

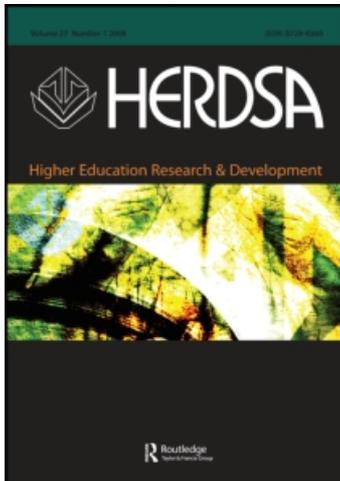
This article was downloaded by: [Harrell, Maralee]

On: 3 May 2011

Access details: Access Details: [subscription number 937173923]

Publisher Routledge

Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK



## Higher Education Research & Development

Publication details, including instructions for authors and subscription information:

<http://www.informaworld.com/smpp/title~content=t713423834>

### Argument diagramming and critical thinking in introductory philosophy

Maralee Harrell<sup>a</sup>

<sup>a</sup> Department of Philosophy, Carnegie Mellon University, Pittsburgh, USA

Online publication date: 03 May 2011

**To cite this Article** Harrell, Maralee(2011) 'Argument diagramming and critical thinking in introductory philosophy', Higher Education Research & Development, 30: 3, 371 – 385

**To link to this Article:** DOI: 10.1080/07294360.2010.502559

**URL:** <http://dx.doi.org/10.1080/07294360.2010.502559>

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: <http://www.informaworld.com/terms-and-conditions-of-access.pdf>

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

## **Argument diagramming and critical thinking in introductory philosophy**

Maralee Harrell\*

*Department of Philosophy, Carnegie Mellon University, Pittsburgh, USA*

*(Received 4 January 2010; final version received 19 May 2010)*

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 post-tests containing argument analysis tasks. During the semester, the treatment group was taught AD, while the control group was not. Methodological problems with the first study were addressed in the second. The results were that among the different pre-test 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 high-achieving students did not differ significantly between the treatment and control groups. The results for intermediate-achieving students were mixed. The implication of these studies is that learning AD significantly improves low- and intermediate-achieving and students' ability to analyze arguments.

**Keywords:** argument diagramming; argument mapping; critical thinking; graphic organizers; philosophy; schema

### **Introduction**

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 developed a CT-focused curriculum to supplement the content-focused curriculum traditionally used in this course.

---

\*Email: mharrell@cmu.edu

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; Fisher & Scriven, 1997; Kuhn, 1991).

The first step in argument analysis is reading a text for the argument, as opposed to, for example, 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 three 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), Rumelhart and Ortony (1977) and Mandler (1984), 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 schemas, 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 schemata better than traditional ways that argument analysis is taught. For our students, argument analysis consists in identifying which statements are the main conclusion, the sub-conclusions and 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 they are true.

Our supplementary curriculum focuses on reading text for an argument and constructing argument diagrams (ADs). An AD is a visual representation of the content and structure of an argument (for an overview of the development of argument diagramming see Reed, Walton and 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 AD, 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, & Cumming, 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 pre-tests at the beginning of the semester and post-tests at the end, each consisting of a series of argument analysis tasks. We repeated this study, with modifications, the following semester.

To 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 hypothesis is that students who began the semester with poor argument analysis abilities will benefit from AD instruction and so gain more from pre-test to post-test than students who do not receive AD instruction.

