

DOES PHILOSOPHY IMPROVE CRITICAL THINKING SKILLS?

Claudia María Álvarez Ortiz

**Submitted in total fulfilment of the requirements of the degree of Master of
Arts**

February, 2007

Department of Philosophy- Faculty of Arts

The University of Melbourne

Abstract

It is widely assumed, and often asserted, that studying philosophy improves critical thinking skills. This idea is the most widely-presented rationale for studying philosophy at university, and is therefore a key pillar of support for the existence of philosophy as a discipline in the modern university. However, the assumption has never been exposed to rigorous scrutiny. This is ironic since philosophers claim to be highly critical and, more than other disciplines, to challenge deep assumptions wherever they are found.

This thesis makes a first attempt to subject the assumption that studying philosophy improves critical thinking skills to rigorous investigation.

The first task, in Chapter 2, is to clarify what the assumption amounts to, i.e., the meaning of the sentence "studying philosophy improves critical thinking." This requires us to determine the relevant senses of key terms. The thesis arrives at the following interpretation: the assumption is making the empirical claim that studying Anglo-American analytic philosophy (i.e., doing those things that a conscientious student would generally do when enrolled in a philosophy subject at a typical Anglo-American university) is especially effective in producing gains, in critical thinking skills, where gains are interpreted as detectably higher levels of skill after studying than before. "Especially" is a comparative claim, and the relevant comparisons are deemed to be university study in general, studying critical thinking in its own right, and studying critical thinking using a particularly effective method ("LAMP").

The assumption has a certain initial plausibility. Thus the second task, in Chapter 3, is to articulate and critically examine the standard arguments that are raised in support of the assumption (or rather, would be raised if philosophers were in the habit of providing support for the assumption). These arguments are found to be too weak to establish the truth of the assumption. The failure of the standard arguments leaves open the question of whether the assumption is in fact true. The thesis argues at this point that, since the assumption is making an empirical assertion, it should be investigated using standard empirical techniques as developed in the social sciences.

In Chapter 4, I conduct an informal review of the empirical literature. The review finds that evidence from the existing empirical literature is inconclusive.

Chapter 5 presents the empirical core of the thesis. I use the technique of meta-analysis to integrate data from a large number of empirical studies. This meta-analysis gives us the best-yet fix on the extent to which critical thinking skills improve over a semester of studying philosophy, general university study, and studying critical thinking. The meta-analysis results

indicate that students do improve while studying philosophy, and apparently more so than general university students, though we cannot be very confident that this difference is not just the result of random variation. More importantly, studying philosophy is less effective than studying critical thinking, regardless of whether one is being taught in a philosophy department or in some other department. Finally, studying philosophy is much less effective than studying critical thinking using techniques known to be particularly effective such as LAMP.

The results of our review of the standard arguments, informal review of the literature, and meta-analysis, suggest four basic conclusions. First, there is insufficient evidence to be confident that studying philosophy improves critical thinking skills any more than studying other academic disciplines. Second, the results indicate that studying philosophy appears to be less effective than studying critical thinking in its own right. Third, there appear to be techniques which, if introduced into philosophy teaching, could improve the extent to which studying philosophy improves critical thinking. Fourth, further research is desirable to gather better evidence in a number of areas. In the light of these findings, though it may sound bold to suggest it, perhaps philosophers themselves more fully live up to their own ideals, by leading the search for more and better evidence regarding the impact of their discipline on the development of critical thinking skills.

TABLE OF CONTENTS

| | |
|--|------|
| Abstract | ii |
| Declaration | iv |
| Acknowledgements | v |
| TABLE OF CONTENTS | viii |
| Table of Figures..... | xii |
| 1 Introduction | 1 |
| 2 Clarifying the Assumption | 5 |
| 2.1 Defining the Terms..... | 5 |
| 2.1.1 “Studying Philosophy” | 5 |
| 2.1.1.1 What is philosophy?..... | 5 |
| 2.1.1.2 <i>Studying</i> Philosophy at University | 9 |
| 2.1.2 “Critical Thinking” | 10 |
| 2.1.2.1 The Prevailing View of Critical Thinking | 11 |
| 2.1.3 “Improves”..... | 14 |
| 2.1.3.1 What is meant by ‘Improvement’ | 14 |
| 2.1.3.2 ‘Causing’ Improvement | 16 |
| 2.1.3.3 Improving CTS ‘through the study of philosophy’ | 17 |
| 2.2 The Nature of the Assumption that Studying Philosophy improves Critical Thinking | 19 |
| 2.2.1 The Nature of the Assumption..... | 19 |
| 2.2.2 Key Questions | 20 |
| 3 Questioning the Assumption..... | 21 |
| 3.1 Why the Assumption Should be Questioned | 21 |
| 3.2 Arguments for the Assumption..... | 21 |
| 3.2.1 Conventional Wisdom..... | 22 |

| | | |
|---------|---|----|
| 3.2.2 | Anecdotal Evidence..... | 23 |
| 3.2.3 | A Priori Grounds..... | 23 |
| 3.2.4 | Philosophy Involves Practicing CT Skills..... | 24 |
| 3.2.5 | Instruction by Experts..... | 25 |
| 3.2.6 | Adequacy of Framing the Problem in These Terms..... | 26 |
| 3.3 | Problems with the Arguments..... | 26 |
| 3.3.1 | The Argument from Conventional Wisdom..... | 27 |
| 3.3.2 | The Argument from Anecdotal Evidence..... | 28 |
| 3.3.3 | The Argument by Definition..... | 28 |
| 3.3.4 | The Argument that Philosophy Provides the Right <i>Practice</i> | 29 |
| 3.3.5 | The Argument from the Expertise of Philosophers..... | 30 |
| 3.3.6 | Summation..... | 31 |
| 3.4 | The Need for Empirical Scrutiny..... | 33 |
| 3.4.1 | This is an Empirical Question..... | 33 |
| 3.4.2 | Two problems for the investigator..... | 33 |
| 4 | Review of the Existing Evidence..... | 35 |
| 4.1 | A Review of the Literature..... | 36 |
| 4.2 | Evidence from Primary and Secondary Students..... | 36 |
| 4.2.1 | The Effectiveness of Philosophy for Children..... | 37 |
| 4.2.1.1 | The Evidence and its Characteristics..... | 38 |
| 4.2.2 | Criticisms of Philosophy for Children’s Empirical Research..... | 39 |
| 4.2.2.1 | Statistical Considerations:..... | 40 |
| 4.2.3 | Lipman and Null Hypothesis Testing..... | 41 |
| 4.3 | Evidence from Undergraduate Students..... | 42 |
| 4.3.1 | Studies Bearing Directly on the Topic..... | 43 |
| 4.3.1.1 | Divergent Findings on Under-Graduates..... | 43 |
| 4.3.2 | Studies Bearing Indirectly on the Topic..... | 45 |
| 4.4 | Evidence from Graduate Students..... | 48 |
| 4.5 | General Conclusions..... | 50 |

| | | |
|---------|---|----|
| 5 | Meta-Analytical Review | 51 |
| 5.1 | The Need for a Meta-Analysis | 51 |
| 5.2 | The Concept of Meta-Analysis..... | 52 |
| 5.2.1 | Meta-analysis vs. Literature Review or Vote-Counting | 53 |
| 5.2.2 | Meta-Analysis Technically Challenging..... | 56 |
| 5.3 | Meta-Analysis of the Field..... | 57 |
| 5.3.1 | Defining the Research Questions:..... | 58 |
| 5.3.2 | Study Selection Criteria:..... | 58 |
| 5.3.2.1 | Independent Variables:..... | 58 |
| 5.3.2.2 | Dependent Variable: | 60 |
| 5.3.2.3 | Research Respondents (Subjects):..... | 60 |
| 5.3.2.4 | Research Methods:..... | 60 |
| 5.3.2.5 | Publication Types:..... | 61 |
| 5.3.3 | The search strategy:..... | 61 |
| 5.3.3.1 | Internet Databases for Published Empirical Studies. | 61 |
| 5.3.3.2 | Indexes of Relevant Research Journals:..... | 62 |
| 5.3.4 | Code study features of relevance..... | 62 |
| 5.3.5 | Statistics Information | 64 |
| 5.4 | Results of the Meta-Analysis | 68 |
| 5.4.1 | To what extent do critical thinking skills increase for students studying Anglo-American analytic philosophy? | 72 |
| 5.4.2 | To what extent do critical thinking skills increase for students studying subjects other than Anglo-American analytic philosophy? | 72 |
| 5.4.3 | To what extent do critical thinking skills increase for students studying CT, either as a philosophy subject or outside philosophy? | 73 |
| 5.4.3.1 | CT improvement for students studying CT in philosophy:..... | 73 |
| 5.4.3.2 | CT improvement for students studying CT outside philosophy: | 73 |
| 5.4.4 | <i>Relevant Comparisons:</i> | 74 |
| 5.4.4.1 | Is philosophy better than other subjects? | 74 |
| 5.4.4.2 | Is philosophy better than CT courses? | 75 |

| | | |
|-----|--|-----|
| 6 | General Discussion..... | 81 |
| 6.1 | Question one: Does (Anglo-American analytic) philosophy improve critical thinking skills? 81 | |
| 6.2 | Question Two: Does (Anglo-American analytic) philosophy improve critical thinking skills over and above university education in general?..... | 84 |
| 6.3 | Question three: Do critical thinking courses make a difference to critical thinking skills, whether or not such courses take place within the discipline of philosophy?..... | 86 |
| 7 | Conclusions | 90 |
| 7.1 | Summary of the case | 90 |
| 7.2 | Future Directions for Research | 91 |
| 8 | References in the Text..... | 93 |
| 9 | References used in the Meta-Analysis | 97 |
| 10 | APPENDICES..... | 101 |

5 Meta-Analytical Review

In chapter four, I conducted an informal review of available empirical evidence regarding the impact of philosophy on the development of critical thinking skills. The conclusions showed that the available evidence overall does not make a compelling case that philosophy improves CTS. There are two main reasons for this. On the one hand, the findings of different investigations of the matter are quite divergent. On the other hand, it is difficult to compare and reconcile these divergent findings. These problems are especially notable with regard to undergraduate studies, which are our main concern.

This chapter presents a meta-analytical review. This is a new methodological approach that allows us to integrate divergent findings. Thereby, more solid conclusions can be drawn about the impact of philosophy on the development of critical thinking skills. This chapter has been divided into three sections. First, the argument that supports the need for a meta-analysis; second, the meta-analysis in itself; third, the results.

