hypothesis is that students who receive AD instruction will gain more from pretest to posttest than students who did not. Our second hypothesis is that students who began the semester with poor argument analysis abilities will benefit more from AD instruction, and so gain more from pretest to posttest, than students who began with good argument analysis abilities.

Study 1

Method

Participants and Design
In the Spring of 2004, 139 (46 women, 93 men) in four distinct sections of Carnegie Mellon University’s introductory philosophy course took a pretest at the beginning of the semester and a structurally identical posttest at the end. These tests each consisted of several short arguments to be analyzed. Each distinct section had a different lecturer and teaching assistant, and the students chose their section at the start of the term. The students attended class with the lecturer twice a week and a recitation section with the teaching assistant once a week.

The students in Section 1 of the course (35 students—13 women and 22 men) were explicitly taught AD to represent arguments from the text. In contrast, students in Sections 2, 3 and 4 (104 students—33 women and 71 men) were not taught AD, but rather were taught (only implicitly) to use more traditional kinds of representations (e.g. lists of statements).

Materials
To evaluate the development of an “argument schema” over the course of a semester, we developed a pretest to be taken at the beginning of the semester, and a companion posttest to be taken at the end. Each question on the pretest consisted of a series of questions
about a particular argument, and for each argument on the pretest, there was a structurally identical argument on the posttest. 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) 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.

Results

Test Coding
The students who either only took the pretest or only took the posttest were not included in the study, but their tests were used for coder-calibration, prior to each session of coding. Included pretests and posttests were coded by two coders, and each coder independently coded all pairs of tests.

Questions 1 and 2, and each question part of 3-6, except for part (d), was coded 1 for a correct answer, and 0 for an incorrect answer. Thus, there were 18 question-parts that were coded either 1 or 0. During the coder-calibration session, we determined that there were some standard representations that students used, so 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, Translation into logical symbols (like a proof), Venn diagram, Concept map, Schematic (like: P1 + P2/C), Other or blank.

The Percentage Agreement (PA) between the coders was .85 for both tests, so the inter-coder reliability was good. We also wanted to allow for a more nuanced scoring of
each question than either coder alone could give, so for each test, the codes from the two coders on these questions were averaged.

Since we were interested in whether the use of AD aided the student in answering each part of each question correctly, the code a student received for part (d) of each multi-part question was set aside, while the sum of the codes received on each of the other 18 question-parts determined the raw score a student received on the test. This raw score was converted to a percentage (the “score”).

The primary variables of interest were the pretest and posttest scores, whether the student was taught argument diagramming, and the kind of visual representation the student provided for each multi-part question on the posttest for each semester. In addition, the following data were recorded for each student in each semester: which section the student was enrolled in, the student’s final grade in the course, the student’s year in school, the student’s home college, the student’s sex, and whether the student had taken the concurrent honors course associated with the introductory course.

**Student Characteristics**
To determine whether the students in the study differed in any statistically significant characteristic other than being taught AD, we tested how well we could predict students’ gains from pretest to posttest based on the variables we had collected. We performed a regression for Posttest using Pretest, Section, Gender, Honors, Grade, Year, and College as regressors. The results indicate that none of the variables besides Pretest and Section was a factor in a student’s posttest score. Thus, we are confident that the students who were taught AD were not different in any important aspect from the students who were not taught AD.
Grouping Students by Pretest Score
The results of the pretests and posttests were analyzed using a classification of the
students’ academic level as reflected in their pretest results. We divided the students into
three roughly equal groups based on the pretest scores: Low Academic Level,
Intermediate Academic Level and High Academic Level. The dividing lines between the
groups were decided based on the mean test score: the intermediate level included scores
one half of a standard deviation on either side of the mean, the low level and high level
consisted of the scores below and above the intermediate level, respectively. In this study,
scores less than or equal to 0.5 fell into the Low Academic Level group, scores greater
than 0.5 and less than 0.67 fell into the Intermediate Academic Level group, scores
greater than or equal to 0.67 fell into the High Academic Level group.