## Study 1

### Methods

#### *Participants and design*

In the Spring of 2004, 139 students (46 women, 93 men) in four distinct sections of an introductory philosophy course took a pre-test at the beginning of the semester and a structurally identical post-test at the end. These tests each consisted of several short arguments to be analyzed. Each distinct section had a different lecturer and teaching

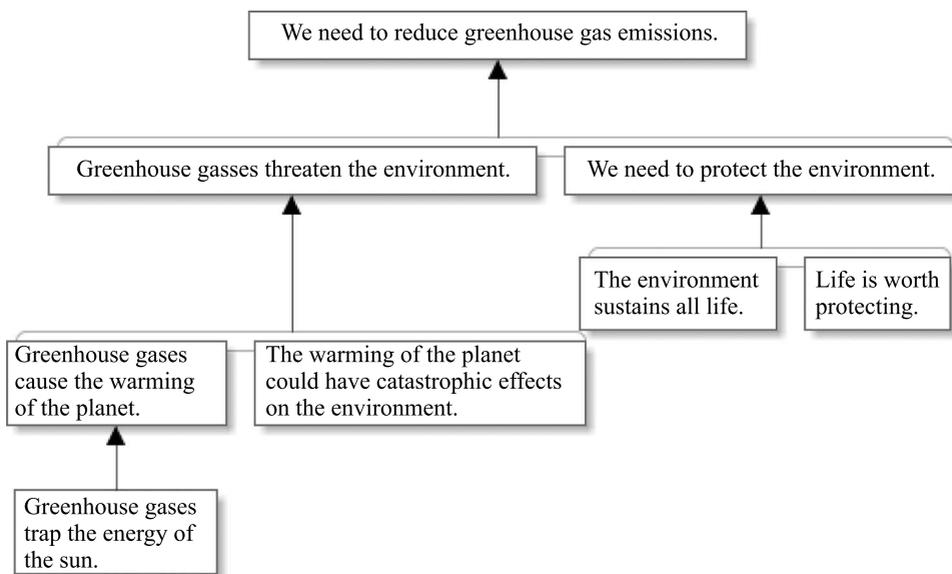


Figure 1. An argument diagram representing an argument for reducing greenhouse gas emissions.

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 class 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 the AD curriculum. In contrast, students in Sections 2, 3 and 4 (104 students – 33 women and 71 men) were not taught the AD curriculum, but, rather, were taught (only implicitly) to use more traditional kinds of representations (e.g. lists of statements).

### *Materials*

To determine whether students developed an ‘argument schema’ over the course of a semester, we developed a pre-test to be taken at the beginning of the semester and a companion post-test to be taken at the end. Each question on the pre-test consisted of a series of questions about a particular argument and for each argument on the pre-test, there was a structurally identical argument on the post-test. In Questions 1 and 2 the student was only asked to state the conclusion of the argument. Questions 3–6 each had five parts: (1) state the conclusion of the argument, (2) state the premises (reasons) of the argument, (3) indicate (via multiple choice) how the premises are related, (4) provide a visual, graphical, schematic or outlined representation of the argument and (5) decide whether the argument is good or bad and explain this decision.

### *Results*

#### *Test coding*

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

Questions 1 and 2, and each question part of 3–6 except for part (4), 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 (4) 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. But one coder was systematically a ‘tougher grader’ than the other and we 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 each question 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 (4) 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 pre-test and post-test scores, whether the student was taught AD and the kind of visual representation the student provided for each multi-part question on the post-test 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 pre-test to post-test based on the variables we had collected. We performed a regression for Gain using Pre-test, Section, Gender, Honors, Grade, Year and College as regressors. The results indicate that none of the variables besides Pre-test and Section was a factor in a student's post-test score. Thus, we are confident that the students in the treatment group were not different in any important aspect from the students in the control group.

### *Grouping students by pre-test score*

The results of the pre-tests and post-tests were analyzed using a classification of the students' academic level as reflected in their pre-test results. We divided the students into three roughly equal groups based on the pre-test 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, Low was score  $\leq 0.5$ , Intermediate was  $0.5 < \text{score} < 0.67$  and High was score  $\geq 0.67$ .

### *Comparison of students by treatment and by argument diagram use*

We first determined whether, overall, the average gain of the students who were taught AD was statistically significantly higher than that of the students who were not. The scores are given in Table 1.