5.1 The Need for a Meta-Analysis

What makes it so difficult to compare and reconcile the divergent findings and thus reach a determinate conclusion? First, the studies done so far used different instruments for measuring critical thinking skills. If we consider only those few studies that measured the impact of philosophy on CTS, there is not actually much difference in the instruments used. However, to be able to determine if philosophy improves critical thinking skills over and above university education in general, it is necessary to compare these results to those yielded by non-philosophy courses. And it is in this pool of studies where the greater diversity of measuring instruments appears.

Why does it matter that various investigators used different instruments to measure CTS? It is because different measuring instruments generate results that are recorded according to different scales. This makes them inherently difficult to compare. For instance, the California Critical Thinking Skills Test (CCTST) uses a scale of 34 points while the traditional Watson-Glaser Critical Thinking Appraisal (WGCTA) uses a scale of 80 points. Thus, while a twenty point difference on the WGCTA scale might look bigger than a ten-point difference on the CCTST scale, it may actually be smaller.

Second, the studies done so far, whether of philosophy or non-philosophy, often used different research designs. Undergraduate studies that have measured CTS are diverse. Although they have in common the goal of measuring CTS, they have used different research

questions and different methods to obtain their answers. Many key variables differ from one study to another. For instance, different statistical tools are used to validate results: p-values, t-tests, analysis of covariance. Or again, different critical thinking teaching strategies - lectures, debates, the questioning technique, argument mapping – are under examination. Then there are different measuring instruments (quantitative, and qualitative); and different methodological designs (longitudinal and cross sectional). Finally, there are different sample sizes, as well as different methods to select subjects from the samples. This heterogeneity of crucial variables makes the task of comparing and reconciling the results of the various studies exceptionally difficult.

Third, and not surprisingly, the findings of these divergent studies are themselves divergent. In the case of the studies that have measured the impact of philosophy on CTS, some studies show encouraging results (Ross & Semb 1981, Harrell 2004), some show negative results (Facione 1990, Reiter 1994) and some show inconclusive results (Annis & Annis 1979). On the other hand, those studies that have measured the impact of non-philosophy courses on CTS, while they contain many suggestions about how to promote critical thinking, provide limited evidence regarding the effectiveness of specific strategies. The findings of these studies of non-philosophy courses are also divergent and are difficult to compare and reconcile for the very same reasons that make the studies of the impact of philosophy on CTS difficult in this regard (Mc Millan, 1987; Pascarella & Terenzini, 2005; Williams, 2001).

The present study aims to reach more confident, if not yet definitive conclusions about the following research questions:

- Does philosophy improve critical thinking skills?
- Does philosophy improve critical thinking skills over and above university education in general?
- Do critical thinking courses as such improve critical thinking skills more than philosophy or university education in general?

To answer these questions and draw sound conclusions, it is necessary to be able to integrate the divergent findings in the literature. This is what a meta-analysis enables us to do and it is why one needs to be conducted. To better understand why it is needed, we need to understand what a meta-analysis actually is and how it standardizes different measures through the calculation of effect sizes.

5.2 The Concept of Meta-Analysis

A meta-analysis is a quantitative technique used to summarize, integrate, and interpret selected sets of scholarly works that address a similar outcome. It has an important, but somewhat circumscribed domain. First, it applies only to empirical research studies. Second, it

applies only to research studies that produce quantitative findings; i.e. studies using quantitative measurement of variables and reporting descriptive or inferential statistics to summarize the resulting data. Third, meta-analysis is a technique for encoding and analysing the statistics that summarize research findings as they are typically presented in research reports (Lipsey & Wilson, 2001). In short, a meta-analysis is the statistical analysis of the overall findings of a set of empirical studies (Glass, Mc Graw, & Smith, 1981).

A meta-analysis is able to compare and reconcile divergent findings by means of the calculation of what is known as an *effect size* (ES). An effect size is a statistic that encodes the critical quantitative information from the results of each relevant study. It produces a statistical standardization of these results. This standardization enables us to interpret the results of various studies in a consistent fashion across all the variables and measures involved.²⁵

In short, the ES standardizes divergent findings, because it can represent them on the same scale. It gives us a common reference point by which to compare and reconcile the divergent findings of different studies. To accomplish this, it is necessary to transform the measures of interest into the same statistical terms, namely, “standard deviation units” (Hunt, 1997, p.30).

5.2.1 Meta-analysis vs. Literature Review or Vote-Counting

There simply is no method other than a meta-analysis which enables us to achieve a common reference point such as that provided by the determination of effect sizes. Or, to be more precise, there is no method that enables us to do so with the same degree of precision. In most fields of science, the standard ways of dealing with a multiplicity of studies and divergent findings have been the *literature review* and the *vote-counting* technique (Hunt, 1997). However, both techniques are inadequate to achieve a common reference point.

A literature review, for example, the classical means for comparing divergent studies, provides many advantages for the reader: a convenient source of references, a conceptual orientation to the field, a discussion of methodological strengths and weaknesses found in published studies, a summary of major findings in the data, suggestions for building explanatory theory, and an invitation to explore primary sources for additional information (Wittrock, 1986). However, this older way of summarizing information yields, *at best, an impression* as to what

²⁵ The term ‘effect size’ was coined by Gene Glass (1976), one of the first exponents of meta-analysis. (Hunter, 1982). Seeking to determine the effectiveness of psychoanalysis, Glass found a great variation of outcome measures in a total of 375 studies of the matter. In order to reconcile these various measures, Glass realized that he needed to do a ‘meta-analysis’ of the analyses they presented. He came up with a means to ‘standardize’ them into a common coin, a common statistical unit, so that they could be added, averaged, divided, or otherwise manipulated. Glass called this unit the “effect size”, because it is a way of expressing the effects that different treatments had on scores. It is this method for resolving the kind of problem Glass had encountered which constitutes ‘meta-analysis’. (Hunt, 1997, p.30)

the literature is saying; without being able to rigorously weigh divergent findings that are to be found in it.

Equally, the technique of vote-counting is inadequate for our purposes. Vote-counting divides the studies of some treatment into two piles: those showing that the treatment worked, those showing that it did not; *the bigger pile being the winner* (Hunt, 1997). A major flaw is that, in vote-counting, every study counts as much as every other, even though one might be based on twenty cases, another on two thousand.

Common sense, as well as elementary statistical theory, tells us that we cannot have as much confidence in the findings of a small sample as those of a large one, since the likelihood of sampling error is higher for small samples. (p.23). In addition, this technique does not measure the size of the effect, in any given study, of one variable on another. If, for example, sample sizes are large, we may correctly conclude that, taken together, the studies reveal a statistically significant positive effect, but we will still have failed to show how great the *average* effect is. (p.25)

A meta-analysis, by contrast with these techniques, not only enables us to calculate the effect size, as an objective measure of findings across studies, but also provides a measure of both the *magnitude* and the *direction* of a relationship. The magnitude is the size of the effect that one variable has on another. The direction, on the other hand, indicates whether that causal relationship is positive or negative. In this case, the relationships being measured are those between the study of philosophy and the development of CTS, between the study of non-philosophy courses and the development of CTS. These outcomes of a meta-analysis enable us to reach a better grounded conclusion than can be provided by the alternative techniques. The argument for the need of a meta-analysis is represented in Figure 3.

Figure 3. The need for a meta-analysis

5.2.2 Meta-Analysis Technically Challenging

In spite of the advantages that a meta-analysis offers, it is technically quite a challenge. One reason for this is that it is sensitive to the GIGO (garbage in, garbage out) effect.²⁶ The worry is that in combining or integrating studies, one is mixing apples and oranges. This can happen in either of two main ways. First, meta-analyses can attempt to integrate studies that don't deal with the same constructs or terms. A second and perhaps more troubling issue is the mixing of study findings of different methodological quality in the same meta-analysis.

As regards constructs and terms, there would be little sense in calculating effect sizes for differences in, for example, academic achievement, social skills, and athletic performance. This would, of course, represent an extreme case of comparing apples and oranges. (Lipsey & Wilson, 2001). More subtle cases can readily be imagined. The problem arises, for instance, in those meta-analyses in which one is trying to compare findings regarding ambiguous or ill-defined variables. One thinks of things such as "progressive education", "teacher warmth" or "pupil self-esteem".

Meta-analyses can generate different results, depending on which kinds of study are used for the mix. Also, data samples for any meta-analysis will mean different things, depending on whether the collection of such data has been based on strictly the same concept or operation, rather than only vaguely or approximately the same ones. (Wittrock, 1986). The consistency and reliability of studies can all too easily be confused by vagueness in the definition of key terms or problems.

Clarifying what is meant by key terms and problems is important, then, if we are to avoid comparing apples with oranges. The quality of methodologies used in different studies is even more important in this regard. A meta-analysis can include both high-quality and lesser-quality studies, but this runs a definite risk of comparing apples with oranges. Considerable care must be exercised, therefore, in discriminating between studies of variable methodological quality. This is all the more so because, in many areas of research, especially those that deal with applied topics (as is the case with CTS) there are genuine methodological challenges in conducting studies at all.

To overcome the problem of comparing apples with oranges requires, firstly, that one decide, from the outset, what standard of rigor one is seeking through the meta-analysis. One then needs to proceed in a manner consistent with this decision. The GIGO effect will be avoided

²⁶ The garbage in, garbage out effect (abbreviated to GIGO) is an aphorism in the field of computer science. If a computer is given incorrect data, incorrect results will be produced. In the same way if we mix apples and oranges in a meta-analysis, the results will be of little value.

here just to the extent that one subjects to a meta-analysis the findings of studies that can, in fact, be meaningfully compared. This means that they must be both conceptually and methodologically comparable. To this end, the meta-analyst must make explicit judgments about the eligibility criteria for inclusion of studies in the meta-analysis.

It is worth noting that this problem of comparing apples and oranges is not peculiar to meta-analysis. Both the literature review and the vote-counting techniques are beset by this problem. Indeed, they are more susceptible to it than is meta-analysis, since they lack any systematic method for overcoming it. Not only does meta-analysis, by its very nature, entail a reconsideration of the comparability of different studies, but it also requires that each step be documented and open to scrutiny.

As one authority has written, "meta-analysis represents key study findings in a manner that is more differentiated and sophisticated than conventional review procedures that rely on qualitative summaries or vote-counting." (Lipsey & Wilson, 2001) Seen in this light, a meta-analysis is less a matter of comparing apples and oranges than of addressing precisely the tendency in other comparative methodologies to do just this.

Somewhat surprisingly, given these benefits of meta-analysis, as a means for checking the efficacy of studies, no meta-analysis seems to have been done so far to measure the impact of the study of philosophy on the development of CTS. It seems conceivable, even probable, that this oversight is due to the commonly accepted assumption that philosophy, of its nature, not only helps develop critical thinking skills but does so more directly than other disciplines.