Comparison of Students by Treatment and by Argument Diagram Use
The first part of the hypothesis was that in general, across all ability levels, the students
who learned AD would gain more from pretest to posttest than students who did not learn
AD. We tested this part of the hypothesis by determining whether the posttest scores of
the students who were taught AD was statistically significantly higher than those of the
students who were not taught AD. The results of this analysis indicate that this part of the
hypothesis is not confirmed. The scores are given in Table 1.

Table 1
Number of Participants, Mean Scores (SD) for the Pretest and the Posttest for Students

<table>
<thead>
<tr>
<th>Treatment</th>
<th>N</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taught</td>
<td>35</td>
<td>0.64 (0.14)</td>
<td>0.87 (0.08)</td>
</tr>
<tr>
<td>Not taught</td>
<td>104</td>
<td>0.58 (0.14)</td>
<td>0.76 (0.12)</td>
</tr>
</tbody>
</table>
Overall, the pretest scores of the students who were taught AD were higher than the pretest scores of the students who were not taught AD, $t(61) = 2.50, p = .015, d = .68$, and the posttest scores of the students who were taught AD were higher than the posttest scores of the students who were not taught AD, $t(89) = 6.05, p < .001, d = 1.22$. However, the gains from pretest to posttest of the students who were taught AD were not different from the gains from pretest to posttest of the students who were not taught AD, $t(53) = 1.35, p > .05$.

The second part of the hypothesis was the differences in improvement on performance of argument analysis tasks between students who do and do not learn AD will be greatest for students who are low-achieving. We tested this part of the hypothesis by determining whether, for each academic level, the mean gain of the students who were taught AD was statistically significantly higher than the mean gain of the students who were not taught AD. The results of this analysis indicate that this second part of the hypothesis is confirmed. The scores are given in Table 2.

Table 2
Number of Participants, Mean Score (SD) for the Pretest and the Posttest Among Academic Levels for Students Who Were Either Taught or Not Taught Argument Diagramming in the Spring of 2004.

<table>
<thead>
<tr>
<th>Level/Treatment</th>
<th>$n$</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taught</td>
<td>6</td>
<td>0.44 (0.06)</td>
<td>0.83 (0.11)</td>
</tr>
<tr>
<td>Not taught</td>
<td>32</td>
<td>0.41 (0.10)</td>
<td>0.69 (0.13)</td>
</tr>
</tbody>
</table>
A two-way ANOVA was conducted for the pretest score with factors “Taught” and “Academic Level”. The main effect of Taught was not statistically significant (as was expected, since the students had not actually been taught anything at the time of the pretest), $F(1, 133) = 0.7, p > .05$, while the main effect of Academic Level was statistically significant (as expected, since Academic Level was based on the pretest scores), $F(2, 133) = 151.2, p < .001$, $\eta^2 = .77$. The interaction between the factors was also not statistically significant, $F(2, 133) = 0.2, p > .05$.

A two-way ANOVA was also conducted for the posttest score with the same factors, and using the pretest score as a covariate. The main effect of Taught was statistically significant, $F(1, 133) = 22.1, p < .001$, $\eta^2 = .16$, as was the main effect of Academic Level, $F(2, 133) = 4.5, p = .01$, $\eta^2 = .11$. The interaction was also statistically significant, $F(2, 133) = 2.9, p < .05$, $\eta^2 = .03$, indicating that the effect of being taught AD was greater for the Low Academic Level students than it was for either the Intermediate or High Academic Level students.

Tukey post-hoc comparisons of the treatment group to the control group among academic levels indicate that the pretest scores of the students who were taught AD were
not different from the pretest scores of the students who were not taught AD, among the low academic level students, $p > .05$, and the intermediate level students, $p > .05$.

However, the posttest scores of the students who were taught AD were statistically significantly higher than the posttest scores of the students who were not taught AD, among both the low academic level students, $t(7) = 3.22, p = .02, d = 1.7$, and the intermediate students, $t(32) = 3.88, p = .002, d = 2.3$. In contrast, the post-hoc comparisons indicate that among the high academic level students, the pretest scores of the students who were taught AD were not different from the pretest scores of the students who were not taught AD, $p > .05$, and neither were the posttest scores, $p > .05$.