Overall, the post-test scores of the students who were taught AD were higher than those of the students who were not,  $t(89) = 6.05$ ,  $p < .001$ . However, this can be explained by the fact that, overall, the pre-test scores of the students who were taught AD were higher than those of the students who were not,  $t(61) = 2.50$ ,  $p = .015$ . This

Table 1. Number of participants, mean scores (95% Confidence) (SD) for the pre-test, post-test and gain for students who were either taught or not taught argument diagramming in the spring of 2004.

Treatment	<i>N</i>	Pre-test	Post-test	Gain
Taught	35	0.64(0.046)(0.14)	0.87(0.026)(0.08)	0.22(0.052)(0.15)
Not taught	104	0.58(0.028)(0.14)	0.76(0.023)(0.12)	0.18(0.027)(0.13)

explanation is further supported by the fact that the gains from pre-test to post-test of the students who were taught AD were not significantly different from the gains of the students who were not taught AD,  $t(53) = 1.35, p > .05$ .

The hypothesis was that 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 by determining whether, for each academic level, the mean gain from pre-test to post-test of the students who were taught AD was statistically significantly higher than the mean gain of the students who were not. The results of this analysis indicate that the hypothesis is confirmed. The scores are given in Table 2.

A two-way ANOVA was conducted for Pre-test 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 pre-test),  $F(1,133) = 0.7, p > .05$ , while the main effect of Academic Level was statistically significant (as expected, since Academic Level was based on the pre-test scores),  $F(2,133) = 151.2, p < .001$ . The interaction between the factors was not statistically significant,  $F(2,133) = 0.2, p > .05$ .

An ANCOVA was conducted for Post-test with the same factors and using Pre-test as a covariate. The main effect of Taught was statistically significant,  $F(1,133) = 21.1, p < .001$ . The main effect of Academic Level was not significant,  $F(2,133) = 2.1, p > .05$ , but the effect of the interaction was,  $F(2,133) = 2.9, p < .05$ . In addition, an ANOVA was conducted for Gain with the same factors. The main effect of Taught was statistically significant,  $F(1,133) = 14.6, p < .001$ , as was Academic Level,  $F(2,133) = 40.8, p < .001$ , but not the interaction,  $F(2,133) = 2.4, p > .05$ .

In order to explain these results, Tukey post-hoc comparisons of the treatment group and the control group among academic levels were performed. These results indicate that the pre-test scores of the students who were taught AD were not significantly different from the pre-test scores of the students who were not, among any of the academic levels ( $p > .05$  in all cases). However, the post-test scores of the students who were taught AD were significantly higher than the post-test scores of the students who were not, among both the low academic level students,  $t(7) = 3.10, p = .02$ , and the intermediate students,  $t(32) = 3.93, p < .01$ , as were the gains for the low academic level students,  $t(7) = 2.48, p = .14$  and the intermediate students,

Table 2. Number of participants, mean score (95% Confidence) (SD) for the pre-test, post-test and gain among academic levels for students who were either taught or not taught argument diagramming in the spring of 2004.

Level/treatment	N	Pre-test	Post-test	Gain
Low				
Taught	6	0.44(0.047)(0.06)	0.83(0.09)(0.11)	0.40(0.10)(0.12)
Not taught	32	0.41(0.035)(0.10)	0.69(0.046)(0.13)	0.28(0.048)(0.13)
Intermediate				
Taught	13	0.60(0.024)(0.04)	0.90(0.027)(0.06)	0.30(0.035)(0.06)
Not taught	47	0.60(0.015)(0.05)	0.78(0.031)(0.09)	0.18(0.029)(0.10)
High				
Taught	16	0.76(0.041)(0.08)	0.85(0.038)(0.08)	0.10(0.05)(0.10)
Not taught	25	0.75(0.024)(0.06)	0.82(0.043)(0.11)	0.07(0.044)(0.11)

Table 3. Number of participants, Cohen's  $d$  and effect size  $r$  for each academic level for students who were either taught or not taught argument diagramming in the spring of 2004.