This assumption, as we have said, is not unnatural. What is interesting is that it should have gone for so long without being sceptically or rigorously tested. It pretty clearly constitutes the conventional wisdom. What is required, though, is a rigorous process for examining the basis of that conventional wisdom. Meta-analysis, for the reasons I have given, is the best kind of process currently available for attempting this.

5.3 Meta-Analysis of the Field

A meta-analysis requires the following steps (Lipsey and Wilson, 2001):

1. Define the research questions
2. Define the study selection criteria
3. Define the search strategy
4. Code the study features of relevance
5. Specify statistical procedures
6. Report the results.

5.3.1 Defining the Research Questions:

The three major research questions addressed in this thesis are:

- Does (Anglo-American analytic) philosophy improve critical thinking skills?
- Does (Anglo-American analytic) philosophy improve critical thinking skills over and above university education in general?
- Do critical thinking courses make a difference to critical thinking skills, whether or not such courses take place within the discipline of philosophy?

Answering these questions requires us to address a number of more specific statistical questions, questions which can be answered via a meta-analysis:

- To what extent do critical thinking skills increase for students studying Anglo-American analytic philosophy?
- To what extent do critical thinking skills increase for students studying subjects other than Anglo-American analytic philosophy?
- To what extent do critical thinking skills increase for students studying CT, either as a philosophy subject or outside philosophy?

5.3.2 Study Selection Criteria:

There are 5 criteria for including studies in this meta-analysis: independent variables, dependent variables, research respondents, research methods and study publication types (Lipsey and Wilson, 2001).

5.3.2.1 Independent Variables:

To address the first question, eligible studies must involve the use of formal instruction in Anglo-American analytic philosophy for undergraduate students. Philosophy departments in universities in the English-speaking world typically offer such instruction. These undergraduate courses deal with core philosophical ideas, the clarification of concepts, the analysis of arguments, and the inculcation of a critical attitude. Some examples of these courses might include: Ethics, Introduction to Philosophy, and the like. These courses have been grouped in this study under the name of 'Pure Philosophy'. Any given student might take one or more than one such course, over one semester, two semesters, or whole degree.

In this sense, the independent variable is the *amount* of philosophy instruction that the students receive. For the purpose of comparing the impact of philosophy courses with that of courses in other disciplines, and thus to address the second question, eligible studies also include the impact of formal instruction in non-philosophy undergraduate courses; for instance, courses in literature, history, languages, nursing or the basic physical sciences. Such courses

may or may not include elements of specific CT instruction. Where they do not, they might usefully be categorized as 'No Phil, No CT' courses. Naturally, students might study one or more than one of these disciplines and do so in greater or lesser depth over the course of their university studies. Consequently, it is important to take into account, in any given instance, the amount of study in a non-philosophy course, including no-CT courses, relative to the putative impact on the development of CTS.

To address the third question, the pool of eligible studies must include undergraduate critical thinking (CT) courses. The independent variable here is *the amount of CT instruction* that the students receive. Two broad types of CT course are considered: (1) CT courses offered by philosophy departments, (2) CT courses offered by Non-philosophy departments.

(1) CT courses offered by philosophy departments are divided into three groups. First, courses dedicated to explicit instruction in CT, but without the use of argument mapping. Such courses utilize *traditional* didactic techniques, such as lectures, discussions, questioning techniques and the like. These courses have been categorized as "Phil CT, No AM". Second, courses dedicated to such instruction, but *including* the use of argument mapping. This technique enables students to represent and grasp the logical structure of informal reasoning in a visually explicit, diagrammatic form. Such instruction constitutes a marked departure from traditional didactic approaches to teaching CT. These courses have been grouped under the name of "Phil CT AM". Third, courses teaching CT which emphasize dedicated practice in argument mapping and require the students to do substantially more of it than do Phil CT AM courses. These courses are distinct from the second group because of the particular emphasis on the amount of practice in argument mapping that the students receive and the correlation between the amount of practice and the improvement in CT. For this reason, the third group of courses has been called philosophy with Lots of Argument Mapping Practice (Phil LAMP).

(2) CT courses offered by Non-philosophy departments are divided into two general groups. First, there are courses exclusively dedicated to promoting critical thinking skills ("No Phil, Ded-CT" courses), for instance, "Introduction to Reasoning", "Informal Logic", "Critical Thinking", or "Analysis of Information". Second, there are courses that have been designed and implemented to promote other abilities and knowledge, but with the inclusion of some pedagogical strategies intended to accelerate the growth in the students' CTS ("No Phil, Some-CT" courses). These would include courses such as nursing, classics and history, psychology, politics and sociology, or mathematics. The didactic techniques implemented in such courses might vary from the use of software, critical writing and reading, to debates, analysis of information, argumentation, and exercises in clear reasoning. Any such course must be of at least one-semester's duration to be eligible.

In summary, we classified the studies into seven groups. These groups will make it possible for us to measure the impact of the two major independent variables selected for the purposes

of this inquiry: the amount of philosophy and CT instruction the students have received. These groups are as follows:

1. Courses offered by philosophy departments consisting of formal instruction in Anglo-American analytic philosophy, or what I shall call 'pure philosophy' (Pure Phil).
2. Critical thinking courses offered by philosophy departments with no instruction in argument mapping (Phil CT No AM).
3. Critical thinking courses offered by philosophy departments with some instruction in argument mapping (Phil CT-AM).
4. Critical thinking courses offered by philosophy departments with lots of argument mapping practice (Phil LAMP). These are courses fully dedicated to teaching CT with argument mapping.
5. Courses offered by non-philosophy departments and wholly dedicated to explicit instruction in CT (No Phil, Ded-CT).
6. Courses offered by non-philosophy departments with some form of conventional CT instruction embedded (No Phil, Some-CT).
7. Courses offered by non-philosophy departments with no special attempts being made to cultivate CT skills (No Phil, No-CT).

5.3.2.2 Dependent Variable:

The purpose of this meta-analysis is to examine the effect of philosophy and CT instruction on students' critical thinking skills. Therefore, the dependent variable in this study is critical thinking skills gain.

5.3.2.3 Research Respondents (Subjects):

Since it is the CT skills of undergraduate students that we are seeking to assess, only studies of undergraduate students, not graduate students or pre-university students are eligible for inclusion in the meta-analysis.

5.3.2.4 Research Methods:

The studies to be included in this meta-analysis are only those that report *quantitative* results of efforts to measure CT skills. Such measures must, in turn, be about demonstrable *abilities*, rather than simply the *dispositions* of students, or their attitudes toward critical thinking. Also, to calculate the overall effect size, eligible studies must provide sufficient statistical data: pre and post test means, standard deviations, and sample sizes. Alternatively, they must report

sufficient information to allow the gain in CT, expressed as an effect size in appropriate SD units, to be calculated. This calculation of effect size is discussed below.

To assure that the studies included were of high methodological quality, it was determined that they must have used a pre-test, post-test (longitudinal) research design. A pre-post design compares the central tendency (e.g. mean or proportion) on a variable measured at one time with the central tendency of that same variable measured the same way *on the same sample* at a later time. Further, as a standard of empirical rigor, such studies must have used objective multiple-choice tests of critical thinking.

5.3.2.5 Publication Types:

To help counteract the file-drawer effect, whereby only positive results get published while negative ones are left in the filing cabinet, both published and unpublished studies have been considered eligible in this inquiry. They might include journal articles, dissertations, technical reports, unpublished manuscripts, conference presentations, and the like.

5.3.3 The search strategy:

Multiple strategies were used to ensure the collection of the widest possible pool of existing studies. These strategies included internet databases, relevant research journals, the reference lists of published studies, email communication with CT interest groups and known CT researchers; and web publication of the list of studies, accompanied by contact details and an invitation to contribute.

5.3.3.1 Internet Databases for Published Empirical Studies.

Engines targeting philosophy, education, psychology, and social science journals were all utilised, including:

- Current Issues in Education
- Current Contents
- Dissertation Abstracts
- ERIC
- Expanded Academic ASAP
- Google Scholar
- JSTOR
- Philosopher's Index

- Project Muse
- Psych INFO
- Research in Education
- Social Sciences Plus Education Complete (ProQuest 5000)
- Web of Science

5.3.3.2 Indexes of Relevant Research Journals:²⁷

- Current Issues in Education
- Informal Logic
- Research in Education
- Teaching philosophy

Keywords were selected with the assistance of two research librarians. Three different groupings of keywords were combined on the databases: critical thinking, higher education, research design. Keyword terms in the critical thinking grouping were: critical thinking skills, critical thinking gain or growth. This CT grouping also included searches for instruments designed to measure this construct (e.g. California Critical Thinking Skills Test, Watson-Glaser Critical Thinking Appraisal, Cornell Critical Thinking Test, and the Collegiate Assessment of Academic Proficiency). Keyword search terms for the higher education grouping were: undergraduate, college, university, and postsecondary. Keyword terms in the research design category were: longitudinal, pre-test post-test or pre-post test.

5.3.4 Code study features of relevance

For each potentially relevant study, the features shown in this section were coded. Table # , in Appendix A, sets out the study characteristics, the course information, and the research method information for each of these studies. Table #, in Appendix B, divides the pool of studies into the seven groups established in the section “Study Selection Criteria” (5.3 ‘Meta-analysis of the Field’). This table also displays the statistical information (sample sizes, pre and post test means, standard deviations and CT gain) for every available study; and also the effect sizes calculated following the two methods used in this thesis (the study SD and the test SD).

The Study Characteristics

Research Number (e.g. Adams99)

²⁷ In several of these categories, I am very much indebted to the work of Dr Melanie Bissett, on whose work I am grateful to have been able to draw in the course of the present inquiry.

Study Identification (e.g. Adams99-1, Adams99-2)

Article & Source's Name (author's name, year of publication, article's title, source's name)

Status

Published (P)

Unpublished (UNP)

Type of Publication:

Book/Book Chapter (B)

Journal Article (JA)

Thesis/Dissertation (T/DISS)

Technical Report (TR)

Other

Course Information

Undergraduate course's name (e.g. Introduction to Philosophy, Nursing, History).

Philosophy category

Philosophy courses (Phi)

Non-philosophy courses (Non-Phi)

Critical Thinking category

Dedicated-Critical Thinking (Ded-CT)

Some -Critical Thinking (Some-CT)

No-Critical Thinking (No-CT)

Argument Mapping category

CT courses with some argument mapping (AM)

CT courses with lots of argument mapping practice (LAMP)

CT courses without argument mapping (No AM)

Educational level of subjects (e.g. freshman, sophomore, etc.)