**Discussion**

Two main conclusions can be drawn from the results of the pilot study that compare the gains from pretest to posttest of the treatment group (students who were taught AD) and the control group (students who were not taught AD). First, while all students regardless of section gained significantly over the course of the semester, the students in the treatment group did not gain statistically significantly more than the students in the control group. This indicates that the first part of the hypothesis is not confirmed: If we consider all students together, learning AD does not aid in the improvement of performance on argument analysis tasks over a semester-long course in introductory philosophy.

The second conclusion is that students who begin the introductory philosophy course with poor or moderate argument analysis skills benefited the most from being taught AD. That is, among the low and intermediate academic levels, students in the treatment group gained significantly more from pretest to posttest than students in the
control group. This is in contrast to the results that, among the high academic level, the gains of the students in the treatment group and the students in the control group were the same. This indicates that the second part of the hypothesis is confirmed: learning AD aids low- and moderately-achieving students in the improvement of performance on argument analysis tasks over a semester-long course in introductory philosophy.

While these results are very encouraging, this pilot study suffered from several methodological flaws, all of which were addressed in Study 2. First, the number of students in the treatment group was significantly lower than the number of students in the control group. Moreover, the students in the treatment group all received instruction from the same lecturer, while the students in the control group received instruction from 3 different lecturers. In addition, while all the students completed the pretests on the same day, different students completed the posttest on different days, depending on the student’s section.

Finally, nearly all of the students in both the treatment and control groups answered Questions 1 and 2 correctly on both the pretest and the posttest. Also, part (c) of Questions 3-6 asked the students to choose one characteristic to describe the structure of the argument, when in fact several of the arguments had more than one of the characteristics listed. This lead to the determination that the coders should give a code of 1 to part (c) of each question if the answer correctly identified one of the structural characteristics of the argument.
Study 2

Method

Participants and Design
In the Fall of 2004, 130 students (36 women, 94 men) in five distinct sections of the introductory philosophy course were studied. Each distinct section had a different lecturer and teaching assistant, and the students chose their section at the start of the term. Three of the five instructors were also instructors in the first study; one of these instructors taught AD in both studies, one taught AD in the second but not the first, and one did not teach AD in either study. Sixty-eight students (21 women and 47 men) in 3 different sections were explicitly taught AD (the treatment group), while 62 students (17 women and 45 men) in 2 different sections were not taught AD (the control group). This distribution addressed two of the flaws of the previous study: (a) the treatment group and the control group have roughly even numbers, and (b) both the treatment group and the control group were taught by multiple instructors.

Materials
We used modified versions of the pretest and posttest from the previous study. The tests in this study did not have any questions in which the student was only asked to state the conclusion (thesis) of the argument, thus addressing another of the flaws in the previous study. Instead, each test consisted of five questions in which the student had to analyze a short argument, and each question 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) 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. For part (c), the students were asked to indicate all the
characteristics of the structure of the argument that apply, thus correcting a flaw in the
tests of the previous study.

**Results**

*Test Coding*
The procedure for test coding was identical to that in this first study, except that two
different coders were used. In this study, PA = .88 for the pretest, and PA = .89, for the
posttest. Thus, the inter-coder reliability was better than for the first study. And again, for
each test, the codes from the two coders on these questions were averaged. The score for
each test was determined the same way as in the first study, and the same additional data
were recorded for each student.

*Student Characteristics*
Again we tested how well we could predict students’ gains from pretest to posttest based
on the variables we had collected. We performed a regression for Posttest using Pretest,
Section, Sex, Honors, Grade, Year, and College. These results were the same as those in
the first study: as expected, a student’s pretest score and section were statistically
significant predictors of the student’s posttest score, while none of the other variables we
collected was a factor. Thus, we are confident that in this study as well the students who
were taught AD were not different in any important aspect from the students who were
not taught AD.