Level/treatment	$N$	Cohen's $d$	$r$
Low			
Taught	6	4.5	0.92
Not taught	32	2.4	0.77
Intermediate			
Taught	13	6.1	0.95
Not taught	47	2.4	0.77
High			
Taught	16	1.2	0.52
Not taught	25	0.84	0.39

$t(32) = 3.51, p < .01$ . In contrast, among the high academic level students, neither the post-test scores nor the gains of the students who were taught AD were different from the pre-test scores of the students who were not taught AD ( $p > .05$  in both cases).

More importantly, though, we are interested in the effect sizes for the treatment group versus the control group. Cohen's  $d$  (the difference in average means of the pre-test and post-test divided by the standard deviation of the pre-test) was calculated for each academic level. The results are given in Table 3.

### Discussion

The main conclusion that can be drawn from the results of the pilot study is that by dividing the participants into groups based on their pre-test scores, we can see clearly whom the AD instruction is helping the most. These results indicate that students who begin the introductory philosophy course with poor or moderate argument analysis skills benefit from being taught AD. That is, among the low and intermediate academic levels, students in the treatment group gained significantly more from pre-test to post-test than students in the control group. The difference in the effect size of the instruction over the course of the semester is extremely large – for the low and intermediate levels, 6.9 and 6.9, respectively, for the students who were taught AD, but only 2.8 and 3.5, respectively, for the students who were not.

This is in contrast to the result 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 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 three different lecturers. In addition, while all the students completed the pre-tests on the same day, different students completed the post-test on different days, depending on the student's class.

Finally, nearly all of the students in both the treatment and control groups answered Questions 1 and 2 correctly on both the pre-test and the post-test. Also, part (3) 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 led to the determination that the coders should give a code of 1 to part (3) of each question if the answer correctly identified one of the structural characteristics of the argument.

## Study 2

### *Methods*

#### *Participants and design*

In the Fall of 2004, 130 students (36 women, 94 men) in five distinct classes of the introductory philosophy course were studied. Each distinct class had a different lecturer and teaching assistant and the students chose their class 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 classes were explicitly taught AD (the treatment group), while 62 students (17 women and 45 men) in 2 different classes were not taught AD (the control group). This distribution addressed two of the flaws of the previous study: (1) the treatment group and the control group have roughly even numbers and (2) both the treatment group and the control group were taught by multiple instructors.

#### *Materials*

We used modified versions of the pre-test and post-test 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 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 again had five parts: (1) state the conclusion of the argument, (2) state the premises (reasons) of the argument, (3) indicate (via multiple choice) how the premises are related, (4) provide a visual, graphical, schematic or outlined representation of the argument and (5) decide whether the argument is good or bad and explain this decision. For part (3), 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 pre-test and  $PA = .89$  for the post-test. Thus, the inter-coder reliability was better than for the first study. Again, for each test the codes on each question 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 pre-test to post-test based on the variables we had collected. We performed a regression for Gain using Pre-test, Section, Sex, Honors, Grade, Year and College. These results were the same as those in the first study: as expected, a student's pre-test score and section were statistically significant predictors of the student's post-test score, while none of the other variables we collected was a factor. Thus, we are confident that, as in the first study, the students who were taught AD were not different in any important aspect from the students who were not.

*Grouping students by pre-test score*

Again, we divided the students into three groups based on the pre-test scores: Low Academic Level, Intermediate Academic Level and High Academic Level. While this set of tests was more difficult (because of the change of the question in part [3]), the division between the groups were determined in the same way as in the first study. In this study, Low was score  $\leq 0.35$ , Intermediate was  $0.35 < \text{score} < 0.55$  and High was score  $\geq 0.55$ .

*Comparison of students by treatment and by argument diagram use*

We first determined that, unlike in the previous study, overall, the average gain of the students who were taught AD was not statistically significantly higher than that of the students who were not. The scores are given in Table 4.