Teaching features (e.g. traditional philosophy, questioning technique, computer based course)

Research Method Information

Interval (e.g. 1 semester, 1 year, 2 years)

Test

California Critical Thinking Skills (CCTST)

Cornell Critical Thinking Test (Cornell)

Watson-Glaser Critical Thinking Appraisal (Watson-Glaser)

Collegiate Assessment of Academic Proficiency (CAAP)

Test of Critical Thinking (TCT)

Graduate Skills Assessment (GSA)

Home made test (Home made test)

Sampling Procedure

Experimental (Randomized) (Exp)

Quasi-experimental (QExp)

Methodological Design

Within group (WG)

Information to calculate Effect Size

Sample Size

Pre Test Mean

Post Test Mean

Standard Deviation Pre Test

Standard Deviation Post Test

5.3.5 Statistics Information

Effect sizes (ESs) and their confidence intervals were calculated for every study included in the meta-analysis. It is important to make clear that there is more than one possible answer about how to calculate effect sizes, but some answers are better than others. The variance in outcome is due to the fact that an ES is not the calculation of an absolute quantity, but only a reasonable estimation of the magnitude and strength of a *relationship* between two variables (Cooper, 1998, in(Gellin, 2003)). The difference between the methods to calculate effect sizes lies in which standard deviation (SD) you wish to use as the ES measuring unit.

Here we have used two measuring units, that is to say two methods to calculate an ES. The first method, a widely used one, is known as the Standardized Mean Difference approach. It uses as the measuring unit the average SD of the pre- and post-test scores reported from each study. The second method uses the SD for all students about which there is information, using a particular test. In this method, we group all the studies by the critical thinking *test* used in them to estimate standard deviations. We regard this second method as a better method than the first one for estimating the SD for the whole population of potential students. For future reference, in this study, the two methods will be referred to as the Study SD and the Test SD, respectively. In each case, a weighted mean ES is calculated for each category of study.

The study SD:

The Study SD is generally known as 'Cohen's d '. It is defined in various ways by different authors, but the most common usage is to regard it as an effect (in original units), divided by an appropriate SD. In this case, we have chosen as the SD unit, the average standard deviation of the pre- and post-test scores reported from each study.

This Standardized Mean Difference method is traditionally one of the methods most used in meta-analysis. This method takes a standard deviation representative of the population from which the study sample was taken. For this reason, it uses the standard deviations reported from each individual study.

The following formulas were employed to calculate effect sizes for individual studies and overall effect sizes for each group of studies:

Effect size (the Cohen's d) for individual studies:

$$d = (\text{mean post-test} - \text{mean pre-test}) / \text{average SD}$$

Where average SD is the average of the pre-test and post-test standard deviations, calculated for a particular study. The formula used was:

$$\text{Average SD} = (\text{SD post-test} + \text{SD pre-test}) / 2$$

Results from studies with pre-post intervals exceeding one semester were divided by the number of tested semesters to ascertain a single semester effect. We are assuming equal improvement over the semesters.

SDs were taken from the individual studies when such studies reported them. However, there were cases in which the SDs had to be derived from other data presented in the studies. Among these kinds of data were t-tests, P-values, and ranges. Those studies for which SDs

were derived from other data are marked with an asterisk in the tables.²⁸ (See Appendix C, for a brief description of these calculations.)

Overall effect size:

The overall Effect Size for *each category* of study was calculated using a weighted average, where individual study Effect Sizes were weighted by sample size, since larger samples provide a better estimate of population values than small samples. This weighting of d values by sample size amounts to the weighting of studies by their inverse variances, as is standard practice in meta-analysis (Lipsey & Wilson, 2001)

The formula used was:

Overall effect sizes: $\sum [d_i \times (n_i / \sum n_i)]$, where d_i is the standardised ES for study i , n_i is the sample size for study i , \sum is the sum over $i=1$ to k , and where k is the number of studies in the group.

Confidence Intervals:

For each group of studies, the 95% confidence interval (CI) was calculated by estimating the SD of d for that group, then using this to calculate the margin of error of the CI (i.e., the length of one arm of the CI).

The SD of d for a group of k studies was estimated as:

$$SD_d = \sqrt{\frac{\sum_{i=1}^k n_i (d_i - \bar{d})^2}{\sum_{i=1}^k n_i}}$$

Then the margin of error of the CI was calculated as $t_{95,k-1} SD_d / \sqrt{k}$ where $t_{95,k-1}$ is the critical value of t for a 95% CI, for $(k-1)$ degrees of freedom. Note that this method of calculating the CI does not assume population effect size is homogeneous over studies within a group, and is thus a conservative (and realistic) way to show a CI that gives a good indication of where overall population mean ES is likely to lie, for the whole population of potential studies of the particular type of course.

Despite the fact that the Standardized Mean Difference is one of the more commonly used ES measures, there are some criticisms of it as a procedure. Perhaps the most important such

²⁸ For these calculations, I obtained the help of the Statistical Consulting Centre at the University of Melbourne and, in particular, of Dr. Sue Finch.

criticism, at least for present purposes, is that, when we standardize the effect using SDs derived in this manner, the estimate of standardised ES is influenced by sampling error in the SD, as well as sampling error in the mean difference. Greenland et al. argue that this error in estimating standard deviation makes standardised ES, calculated using SD from individual studies, an unacceptable measure (Greenland, 1986).

To briefly illustrate this point, let me give an example provided by Dr. Sue Finch from the Statistical Consulting Centre at Melbourne University. Suppose we have the outcomes from two different studies which both use the same measure of critical thinking:

Study 1: mean change = 10, SD = 2

Study 2: mean change = 10, SD = 2.5

Both studies have the same effect - a change of 10 units. However, if we standardize the effect size, for the first study it is 5 and for the second study it is 4. Our estimate for the second study is less precise than the first, but the actual change is the same.

In order to minimize the sampling error that the Standardized Mean Difference procedure generates, we also calculated Effect Sizes using the Test Standard Deviations (Test SD) method. There are good reasons for believing that this yields the more reliable effect size estimates.

The test SD:

In order to calculate the SD for each test, we used the following procedure:

We divided the studies into critical thinking test categories. 7 test categories in total, each representing one kind of test that had been used to measure critical thinking abilities. The seven categories were: the California Critical Thinking Skills Test (CCTST), Watson-Glaser Critical Thinking Appraisal (WGCTA), Cornell Critical Thinking Test (Cornell), Collegiate Assessment of Academic Proficiency (CAAP), Graduate Skills Assessment (GSA), the Test of Critical Thinking (TCT), and home made tests.

We collected all the data for each test to estimate a weighted standard deviation. The formula used was: Test SD: $\sum [\text{average } SD_i \times (n_i / \sum n_i)]$ where average SD_i is the average SD for study i , as defined above, and the summation is over all studies in a particular test category.

We used the test SD to calculate new effect sizes for individual studies and an overall effect size for each category. For this we used the same formulas employed in the study SD, but with test SD in place of the individual study SDs.

Before presenting the results, it is worth noting that this meta-analysis could be developed further statistically. Another analysis that could be performed is a homogeneity test to assess the homogeneity of the effect size distribution for each selected group. In a homogeneous distribution, any individual effect size differs from the population mean only by sampling error. In other words, if the variability of the effect sizes is larger than would be expected from sampling error there are differences among the effect sizes that have some source other than subject-level sampling error. (Lipsey and Wilson, 2001)

Basically, a homogeneity test indicates that one or more moderator variables are likely to be causing variance in the effect sizes. It is beyond the scope of this project to detect and analyse any such variables. (As noted earlier, however, our method of calculating the CI of a group of studies does not assume the absence of moderator variables.) In the present study, we have concentrated on trying to determine if the study of philosophy and, more broadly, the study of critical thinking, bring a change in the development of critical thinking skills in university students. For this purpose, a calculation of effect sizes that indicates the magnitude and direction of any change is sufficient. The detection of moderated variables which explain any heterogeneity between the effect sizes could, however, be an interesting subject for a future investigation based on the present study.

The statistics body of statistical data on which the foregoing meta-analysis has been based is set out in Appendix B. This includes, in several tables, the pool of studies concerning each of the seven groups selected for the thesis; the effect sizes for every available study meeting the criteria, using the formulas presented for both analyses (the study SD, and the test SD); and pre- and post-test means, standard deviations, sample sizes, and raw score gain.

5.4 Results of the Meta-Analysis

Fifty-two studies met the criteria for consideration in this meta-analysis. These studies reported a total of one hundred and nineteen research findings²⁹. These studies sought to measure the gain in university students' critical thinking skills by examining two key, independent variables. Those two variables were the amount of instruction in philosophy and instruction in CT over different intervals of time. The variations within these two types of instruction were categorized into seven groups of studies (see section 5.3.2, "Study Selection Criteria"). Although the time spent by the students in philosophy or CT instruction varied among

²⁹ For purposes of meta-analysis, a single *research finding* is a statistical representation of one empirical relationship involving the variables of interest to the meta-analyst, measured *on a single subject sample*. For instance, for those studies that used an experimental-control group research design, one research finding corresponds to the experimental group, and another research finding to the control group.

the studies, a single semester effect was calculated in all cases, in order to establish a basis for comparison between the groups.

The information in the fifty-two studies was coded, following the coding protocol indicated in section 5.3, A Meta-Analysis of the Field. Table 2, in Appendix A, shows the Master List of studies used in this meta-analysis. This Table sets out the study characteristics, course information, and research method information for each of these studies. Table 3, in Appendix B, divides the pool of studies into the seven groups established in the section “Study Selection Criteria” (see 5.3.2). This table also displays the statistical information (sample sizes, pre and post test means, standard deviations and CT gain) for every available study; and also the effect sizes calculated following the two methods used in this thesis (the study SD and the test SD).

To facilitate the presentation of the results in this section, the effect sizes calculated using the two methods are displayed in the following figures:

Figure 4. Chart of effect sizes calculated using the SD found in each study.