*Grouping Students by Pretest Score*
Again, we divided the students into three groups based on the pretest scores: Low
Academic Level, Intermediate Academic Level and High Academic Level. The dividing
lines between the groups were determined in the same way as in the first study, and this
time scores less than or equal to 0.35 fell into the Low Academic Level group, scores greater than 0.35 and less than or equal to 0.55 fell into the Intermediate Academic Level group, and scores greater than 0.55 fell into the High Academic Level group.

Comparison of Students by Treatment and by Argument Diagram Use
The first part of the hypothesis was that in general, across all ability levels, the students who learned AD would gain more from pretest to posttest than students who did not learn AD. We tested this part of the hypothesis again by determining whether the posttest scores of the students who were taught AD were statistically significantly higher than those of the students who were not taught AD. The results of this analysis indicate that this part of the hypothesis is confirmed. The scores are given in Table 3.

Table 3
Number of Participants, Mean Scores (SD) for the Pretest and the Posttest for Students Who Were Either Taught or Not Taught Argument Diagramming in the Fall of 2004.

<table>
<thead>
<tr>
<th>Treatment</th>
<th>N</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taught</td>
<td>68</td>
<td>0.48 (0.23)</td>
<td>0.70 (0.16)</td>
</tr>
<tr>
<td>Not taught</td>
<td>62</td>
<td>0.44 (0.17)</td>
<td>0.62 (0.17)</td>
</tr>
</tbody>
</table>

The pretest scores of the students who were taught AD were not higher than the pretest scores of the students who were not taught AD, $t(121) = 1.16, p > .05$, but the posttest scores of the students who were taught AD were statistically significantly higher than the posttest scores of the students who were not taught AD, $t(124) = 2.67, p = .01, d = .66$.

The second part of the hypothesis was the differences in improvement on performance of argument analysis tasks between students who do and do not learn AD
will be greatest for students who are low-achieving. We tested this part of the hypothesis by determining whether, for each academic level, the posttest scores of the students who were taught AD were statistically significantly higher than those of the students who were not taught AD. The results of this analysis indicate that this second part of the hypothesis is also confirmed. The scores are given in Table 4.

Table 4
Number of Participants, Mean Score (SD) for the Pretest and the Posttest Among Academic Levels for Students Who Were Either Taught or Not Taught Argument Diagramming in the Fall 2004.

<table>
<thead>
<tr>
<th>Level/Treatment</th>
<th>n</th>
<th>Pretest</th>
<th>Posttest</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taught</td>
<td>24</td>
<td>0.23 (0.10)</td>
<td>0.62 (0.13)</td>
</tr>
<tr>
<td>Not taught</td>
<td>22</td>
<td>0.25 (0.09)</td>
<td>0.49 (0.17)</td>
</tr>
<tr>
<td>Intermediate</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taught</td>
<td>21</td>
<td>0.48 (0.06)</td>
<td>0.63 (0.14)</td>
</tr>
<tr>
<td>Not taught</td>
<td>27</td>
<td>0.48 (0.05)</td>
<td>0.66 (0.12)</td>
</tr>
<tr>
<td>High</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Taught</td>
<td>23</td>
<td>0.74 (0.12)</td>
<td>0.85 (0.10)</td>
</tr>
<tr>
<td>Not taught</td>
<td>13</td>
<td>0.67 (0.07)</td>
<td>0.75 (0.13)</td>
</tr>
</tbody>
</table>

A two-way ANOVA was conducted for the pretest score with factors “Taught” and “Academic Level”. The main effect of Taught was not statistically significant, $F(1, 124) = 1.3, p > .05$, while the main effect of Academic Level was statistically significant (again, as expected), $F(2, 124) = 291.2, p < .001, \eta^2 = .83$. The interaction effect was also
statistically significant, $F(2, 124) = 3.9, p = .02, \eta^2 = .01$, resulting from, among the High Academic Level students, the difference in pretest scores between those who were and were not taught AD.