The post-test scores of the students who were taught AD were statistically significantly higher than the post-test scores of the students who were not taught AD,  $t(124) = 2.67, p = .01$ . Unlike in the previous study, however, this cannot be explained by differences in pre-test scores as the pre-test scores of the students in the treatment group were not significantly higher those of the control group,  $t(121) = 1.16, p > .05$ . Furthermore, the overall gains from pre-test to post-test of the students who were taught AD were not significantly different from the gains from pre-test to post-test of the students who were not taught AD,  $t(126) = 1.22, p > .05$ . This puzzle leads to the hypothesis that 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 by determining whether, for each academic level, the mean gain from pre-test to post-test of the students who were taught AD was statistically significantly higher than those of the students who were not. The results of this analysis indicate that this hypothesis is confirmed. The scores are given in Table 5.

A two-way ANOVA was conducted for pre-test with factors 'Taught' and 'Academic Level'. The main effect of Taught was not statistically significant (as

Table 4. Number of participants, mean scores (95% Confidence) (SD) for the pre-test, post-test and gain for students who were either taught or not taught argument diagramming in the fall of 2004.

Treatment	N	Pre-test	Post-test	Gain
Taught	68	0.48(0.057)(0.23)	0.70(0.039)(0.16)	0.22(0.046)(0.19)
Not taught	62	0.44(0.043)(0.17)	0.62(0.046)(0.17)	0.18(0.039)(0.15)

Table 5. Number of participants, mean score (95% Confidence) (SD) for the pre-test and the post-test among academic levels for students who were either taught or not taught argument diagramming in the fall 2004.

Level/treatment	<i>N</i>	Pre-test	Post-test	Gain
Low				
Taught	24	0.23(0.04)(0.10)	0.62(0.054)(0.13)	0.39(0.07)(0.17)
Not taught	22	0.25(0.037)(0.09)	0.49(0.075)(0.17)	0.24(0.07)(0.16)
Intermediate				
Taught	21	0.47(0.027)(0.06)	0.63(0.06)(0.14)	0.17(0.057)(0.13)
Not taught	27	0.48(0.022)(0.07)	0.66(0.045)(0.12)	0.18(0.048)(0.12)
High				
Taught	23	0.74(0.048)(0.08)	0.85(0.038)(0.10)	0.10(0.049)(0.12)
Not taught	13	0.67(0.041)(0.07)	0.75(0.071)(0.13)	0.09(0.085)(0.15)

expected),  $F(1,124) = 1.3, p > .05$ , while the main effect of Academic Level was statistically significant (again, as expected),  $F(2,124) = 291.2, p < .001$ . The interaction effect was also statistically significant,  $F(2,124) = 3.9, p = .02$ .

An ANCOVA was conducted for post-test with factors 'Taught' and 'Academic Level', with 'Pre-test' as a covariate. While the main effect of Academic Level was not significant,  $F(2,123) = 2.26, p > .05$ , the main effect of Taught was statistically significant,  $F(1,123) = 5.92, p = .01$ , as was the interaction,  $F(2,123) = 4.51, p = .01$ . In addition, an ANOVA was conducted for Gain with the same factors. The main effect of Taught was statistically significant,  $F(1,124) = 3.05, p = .05$ , as was the main effect of Academic Level,  $F(2,133) = 24.78, p < .001$  and the interaction,  $F(2,124) = 4.88, p = .009$ .

In order to explain these results, Tukey post-hoc comparisons of the treatment group to the control group among academic levels were performed. These results indicate that among each of the low, intermediate and high academic level students, the pre-test scores of the students who were taught AD were not different from the pre-test scores of the students who were not taught AD ( $p > .05$  for all). However, among the low academic level students, the post-test scores of the students who were taught AD were statistically significantly higher than the post-test scores of the students who were not taught AD,  $t(38) = 3.42, p = .01$ , as were the gains,  $t(38) = 3.57, p = .007$ . In contrast, the post-hoc comparisons indicate that, among both the intermediate and high academic level students, neither the post-test scores nor the gains of the students who were taught AD were different from those of the students who were not taught AD,  $p > .05$ .