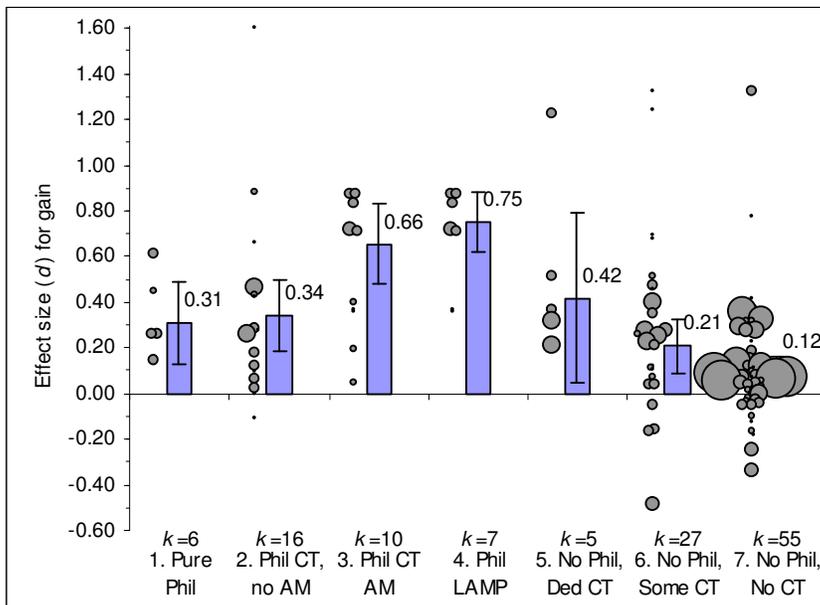
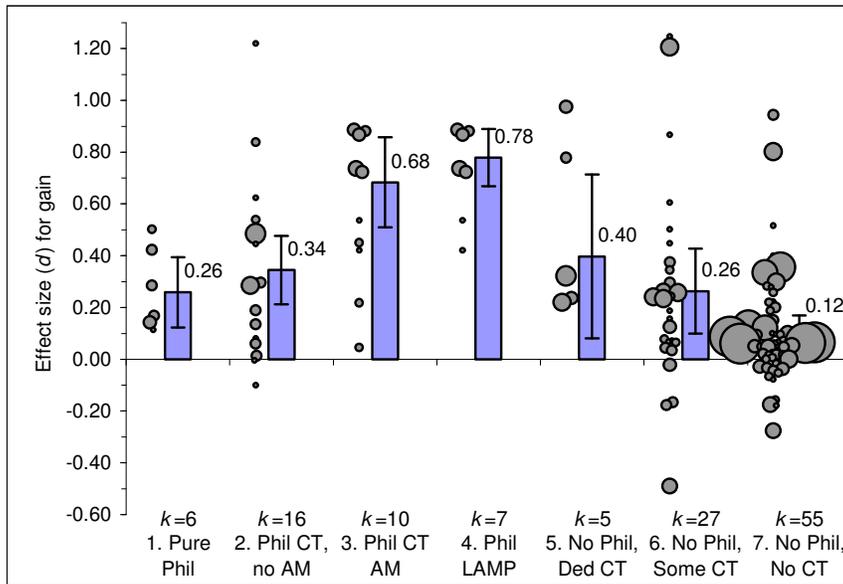


Figure 5. Chart of effect sizes calculated using our best estimates of SD for each test instrument.



It is interesting to notice that, in general, the results from the two approaches – the standardized mean difference (the study SD) and the best possible estimate (the test SD) – are not that different; indeed, the general pattern is one of great consistency. This indicates that the studies are not using radically different populations. However, this consistency of results between the two methods does not apply to all the studies, or more specifically to all the research findings of these studies. For example:

| Study-ID | N (sample) | Effect Size based on study SD | Effect Size based on test SD |
|--------------|------------|-------------------------------|------------------------------|
| Ross81-1 | 64 | 0.61 | 0.42 |
| Solon03-1 | 25 | 1.60 | 1.22 |
| Rimiene02-1 | 77 | 1.22 | 0.97 |
| Spurrett05-1 | 27 | 0.37 | 0.54 |
| Vieira97-1 | 26 | 0.28 | 0.50 |

In these examples, the difference between the effect sizes is considerable. At first glance, this difference is caused simply by the standard deviations used to calculate the effect sizes. In the first three examples (Ross81-1, Solon03-1, Rimiene02-1) the SDs used to calculate the ESs with the “Study SD” method are smaller than those estimated from the CT tests populations. Therefore, the sizes of the effects are greater. Conversely, in the last two examples (Spurret05-1, Vieira97-1) the SDs estimated using the “Test SD” method are smaller than those reported by the individual studies, yielding greater ESs.

The most likely reason that a small proportion of studies show such a difference between ES based on study SD, and ES based on test SD, is sampling variability. A few studies (especially small ones) are bound to give study SD values that happen to be a bit big or a bit small. This interpretation is strengthened if these studies are smaller than average, because the influence of sampling variability on study SD is greater for smaller n .

For the purposes of this thesis, the test SD method provides the best ES estimates. Therefore, any future references in this section to the results of the meta-analysis will refer to Figure 5 and the respective ESs that are displayed in it.

Let us remind ourselves, at this juncture, that the ES (effect size) is the point estimate of the magnitude of the effect of one variable on another. In the case of this study, we have analysed the impact of two main independent variables (philosophy instruction and CT instruction) on CT gain. Also, it is important to bear in mind that the range of values contained in the confidence intervals (CIs) provides an interval estimate of the true value of a parameter (ES) for the population.

The level of confidence is the probability of producing an interval containing the true value for the population. There is a single true value that we never know. The CI is an interval in which we are 95% confident the true value lies. I will be presenting and discussing standardized ESs, meaning that they are expressed in SD units. Also, I will use the convention that the range of values stated in any given CI means a 95% CI.

In what follows, I shall set out the results of the meta-analysis by addressing in turn each of the three questions which were asked at the beginning of the meta-analysis.

5.4.1 To what extent do critical thinking skills increase for students studying Anglo-American analytic philosophy?

The analysis of the three Groups of studies representing courses offered by philosophy departments (see columns 1, 2 and 3 in Figure 5) resulted in an ES of .45 SD; CI [0.37, 0.53]. This is the estimated CT gain over one semester for undergraduate students studying any philosophy course, whether or not including CT instruction.

These figures, however, give a misleading impression of the magnitude of the effect of Anglo-American analytic philosophy taken in itself, because they include CT instruction within the philosophy courses in question. To ascertain the real impact of philosophy in its own right, we must look at it in isolation from CT instruction. This is the importance of Group 1 (in Fig. 5) which represents “Pure Philosophy” courses.

The mean effect size of the pool of studies from this category yielded a value of **0.26 SD; CI [0.12, 0.39]**.

5.4.2 To what extent do critical thinking skills increase for students studying subjects other than Anglo-American analytic philosophy?

The analysis of the three Groups of studies (see columns 5, 6, and 7, in Fig. 5), representing courses offered by non-philosophy departments, resulted in a mean ES of 0.16 SD, CI [0.11, 0.21]. This is the estimated CT gain over one semester for undergraduate students studying non-philosophy courses, whether or not including CT instruction.

These figures, however, also give a misleading impression of the magnitude of the effect of these non-philosophy courses in their own right, because they, also, include some CT instruction. Once again, then, in order to ascertain the real impact of non-philosophy courses in their own right, we must look at them in isolation from CT instruction. This is the importance of the pool of studies in Group 7 (in Fig 5) “No Phil, No CT”.

The mean effect size of the pool of studies from this category yielded a value of **0.12 SD, CI [0.075, 0.17]**.

5.4.3 To what extent do critical thinking skills increase for students studying CT, either as a philosophy subject or outside philosophy?

In order to be able to discuss the effectiveness of philosophy Departments in teaching CTS, we must first distinguish between CTS courses taught within philosophy Departments and those taught in other departments.

5.4.3.1 CT improvement for students studying CT in philosophy:

Traditional CT. In this group we refer to the CT gain for those students taking traditional CT offered by philosophy departments. Traditional, in this case, means CT teaching using lectures and discussion, but excluding argument mapping instruction (see Group 2, “Phil CT, no AM”, in Fig 5). The analysis of the results for this Group yielded a value of 0.34 SD, CI [0.21, 0.48].

CT with some argument mapping: The CT gain for those students taking CT courses teaching some argument mapping (see Group 3, “Phil CT AM”, in Fig 5) is 0.68 SD, CI [0.51, 0.86].

CT with lots of argument mapping practice: The CT gain for those students taking CT courses teaching lots of argument mapping practice (see Group 4, “Phil LAMP”, in Fig 5) is 0.78 SD, CI [0.67, 0.89].

The combined effect: The combined effect of CT change for any philosophy CT course (traditional and argument mapping courses) yielded an effect size of 0.49, CI [0.39, 0.59], (see Fig. 5, columns 2 and 3).

5.4.3.2 CT improvement for students studying CT outside philosophy:

Traditional CT: Traditional CT includes two groups of courses: those with dedicated CT instruction and those consisting simply of some CT instruction. Analysis of the results for those students studying a dedicated CT course (see Group 5, “No Phil, Ded-CT” in Fig 5) yielded a value of 0.40, CI [0.08, 0.71]. Analysis of the results for Group 6 (see Group 6 “Some-CT courses”, in Fig. 5) yielded a value of 0.26 SD, CI [0.09, 0.43].

The combined effect: The combined effect of CT change for any Non-philosophy course with at least some CT (Groups 5 and 6 together) yielded an effect size of 0.30, CI [0.16, 0.43].

5.4.4 Relevant Comparisons:

In order to determine if philosophy does it better than other subjects and better than CT courses, we need to make relevant comparisons among the groups. The tests of statistical significance are the criteria to determine whether or not the difference from a comparison is important.

5.4.4.1 Is philosophy better than other subjects?

a) CT gain in philosophy vs CT gain in Non-philosophy courses:

This first comparison contrasts Groups 1, 2 and 3 taken together with Groups 5, 6 and 7 taken together, to examine the difference between the study of any kind of philosophy course (with or without CT) and any non-philosophy course (again, with or without CT). Analysis of the results shows that Group “All Phil” (1,2,3) yields an ES of 0.45, CI [0.37, 0.53] vs Group “All No Phil, any CT”(5,6,7) with an ES of 0.16 CI [0.11, 0.21]. The difference is statistically significant at $p < .01$.

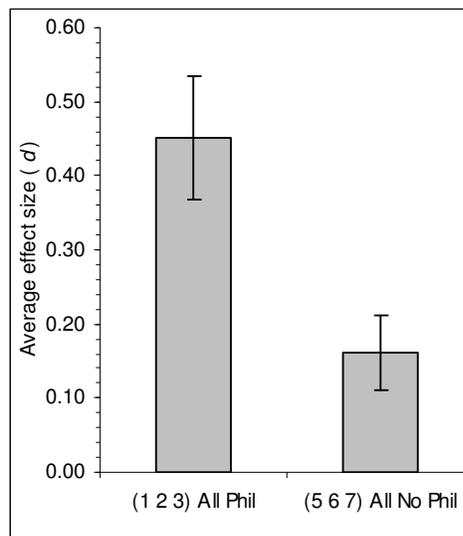


Figure 6. CT gain in philosophy vs CT gain in Non-philosophy courses.

b) CT gain in Anglo-American analytic philosophy vs No Phi No CT (1 vs 7)

Here we are concerned with whether philosophy as such, without any specialized CT component, actually makes any more difference to CTS gains than subjects other than pure (Anglo-American analytic) philosophy without CT. This compares Group 1 with Group 7.