An analysis of covariance was conducted for the posttest score with factors “Taught” and “Academic Level”, with “Pretest” as a covariate. The main effect of Taught was statistically significant, $F(1, 123) = 6.9, p = .01, \eta^2 = .03$, as was the main effect of Academic Level, $F(2, 123) = 33.4, p < .001, \eta^2 = .36$. The interaction effect also was statistically significant, $F(2, 123) = 4.3, p = .01, \eta^2 = .04$, again indicating that the effect of being taught AD was greater for the Low Academic Level students than it was for either the Intermediate or High Academic Level students.

Tukey post-hoc comparisons of the treatment group to the control group among academic levels indicate that among each of the low, intermediate and high academic level students, the pretest scores of the students who were taught AD were not different from the pretest scores of the students who were not taught AD ($p > .05$ for all).

However, among the low academic level students, the posttest scores of the students who were taught AD were statistically significantly higher than the posttest scores of the students who were not taught AD, $t(38) = 3.20, p = .02, d = 1.1$. In contrast, the post-hoc comparisons indicate that, among both the intermediate and high academic level students, the posttest scores of the students who were taught AD were not different from the posttest scores of the students who were not taught AD, $p > .05$.

**Discussion**

Two main conclusions can be drawn from the results of the main study that compare the gains from pretest to posttest of the treatment group (students who were taught AD) and
the control group (students who were not taught AD). First, in contrast to the pilot study, while all students regardless of section gained statistically significantly over the course of the semester, the students in the treatment group gained statistically significantly more than the students in the control group. This indicates that the first part of the hypothesis is confirmed: learning AD aids in the improvement of performance on argument analysis tasks over a semester-long course in introductory philosophy.

The second conclusion is that students who begin the introductory philosophy course with poor argument analysis skills benefited the most from being taught AD. That is, among the low academic level, students in the treatment group gained statistically significantly more from pretest to posttest than students in the control group. This is in contrast to the results that, among the intermediate and high academic level, the gains of the students in the treatment group and the students in the control group were the same. For the low academic level and high academic level, this is the same result as that of the pilot study. For the intermediate academic level, however, the results from Study 1 are different from the results of Study 2.

**General Discussion**

**Findings**

For students taking introductory philosophy at Carnegie Mellon University, our studies have two main results. First, while the pilot study did not confirm the first hypothesis, the main study did confirm it. As we believe that the main study was a better study than the pilot, we cautiously claim that students who are taught argument diagramming as an aid to understanding, analyzing and evaluating arguments improve their argument analysis skills over the course of the semester significantly more than students who are not taught
argument diagramming. Second, both the pilot and main studies showed that students with the poorest argument analysis skills at the beginning of the semester benefit the most from instruction on argument diagramming. The low-achieving students who were taught argument diagramming improved their abilities on these tasks significantly more than the low-achieving students who were taught using more traditional methods, while the high-achieving students who were taught argument diagramming improved about the same as the high-achieving students who were not taught argument diagramming.

The results were unclear for students who were intermediate-achievers. In the pilot study, the intermediate achieving students who were taught argument diagramming improved the abilities on argument analysis tasks significantly more than the intermediate-achieving students who were not taught argument diagramming. In the main study, however, there was no difference in improvement among the intermediate-achieving students between those who were and were not taught argument diagramming. Because the main study addressed all of the identified methodological flaws, it is more reasonable to conclude that teaching argument diagramming to students with average argument diagramming skills does not significantly aid the improvement of these skills over the semester.

Schema theory not only offers a motivation for using argument diagramming in an introductory philosophy course, but also explains the results of these two studies. Instruction on argument diagramming to aid argument analysis provides a new schema and a new cognitive process to students for reading text. Students may begin the introductory philosophy course with only one or two general schemata for reading text—the story schema and perhaps the fact schema—especially if they have never, or rarely,
been taught how to read for an argument. This explains why some students perform poorly on argument analysis tasks at the beginning of the semester.