Again, we are interested in the effect sizes for the treatment group versus the control group and Cohen's  $d$  was calculated for each academic level. The results are given in Table 6.

### Discussion

The results of the main study corroborate the conclusion of the pilot study: that low-achieving students do, in fact, benefit from being taught argument diagramming. First, there was no statistical difference in the pre-test scores between the treatment group

Table 6. Number of participants, Cohen's  $d$  and effect size  $r$  for each academic level for students who were either taught or not taught argument diagramming in the fall of 2004.

Level/treatment	$N$	Cohen's $d$	$r$
Low			
Taught	24	3.3	0.85
Not taught	22	1.7	0.65
Intermediate			
Taught	21	1.4	0.58
Not taught	27	1.9	0.70
High			
Taught	23	0.97	0.44
Not taught	13	0.84	0.39

and the control group. And while the students in the treatment group had statistically significantly higher post-test scores than the students in the control group, the gains from pre-test to post-test of the treatment group were not statistically different from the gains of those in the control group.

The solution to this puzzle is to divide the participants into groups based on their pre-test scores. These results indicate 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 pre-test to post-test 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 not statistically significantly different. For both 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.

This difference can be seen specifically in the comparison of the effect sizes of the treatment in both studies. In both studies, the *difference* between the effect sizes (as measured by Cohen's  $d$ ) is large (2.1, 1.6) for the low achieving students and small (0.36, 0.13) for the high achieving students. The studies differ, however, for the intermediate achieving students (3.7 for Study 1 and -0.5 for Study 2).

## General discussion

### Findings

For students taking introductory philosophy, our studies have one main result, shown in two ways. First, 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 their 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. However, given the evidence from other studies that argument diagramming does significantly aid CT skills development (van Gelder, Bissett, & Cumming, 2004), it seems reasonable to conclude that teaching argument diagramming to students with average argument diagramming skills does, in fact, aid the improvement of these skills over the semester.

Second, the sizes of the effect of the introductory philosophy course on students' argument analysis skills are quite high in each of the achievement levels. This is partly due to the fact that the pre-tests and post-tests were specifically designed to test the skills we were trying to teach our students. But the above results were borne out in the effect sizes. In both studies the size of the effect was far greater for low-achieving students who were taught argument diagramming than who were not. Also, in both studies, the effect size was roughly equal for high-achieving students who were and were not taught argument diagramming. In contrast, in the first study the effect size was quite different among intermediate-achieving students for those who were taught argument diagramming and those who were not, while in the second study there was no difference.

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 ADs 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 – for example *modus ponens*, disjunctive syllogism, argument by analogy and argument from authority. As the students with poor or average 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 well-developed 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 our introductory philosophy course helps students develop certain critical thinking skills – those we call ‘argument analysis skills’. We also conclude that learning how to construct ADs significantly improves the ability of a student with poor or average 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 CT skills (of which argument analysis is a part) only one standard deviation at most during all four years of college (Pascarella & Terenzini, 2005). This discouraging statistic, however, may be much improved by more students enrolling in courses that explicitly focus on the development of CT 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 who 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 pre-tests and post-tests 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 pre-test and post-test pairs.

In addition, there is a potential problem with the use of two tests which are not known to be equivalent for the pre-test and post-test. It could be that any improvement measured is due to the fact that the post-test is easier than the pre-test.<sup>1</sup> In future studies, equal numbers of both tests should be randomly given in each section as a pre-test and then each student would get the test not taken as a post-test.

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 labelled ‘critical thinking’ that this ability may help to improve.

In addition, the arguments we used in testing our students were necessarily short and simple when 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.