Analysis of the results shows that the difference between the two is not statistically significant at $p < .05$. What does this tell us? The apparent difference between the two and the fact that the confidence intervals overlap only slightly, suggests that philosophy may make more difference; but we need better evidence before we can claim to know that this is the case.

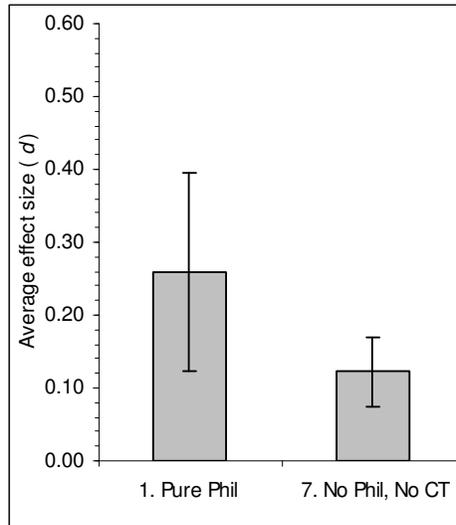


Figure 7. CT gain in Anglo-American analytic philosophy vs No Phi No CT

5.4.4.2 Is philosophy better than CT courses?

In this comparison, we are concerned with whether pure philosophy instruction (Group 1) makes more difference than CT instruction in its own right. There are various sub-comparisons to make here.

a) Anglo-American analytic philosophy (Group 1) vs all CT instruction in philosophy (Groups 2,3). The difference is statistically significant at p value $< .05$.

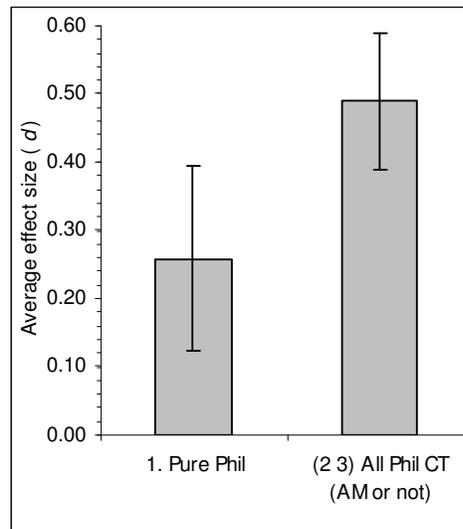


Figure 8. Anglo-American analytic philosophy vs all CT instruction in philosophy.

b) Anglo-American analytic philosophy (Group 1) vs Traditional CT in philosophy (Group 2).
 The difference is not statistically significant ($p = 0.435$) at p value $< .05$

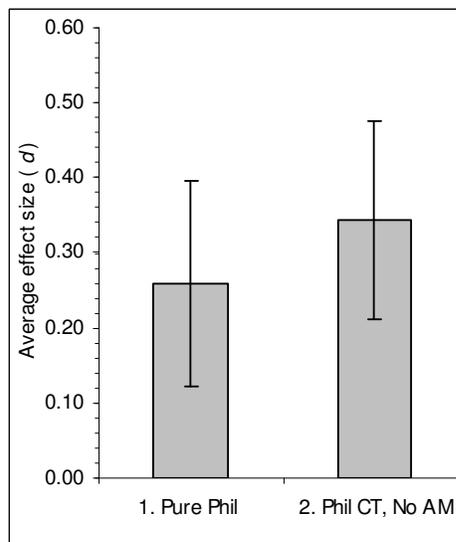


Figure 9. Anglo-American analytic philosophy vs Traditional CT in philosophy.

c) Anglo-American analytic philosophy (Group 1) vs CT instruction in philosophy with argument mapping (Group 3). The difference is statistically significant at p value $< .01$

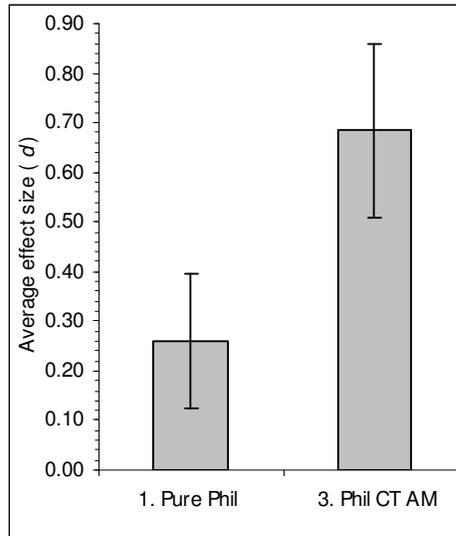


Figure 10. Anglo-American analytic philosophy vs CT instruction in philosophy with argument mapping.

d) Anglo-American analytic philosophy (Group 1) vs Traditional CT instruction outside philosophy (Group 5). The difference is not statistically significant ($p = 0.272$) at p value $< .05$

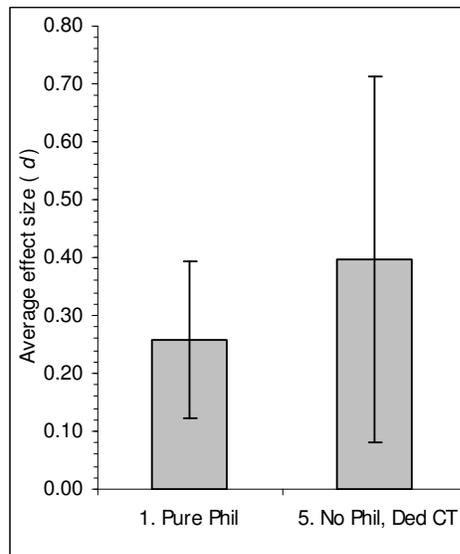


Figure 11. Anglo-American analytic philosophy (Group 1) vs Traditional CT instruction outside philosophy.

e) Anglo-American analytic philosophy (Group 1) vs all CT instruction in outside philosophy (Groups 5 and 6). The difference is not statistically significant (p value = 0.806) at $p < .05$

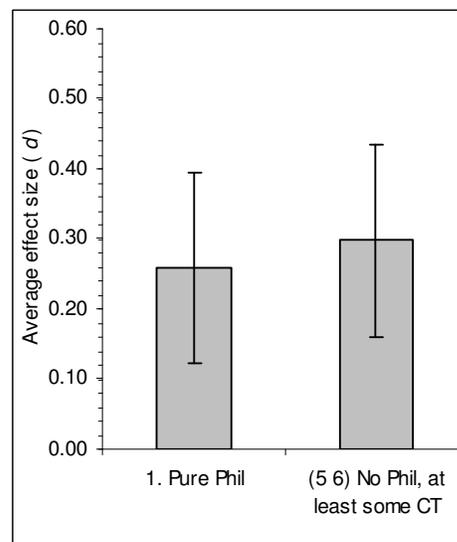


Figure 12. Anglo-American analytic philosophy vs all CT instruction in outside philosophy.

f) Anglo-American analytic philosophy (Group 1) vs Traditional CT in philosophy and outside philosophy (Groups 2 and 5). The difference is not statistically significant (p value = 0.324), at $p < .05$.

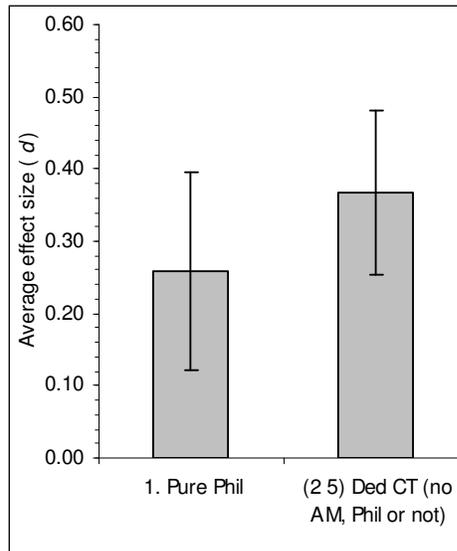


Figure 13. Anglo-American analytic philosophy vs Traditional CT in philosophy and outside philosophy.

g) All CT instruction in philosophy (Groups 2 and 3) vs No Philosophy, No CT instruction (subjects other than Pure Philosophy, Group 7). The difference is statistically significant at p value $< .01$

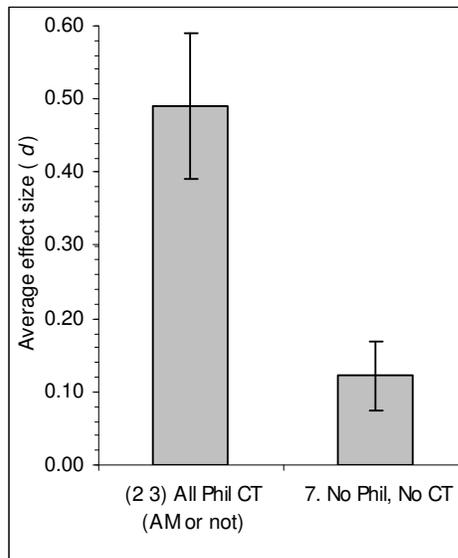


Figure 14. All CT instruction in philosophy (Groups 2 and 3) vs No Philosophy, No CT instruction.

h) All CT instruction outside philosophy (Groups 5 and 6) vs No Philosophy, No CT instruction (subjects other than Pure Philosophy, Group 7). The difference is statistically significant at p value $< .01$.

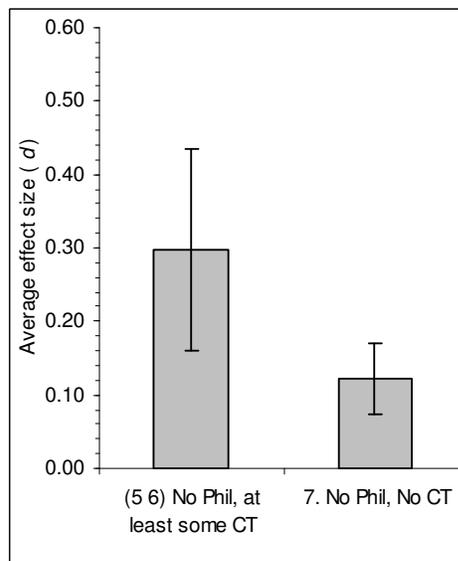


Figure 15. All CT instruction outside philosophy vs No Philosophy, No CT instruction.