The process of learning how to understand written arguments and construct argument diagrams is then equivalent to the process of acquiring a new reading schema—the argument schema. This new structural schema is a picture of the text in which the claims are enclosed in boxes, and the inferential connections between the claims are the arrows between the boxes. In fact, as Kienpointner (1987) and van Eemeren and Kruiger (1987) argue, this general argument schema can itself be restructured into several different argument schemata that cover a wide range of different kinds of arguments—e.g., *modus ponens*, disjunctive syllogism, argument by analogy, and argument from authority. As the students with poor argument analysis skills become more familiar with the new schema and tune it with extensive practice, their argument analysis abilities—especially under time constraints—greatly improve.

Similarly, the likely reason that students who have average or high argument analysis skills at the beginning of the course do not benefit as much from argument diagram instruction and practice is that they come to the course with reasonably well-developed schemata for reading arguments already in place. These students spend the semester tuning their schemata, and so improve during the semester, but the tuning happens equally well no matter what kind of argument analysis construction they receive.

We conclude that taking Carnegie Mellon University’s introductory philosophy course helps students develop certain critical thinking skills—those we call “argument analysis skills.” We also conclude that learning how to construct argument diagrams
significantly improves the ability of a student with poor argument analysis skills to
analyze, comprehend, and evaluate arguments.

*Educational Importance*
The primary educational importance of these studies is two-fold. First, the results indicate
that it is possible to significantly improve students’ argument analysis skills over the
course of just one semester. This finding is important since many studies have shown that
college students in general improve their critical thinking skills (of which argument
analysis is a part) only one standard deviation at most during all four years of college.
This discouraging statistic, however, may be much improved by more students enrolling
in courses that explicitly focus on the development of critical thinking skills.

Second, these results indicate that a relatively small addition to the curriculum of
an introductory philosophy course can have dramatic benefits for the students that begin
the semester with the poorest argument analysis skills. The initial instruction in
understanding arguments and creating argument diagrams can be given in one or two
class-periods, and regular, weekly homework assignments can be added to reading,
summary, and/or reflection assignments. Teaching argument diagramming does not
require a radical reworking of an instructor’s syllabus, course readings or assignments.
This is a great benefit to instructors who may be reluctant to change a curriculum that has
been working reasonably well.

*Limitations and Future Work*
The pretests and posttests completed by the students in these studies do accurately assess
the skills we want our students to develop over the course of the semester in the
introductory philosophy course. Nonetheless, one significant limitation of these two
studies is the fact that these tests are not standard critical thinking skills tests, and thus comparison to other studies may be difficult. In future studies, we plan to use recognized critical thinking assessment tools (such as the Ennis-Weir Critical Thinking Skills Test or the California Critical Thinking Skills Test) for our pretest and posttest pairs.

Our future work also aims to address other areas in which teaching argument diagramming might usefully aid students. For example, while it is clear that the ability to construct argument diagrams significantly improves a student’s critical thinking skills along the dimensions tested, it would be interesting to consider whether there are other skills that may usefully be labeled “critical thinking” that this ability may help to improve.

In addition, the arguments we used in testing our students were necessarily short and simple, compared to arguments encountered in primary source texts. We would like to know what the effect of teaching argument diagramming would be on a student’s ability to analyze longer and more complex arguments. We suspect that the longer and more complex the argument, the more argument diagramming would help.

Acknowledgements
I would like to thank the many people who taught the courses in this study: Wayne Wu, Mark Ravaglia, Barbara Kauffmann, Jiji Zhang and John Mumma. I would also like to thank the graduate students who served as coders for both of the studies: Ryan Muldoon, Jim Soto, Mikel Negugogor, and Steve Kieffer. I would also like to thank Michele DiPietro, Marsha Lovett, Richard Scheines, and Teddy Seidenfeld for their help and advice with the data analysis; and I am deeply indebted to David Danks for detailed comments on many drafts.

Maralee Harrell is Associate Teaching Professor of Philosophy at Carnegie Melon University. She received her Master’s degree in Physics in 1996 and her PhD in Philosophy and Science Studies in 2000 from UC San Diego.
References


Perspectives from cognitive psychology, linguistics, artificial intelligence, and education (pp. 33-58). Hillsdale, NJ: Lawrence Erlbaum Associates.


Word Count: 6761 + 55 = 6816