## Note

1. I thank an anonymous reviewer for highlighting this point.

## References

- Annis, D., & Annis, L. (1979). Does philosophy improve critical thinking? *Teaching Philosophy*, 3(2), 145–152.
- Bartlett, F.C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge: Cambridge University Press.
- Bartlett, F.C. (1958). *Thinking: An experimental and social study*. New York: Basic Books.
- Ennis, R.H. (1987). A taxonomy of critical thinking dispositions and abilities. In J.B. Baron & R.J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 1–26). New York: W.H. Freeman & Co.
- Evans, S.H. (1967). A brief statement of schema theory. *Psychonomic Science*, 8(2), 87–88.
- Fisher, A., & Scriven, D. (1997). *Critical thinking: Its definition and assessment*. Point Reyes, CA: Edgepress.
- Horton, P.B., McConney, A.A., Gallo, M., Woods, A.L., Senn, G.J., & Hamelin, D. (1993). An investigation of the effectiveness of concept mapping as an instructional tool. *Science Education*, 77(1), 95–111.
- Kienpointner, M. (1987). Towards a typology of argumentation schemes. In F.H. van Eemeren, R. Grootendorst, J.A. Blair & C.A. Willard (Eds.), *Argumentation: Across the lines of discipline* (pp. 275–287). Dordrecht: Foris Publications.
- Kuhn, D. (1991). *The skills of argument*. Cambridge: Cambridge University Press.
- Mandler, J.M. (1984). *Stories, scripts and scenes: Aspects of schema theory*. Hillsdale, NJ: Lawrence Erlbaum.
- Means, M.L., & Voss, J.F. (1996). Who reasons well? Two studies of informal reasoning among children of different grade, ability and knowledge levels. *Cognition and Instruction*, 14(2), 139–178.
- Moore, D.W., & Readence, J.E. (1984). A quantitative and qualitative review of graphic organizer research. *Journal of Educational Research*, 78(1), 11–17.
- Pascarella, E. (1989). The development of critical thinking: Does college make a difference? *Journal of College Student Development*, 30(1), 19–26.
- Pascarella, E.T., & Terenzini, P.T. (2005). *How college affects students. Volume 2: A third decade of research*. San Francisco and Oxford: Jossey-Bass.
- Perkins, D.N., Allen, R., & Hafner, J. (1983). Difficulties in everyday reasoning. In W. Maxwell & J. Bruner (Eds.), *Thinking: The expanding frontier* (pp. 177–189). Philadelphia: The Franklin Institute Press.
- Resnick, L.B. (1987). *Education and learning to think*. Washington, DC: National Academy Press.
- Reed, C., Walton, D., & Macagno, F. (2007). Argument diagramming in logic, law and artificial intelligence. *Knowledge Engineering Review*, 22(1), 87–109.
- Rumelhart, D.E. (1980). Schemata: The building blocks of cognition. In R.J. Shapiro, B.C. Bruce & W.F. Brewer (Eds.), *Theoretical issues in reading comprehension: Perspectives from cognitive psychology, linguistics, artificial intelligence and education* (pp. 33–58). Hillsdale, NJ: Lawrence Erlbaum.
- Rumelhart, D.E., & Ortony, A. (1977). The representation of knowledge in memory. In R.C. Anderson, R.J. Spiro & W.E. Montague (Eds.), *Schooling and the acquisition of knowledge*. Hillsdale, NJ: Lawrence Erlbaum.
- Stenning, K., Cox, R., & Oberlander, J. (1995). Contrasting the cognitive effects of graphical and sentential logic teaching: reasoning, representation and individual differences. *Language and Cognitive Processes*, 10(3/4), 333–354.

- Sweller, J. (1994). Cognitive load theory, learning difficulty and instructional design. *Learning and Instruction*, 4(4), 295–312.
- van Eemeren, F.H., & Kruiger, T. (1987). Identifying argumentation schemes. In F.H. van Eemeren, R. Grootendorst, J.A. Blair & C.A. Willard (Eds.), *Argumentation: Perspectives and approaches* (pp. 70–81). Dordrecht: Foris Publications.
- van Gelder, T., Bissett, M., & Cumming, G. (2004). Cultivating expertise in informal reasoning. *Canadian Journal of Experimental Psychology*, 58(2), 142–152.