8 References in the Text

- Annas, J. (2000). *Ancient philosophy : a very short introduction*. Oxford ; New York: Oxford University Press.
- Annis, D., & Annis, J. (1979). Does philosophy improve critical thinking?. *Teaching Philosophy*, 3(2).
- Audi, R. (1981). *Philosophy: A Brief Guide for Undergraduates*: American Philosophical Association. Available: <http://www.apa.udel.edu/apa/publications/texts/briefgd.html>.
- Audi, R. (1999). *The cambridge dictionary of philosophy*. Cambridge: Cambridge University Press.
- Bacon, F. (1905). *Novum organum*. London: Routledge.
- Blackburn, S. (1999). *Think : a compelling introduction to philosophy*. Oxford: Oxford University Press.
- Butchard. (2006). *The monash critical thinking study*: University of Monash. Available: <http://arts.monash.edu.au/phil/research/thinking/>.
- Cebas, E., Garcia Moriyon, F. (2003). What we know about research in philosophy for children. *Unpublished manuscript*.
- Cheng, P., Holyoak, K., Nisbett, R., & Oliver, L. (1986). Pragmatic versus syntactic approaches to training deductive reasoning. *Cognitive Psychology*, 18, 293-328.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. New York: Lawrence Erlbaum Associates, Inc., Publishers.
- Cohen, J. (1994). The earth is round ($p < .05$). *American Psychologist*, 49(12).
- Colman, A. M. (2006). *A dictionary of psychology* (2nd ed.). Oxford ; New York: Oxford University Press.
- Descartes, R., & Clarke, D. M. (1999). *Discourse on method and related writings*. London: Penguin Books.
- Dewey, J. (1910). *How we think*. Boston: D.C. Heath & Co.
- Donohue, A., Van Gelder, T., Cumming, G., & Bissett, M. (2002). *Reason! project studies, 1999-2002 (Tech. Rep. No. 2002/1)*: University of Melbourne.
- Edwards, P., & Pap, A. (1972). *A modern introduction to philosophy: Readings from classical and contemporary sources* (3 ed.). New York: Free Press.
- Ennis, R. (1991). *A Super-Streamlined Conception of Critical Thinking*: Available: <http://www.criticalthinking.net/SSConcCTApr3.html>.
- Ennis, R., Millman, J., & Tomko, T. (1985). *Cornell critical thinking tests level X & level Z- Manual*. Australia: Midwest Publication.
- Facione, P. (1990). *The california critical thinking skills test: College level. Technical report # 1*. California: Santa Clara University.
- Facione, P. (2006). *Critical thinking: What it is and why it counts*: Insight Assessment. Available: http://www.insightassessment.com/pdf_files/what&why2006.pdf.
- Fidler, F. (2005). *From statistical significance to effect estimation: Statistical reform in psychology, medicine and ecology*. Doctoral Dissertation. University of Melbourne.

- Fisher, A. (2001). *Critical Thinking: An Introduction*. Cambridge: Cambridge University Press.
- Fisher, A. (2004). *The logic of real arguments* (2nd ed.). Cambridge, U.K. ; New York: Cambridge University Press.
- Fisher, A., & Scriven, M. (1997). *Critical thinking: Its definition and assessment*. CRIT EdgePress.
- Fisher, R. (2003). *Teaching thinking : philosophical enquiry in the classroom* (2nd ed.). New York ; London: Continuum.
- Garcia-Moriyon, F., Rebollo, I., & Colom, R. (2005). Evaluating Philosophy for Children: A meta-analysis. *Thinking: The journal of Philosophy for Children*, 17(4).
- Garlikov, R. (2002). *Reasoning: What It Is To Be Rational*: Available: <http://www.akat.com/reasoning.htm>.
- Gellin, A. (2003). The effect of undergraduate student involvement on critical thinking: A meta-analysis of the Literature 1991-2000. *Journal of College Student Development*, 44(6).
- Giere, R. N. (1997). *Understanding scientific reasoning* (3rd ed.). Harbor Drive, Orlando: Holt, Rinehart and Winston.
- Gigerenzer, G. (1993). The superego, the ego, and the Id in statistical reasoning. In Keren, G. & Lewis, C. A Hanbook for data analysis in the behavioral sciences. Methodological Issues. In. New Jersey: Lawrence Erlbaum Associates.
- Glass, G., Mc Graw, B., & Smith, M. L. (1981). *Meta-analysis in social research*. California: SAGE Publications Inc.
- Greenland, S., Schlesselman, J. , Criqui, M. (1986). The fallacy of employing standardized regression coefficients and correlations as measures of effect. *Journal of Epidemiology*, 123 (2).
- Hadot, P., & Davidson, A. I. (1995). *Philosophy as a way of life : Spiritual exercises from Socrates to Foucault*. Oxford ; New York: Blackwell.
- Harrell, M. (2004). *The improvement of critical thinking skills in what philosophy is. Technical report CMU-PHIL-158*: Carnegie Mellon University.
- Harris, T. L., Hodges, R. E., & Association, I. R. (1981). *A dictionary of reading and related terms*. Newark, Del.: International Reading Association.
- Hatcher, D. (1999). Why critical thinking should be combined with written composition. *Informal Logic*, 19(2,3).
- Hatcher, D. (2001). Why Percy can't think: A response to Bailin. *Informal Logic*, 21(2).
- Hessler, R. (1992). *Social research methods*. St. Paul: West Pub. Co.
- Hirst, R. J. (1968). *Philosophy: An outline for the intending student*. London: Routledge & Kegan Paul Ltd.
- Hitchcock, D. (2003). The effectiveness of computer-assisted instruction in critical thinking. *Informal Logic*, 24(3).
- Hoekema, D. (1987). Why not study something practical, like philosophy? Available: <http://people.stfx.ca/jmensch/Why%20study%20philosophy.html>.
- Hunt, M. (1997). *How science takes stock: The story of meta-analysis*. New York: Russell Sage Foundation.
- Hunter, J., Schmidt, F., Jackson, G. (1982). *Meta-analysis : cumulating research findings across studies*. California: Sage.
- James, W., & Kallen, H. M. (1911). *Some problems of philosophy : A beginning of an introduction to philosophy*. New York: Longmans Green and co.
- Kuhn, D. (1991). *The skills of argument*. Cambridge ; New York: Cambridge University Press.

- Kurfiss, J. K. (1988). *Critical thinking : Theory, research, practice, and possibilities*. Washington, D.C.: Association for the Study of Higher Education.
- Lipman, M. (1988). *Philosophy goes to school*. Philadelphia: Temple University Press.
- Lipman, M. (2003). *Thinking in education* (2nd ed.). New York: Cambridge University Press.
- Lipman, M., & Bynum, T. W. (1976). *Philosophy for children*. Oxford: Published for the Metaphilosophy Foundation Inc. by B. Blackwell.
- Lipman, M., Sharp, A. M., & Oscanyan, F. S. (1980). *Philosophy in the classroom* (2nd ed.). Philadelphia: Temple University Press.
- Lipsey, M., & Wilson, D. (2001). *Practical meta-analysis*. Thousand Oaks: SAGE publications Ltd.
- Mc Millan, J. (1987). Enhancing college student's critical thinking: A review of studies. *Research in Higher Education, Vol 26*(1).
- McPeck, J. E. (1981). *Critical thinking and education*. Oxford: Robertson.
- Milkov, N. (1992). *Kaleidoscopic mind : an essay in post-Wittgensteinian philosophy*. Amsterdam ; Atlanta, GA: Rodopi.
- Moore, B. N., & Parker, R. (1991). *Critical thinking* (3rd ed.). California: Mayfield Pub. Co.
- Moore, D., McCabe, G. (2003). *Introduction to the practice of statistics* (4th. ed.). New York: W.H. Freeman and Co.
- Nieswiadomy, M. (1998). LSAT scores of economics majors. *Journal of Economic Education., 29*, 377–378.
- Norris, S. P. (1992a). *Bachelors, buckyballs, and ganders: Seeking analogues for definitions of "Critical Thinker"*: Philosophy of education society.
- Norris, S. P. (1992b). *The generalizability of critical thinking : multiple perspectives on an educational ideal*. New York: Teachers College Press.
- Norris, S. P., & Ennis, R. H. (1990). *Evaluating critical thinking*. Cheltenham, Vic.: Hawker Brownlow Education.
- Pascarella, E. (1989). The development of critical thinking: Does college make a difference? *Journal of College Student Development, Vol. 30*.
- Pascarella, E., & Terenzini, P. T. (2005). *How college affects students a third decade of research* (1st ed.). San Francisco: Jossey-Bass.
- Paul, R., Binker, A. J. A., & Willsen, J. (1993). *Critical thinking : How to prepare students for a rapidly changing world*. California: Foundation for Critical Thinking.
- Paul, R., Binker, A. J. A., & Willsen, J. (1994). *Critical thinking : what every person needs to survive in a rapidly changing world* (3rd. ed.). Australia: Hawker Brownlow Education.
- Paul, R., & Scriven, M. (2004). *Defining Critical Thinking*: Available: <http://www.criticalthinking.org/aboutCT/definingCT.shtml>.
- Plewis, I. (1985). *Analysing change : Measurement and explanation using longitudinal data*. New York: J. Wiley.
- Priest, G. (2003). *What is philosophy?* Paper presented at the Paper presented at an Inaugural Lecture delivered at the University of Melbourne., University of Melbourne.
- Rainbolt, G., & Rieber, S. (1997). *Critical thinking on the Web. Information page*: Georgia State University. Available: <http://www.apa.udel.edu/apa/archive/newsletters/v98n1/computers/rainreib.asp>.
- Reiter, S. (1994). *Teaching dialogically: Its relationship To critical thinking in college students*. In P.R. Pintrich, D.R. Brown, & C.E. Weinstein (Eds.), *Student motivation, cognition, and learning : Essays in honor of Wilbert J. McKeachie*. New Jersey: Lawrence Erlbaum.

- Rest, J. (1979). *Development in judging moral issues*. Minneapolis: University of Minnesota Press.
- Reznitskaya, A. (2005). Empirical research in Philosophy for Children: Limitations and new directions. *Thinking: The Journal of Philosophy for Children*, 17(4).
- Ross, G., & Semb, G. (1981). Philosophy can teach critical thinking skills. *Teaching philosophy*, 4(2).
- Russell, B. (1961). *History of western philosophy and its connection with political and social circumstances from the earliest times to the present day*. London: Allen & Unwin.
- Russell, B. (2001). *The problems of philosophy*. London: Oxford University Press.
- Siegel, H. (1988). *Educating reason : rationality, critical thinking, and education*. New York: Published by Routledge in association with Metheun.
- Solon, T. (2003). Teaching critical thinking: The More, the Better! *The Community College Experience*, 9(2).
- Spurrett, D. (2005). Critical thinking and argument mapping. *Unpublished manuscript*.
- 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, 333-354.
- Sykes, J. B. (1976). *The concise Oxford dictionary*: Oxford University Press.
- Talaska, R. A. (1992). *Critical reasoning in contemporary culture*. Albany: State University of New York Press.
- Thayer-Bacon, B. J. (2000). *Transforming critical thinking : Thinking constructively*. New York: Teachers College Press.
- Trickey, S., & Topping, K. J. (2004). Philosophy for Children: A systematic review. *Research Papers in Education*, 19(3).
- Twardy. (2004). *Argument maps improve critical thinking*: Available: <http://cogprints.org/3008/01/reasonpaper.pdf>.
- Van der Pal, J., & Eysink, T. (1999). Balancing situativity and formality: The importance of relating a formal language to interactive graphics in logic instruction. *Learning and Instruction*, 9(327-341).
- Van Gelder, T. (1998). The Roles of Philosophy in Cognitive Science. *Philosophical Psychology*, 11(2).
- Van Gelder, T., Bissett, M., & Cumming, G. (2004). Cultivating expertise in informal reasoning. *Canadian Journal of Experimental Psychology*, 58(142-152).
- Walters, K. S. (1994). *Re-thinking reason : New perspectives in critical thinking*. Albany: State University of New York Press.
- Watson, G., & Glaser, E. (1980). *Critical thinking appraisal: Manual*. San Antonio: The Psychological Corporation.
- Williams, R. (2001). The relationship of critical thinking to success in college. *Inquiry: Critical thinking across the disciplines*, 21(1).
- Wittgenstein, L. (1922). *Tractatus Logico-Philosophicus*. London: Routledge & Kegan Paul Ltd.
- Wittrock, M. (1986). *Handbook of research on teaching: A project of the American Educational Research Association*. New York: Macmillan Publishing Company.

9 References used in the Meta-Analysis

- Adams, M. H., Stover, L. M., & Whitlow, J. E. (1999). A longitudinal evaluation of baccalaureate nursing students' critical thinking abilities. *Journal of Nursing Education, 38*(3), 139-141.
- Allegretti, C., & Frederick, N. (1995). A model for thinking critically about ethical issues. *Teaching of Philosophy, 22*(1).
- Arburn, T., & Bethel, L. (1999). *Assisting at risk community college students' acquisition of CT learning strategies*. Paper presented at the National Association For Research in Science Teaching, Boston, Massachusetts.
- Bartlett, D. J., & Cox, P. D. (2002). Measuring Change in Students' Critical Thinking Ability: Implications for Health Care Education. *Journal of Allied Health, 31*(2), 64-69.
- Beckie, T., Lowry, L., & Barnett, S. (2001). Assessing critical thinking in baccalaureate nursing students: A longitudinal study. *Holistic Nursing Practice, 15*(3), 18-26.
- Berger, M. C. (1984). Critical thinking ability and nursing students. *Journal of Nursing Education, 23*, 306-308.
- Brembeck, W. (1949). The effects of a course in argumentation on critical thinking ability. *Speech Monographs, 16*, 172-189.
- Burbach, M., Matkin, G., & Fritz, S. (2004). Teaching critical thinking in an introductory leadership course utilizing active learning strategies: A confirmatory study. *College Student Journal, 38*(3).
- Butchard. (2006). *The monash critical thinking study*: University of Monash. Available: <http://arts.monash.edu.au/phil/research/thinking/>.
- Dale, P., & Ballotti, D. (1997). An approach to teaching problem solving in the classroom. *College Student Journal, 31*(76-79).
- Daly, W. M. (2001). The development of an alternative method in the assessment of critical thinking as an outcome of nursing education. *Journal of advanced nursing, 36*(1), 120-130.
- Donohue, A., Van Gelder, T., Cumming, G., & Bissett, M. (2002). *Reason! project studies, 1999-2002 (Tech. Rep. No. 2002/1)*: University of Melbourne.
- Facione, P. (1990). *The california critical thinking skills test: College level. Technical report # 1*. California: Santa Clara University.

- Frost, S. H. (1991). Fostering the critical thinking of college women through academic advising and faculty contact. *Journal of College Student Development*, 32.
- Gadzella, B., Ginther, D., & Bryant, W. (1996). *Teaching and learning critical thinking skills*. Paper presented at the International Congress of Psychology, Montreal, Canada.
- Gross, Y., Takazawa, E., & Rose, C. (1987). Critical thinking and nursing education. *Journal of Nursing Education*, 26(8), 317-323.
- Hagedorn, L. S., Pascarella, E. (1999). Institutional context and the development of critical thinking: A research note. *The Review of Higher Education*, 22(3), 273-274.
- Harris, J., & Clemmons, S. (1996). *Utilization of standardized critical thinking tests with developmental freshmen*. Paper presented at the National Conference on Research in Developmental Education, Charlotte, North Carolina.
- Hatcher, D. (2001). Why Percy can't think: A response to Bailin. *Informal Logic*, 21(2).
- Hawai'i, U. o. (1993-1998). *Report on the California Critical Thinking Test*: Office of Assessment and Institutional Research. Available: <http://socrates.uhwo.hawaii.edu/socialsci/oshiro/html/assessment/critthnk.htm>.
- Hawai'i, U. o. (1999-2004). *Report on the California Critical Thinking Test*: Office of Assessment and Institutional Research. Available: <http://socrates.uhwo.hawaii.edu/socialsci/oshiro/html/assessment/critthnk.htm>.
- Hitchcock, D. (2003). The effectiveness of computer-assisted instruction in critical thinking. *Informal Logic*, 24(3).
- Kelly-Riley, D., Brown, G., Condon, B., & Law, R. (1999). *Washington State University Critical Thinking Project*. Available: <http://wsuctproject.wsu.edu/materials/ctm-2.pdf>.
- Kintgen-Andrews, J. (1988). "Development of critical thinking: Career ladder P.N. and A.D. nursing students, pre-health science freshmen, generic baccalaureate sophomore nursing students. *ERIC ED 297 153*.
- Lehmann, I. J. (1963). Changes in critical thinking, attitudes, and values from freshman to senior years. *Journal of Educational Psychology*, 54(6), 305-315.
- Meiss, G. T., & Bates, G. W. (1984). Cognitive and attitudinal effects of reasoning message strategies. *ERIC ED 246 519, U.S.*
- Pascarella, E. (1989). The development of critical thinking: Does college make a difference? *Journal of College Student Development*, Vol. 30.
- Quitadamo, I. J., Brahler, C. J., & Crouch, G. J. (2002). A New Method of Evaluating the Effects of Small Group Collaborative Learning on Critical Thinking in Undergraduate Science and Mathematics Courses. *Submitted to Journal of Research in Science Teaching*.

- Rainbolt, G., & Rieber, S. (1997). *Critical thinking on the Web. Information page*: Georgia State University. Available:
<http://www.apa.udel.edu/apa/archive/newsletters/v98n1/computers/rainreib.asp>.
- Rest, J. (1979). *Development in judging moral issues*. Minneapolis: University of Minnesota Press.
- Richards, M. (1977). One integrated curriculum: An empirical evaluation. *Nursing Research*, 26(2), 90-95.
- Rimiene, V. (2002). Assessing and developing students' critical thinking. *Psychology Learning and Teaching*, 2(1), 17-22.
- Ross, G., & Semb, G. (1981). Philosophy can teach critical thinking skills. *Teaching philosophy*, 4(2).
- Ruff, L. (2005). *The development of critical thinking skills and dispositions in first-year college students: infusing critical thinking instruction into a first-year transitions course*. Unpublished Dissertation, University of Maryland.
- Saucier, B., Stevens, K., & Williams, G. (2000). Critical thinking outcomes of computer-assisted instruction versus written nursing process. *Nursing and Health Care Perspectives*, 21(5), 240-246.
- Scott, J., Markert, R., & Dunn, M. (1998). Critical Thinking: Change during medical school and relationship to performance in clinical clerkships. *Medical Education*, 32(14-18).
- Solon, T. (2001). Improving critical thinking in an introductory psychology course. *Michigan Community College Journal*, 7(2), 73-80.
- Solon, T. (2003). Teaching critical thinking: The More, the Better! *The Community College Experience*, 9(2).
- Soukup, F. (1999). Assessment of Critical Thinking Skills in Associate Degree Nursing Students at Madison Area Technical College-Reedsburg. *ERIC ED430081*.
- South Seattle Community College, W. (1994). Institutional effectiveness assessment process, 1993-94. Executive summary. *ERIC ED381223*.
- Spurrett, D. (2005). Critical thinking and argument mapping. *Unpublished manuscript*.
- Stockard, S., Parsons, M., Hercinger, M., & Andrews, A. (2001). Evaluation of critical thinking outcomes of a BSN program. *Holistic Nursing Practice*, 15(3), 27-34.
- Sullivan, E. (1987). Critical thinking, creativity, clinical performance, and achievement in RN students. *Nurse Educator*, 12(2), 12-16.
- Thompson, C. R., L. M. (1999). Critical thinking skills of baccalaureate nursing students at program entry and exit. *Nursing and Health Care Perspectives*, 20(5).

- Tomlinson-Keasey, C. A., & Eisert, D. (1977). *Second year evaluation of the ADAPT program. In Multidisciplinary Piagetian-Based Programs for College Freshmen: ADAPT:* University of Nebraska.
- Twardy. (2004). *Argument maps improve critical thinking*: Available: <http://cogprints.org/3008/01/reasonpaper.pdf>.
- Vieira, C., & Oliveira, D. M. (1997). *Lab activities in the light of critical thinking*. Paper presented at 1997 NARST Annual Meeting, Oak Brook, Illinois.
- Watson, G., & Glaser, E. (1980). *Critical thinking appraisal: Manual*. San Antonio: The Psychological Corporation.
- Wheeler, L. A., & Collins, S. K. R. (2003). The influence of concept mapping on critical thinking in baccalaureate nursing students. *Journal of Professional Nursing.*, 19(6), 339-346.
- Whitmire, E. (2001). The relationship between undergraduates' background characteristics and college experiences and their academic library use. *College and Research Libraries*, 62(6), 528-540.
- Williams, R. L. (2003). Critical Thinking as a Predictor and Outcome measure in a Large Undergraduate Educational Psychology Course. *ERIC ED478075*.
- Williams, R., Oliver, R., & Stockdale, S. (2004). Psychological versus generic critical thinking as predictors and outcomes measures in a large undergraduate human development course. *The Journal of General Education*, 53(1).

10 APPENDICES

Three appendices are attached to this thesis. These provide much of the data on which the thesis is based. Appendix A is the Master List of studies used in the meta-analysis. Appendix B sets out the statistical information derived from the various studies, which are divided into seven groups, each of which consists of one of the independent variables explored in the thesis. Appendix C explains the methods used to derive the pre- and post-test standard deviations for those studies which did not themselves report standard deviations.