

**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.

## Declaration

This is to certify that:

- (i) the thesis comprises only my original work except where indicated in the acknowledgements,
- (ii) due acknowledgement has been made in the text to all other material used,
- (iii) the thesis is 30,000 words in length, inclusive of footnotes but exclusive of tables, maps, appendices and bibliography.

.....  
**Claudia María Álvarez Ortiz**

## Acknowledgements

Sometimes, things happen because you are in the right place at the right time. That happened in the case of this thesis. Tim van Gelder, my thesis supervisor, had conceived the idea of this project some time before I appeared on the scene. I initially arrived in Australia, planning to stay for only two months. The idea was here ahead of me. Unexpectedly, on my arrival, I encountered not only Tim's idea, but his friend and business partner, Paul Monk. The two things, together, changed both my personal and professional outlook, so that my stay in Australia became permanent.

Tim thus played the key role in triggering this change. Not only did he provide the idea for this project, but he has provided support, as my supervisor, over the past two years. His intelligence, his expertise in the nature of argumentation, and above all his commitment to the discovery and testing of new methods for improving critical thinking have been a constant source of inspiration. How could I not be eternally grateful to him? My thanks, Tim, for the trust you have shown in me. I hope I have not disappointed you.

Coincidentally, after Tim had invited me to take on this project, the University of Melbourne offered me two scholarships (an International Postgraduate Research Scholarship and a Melbourne International Research Scholarship) which enabled me to undertake the project as a Masters degree in the Philosophy Department. I am immensely grateful to the University for giving me the opportunity to go to work on this basis and for seeing the potential in the project itself.

By the time I actually took up the project, Tim and his colleagues had already collected a considerable body of data. My task, at the beginning of the project, was to familiarize myself with all of this data and, then, update it by collecting further studies. I am especially grateful for the support of Melannie Bissett, who has long been part of Tim's team and who, generously, organized for me all the data that she and other members of the team had collected. She did this to such effect that I was able easily to absorb, use and update it. Simply, this thesis would not have been possible without the work she did and the support she has given me.

The work of collecting new data, however, was far from simple. A meta-analysis requires considerable patience, when it comes to conducting an exhaustive search for studies in library and internet data-bases. Here, I must acknowledge the invaluable assistance of Blanca Pizzani, who was the research librarian assigned to the Department of Philosophy at the time, and who was so friendly in spending hours with me searching on the internet. My thanks go, also, to Vija

Pattison, who works for the Interlibrary Loans and Document Delivery section of the Baillieu Library. Vija tirelessly hunted down and collected studies that I had thought could not be found.

Many people contributed to the technical, statistical part of this thesis. I owe a debt of thanks to Sue Finch of the Statistical Consulting Centre at the University of Melbourne. Sue was generous in providing relevant comments on this aspect of the thesis as it developed. She also helped with the derivation of some of the standard deviations, where they had not been reported in the original studies. I must, above all, thank the incredible Neil Thomason, from the Department of History and Philosophy of Science, University of Melbourne. In the course of the project Neil, a fine teacher, has not only been of enormous assistance to me; he has become a good friend. Communicating his passion for both, Neil taught me that there is much more to Rossini than William Tell and that it is possible to find statistical analysis fascinating. Throughout the project, he has provided support in many ways: advice, books and, especially, motivation.

I want to offer a very special vote of thanks to Geoff Cumming from La Trobe University. Geoff became, in effect sizes, as it were, the 'step-supervisor' of this thesis. Thank you, Geoff, for the contribution of your astonishing expertise in statistical science, without which the really fine-grained analysis in the thesis could not have been done. Thank you, also, for your assistance with the graphical representation of the findings. Thank you, finally, for your ever friendly demeanour. Your skill in communicating an understanding of complex matters constantly amazed me.

Then there is Tim's business partner, Paul Monk. Having agreed, at Tim's request, to meet me at Tullamarine, on my arrival in Melbourne, and provide me with 'temporary' accommodation, Paul began as a perfect host and ended up, exactly six months later, as my husband. How can I thank him for all that he has since done for me? I do not have the space here to express my gratitude for everything he has done over the past two and a half years, so I shall limit myself to the difference he has made to my academic endeavours. This thesis would have been easier for me to have written in Spanish; but thanks to Paul's generous assistance, it has emerged in reasonably clear English.

Besides his erudition and great intelligence, Paul has an astonishing capacity to detect 'gaps' in any chain of reasoning. With his help, I was better able both to polish my use of English and to strengthen my process of thinking, as I drafted the thesis. Throughout all of this, he exhibited endless patience, empathy and intellectual rigour. Paul is the best partner a person could have while doing a thesis. It has been my great good fortune to have had him at home, where I could share with him not only the various challenges of completing the thesis, but also the many other activities that partners in life share. And all this was in the context of the deep mutual love we have developed – since Tim set up our meeting at Tullamarine.

Finally, I want to thank those personal friends, both here and overseas, as well as my family, whose emotional support has been of inestimable help to me in being able to make a life in Australia and complete what often seemed a very challenging project. In particular, in Australia, my thanks go to Katharine West, who after a providential encounter became my confidant and friend. Thank you, Katharine, for helping me get through the difficult times that go both with settling in a new country and attempting to do a thesis by research in a second language. And, in Venezuela, I cannot conclude without expressing my gratitude to my soul-mate of many years, Sarah Henderson, whose support throughout these past few years has been amazing.

## 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

## Table of Figures

Figure 1. The Findings of the Thesis.....	4
Figure 2. Questioning the assumption that the study of philosophy improves critical thinking skills.....	32
Figure 3. The need for a meta-analysis .....	55
Figure 4. Chart of effect sizes calculated using the SD found in each study.....	69
Figure 5. Chart of effect sizes calculated using our best estimates of SD for each test instrument.....	70
Figure 6. CT gain in philosophy vs CT gain in Non-philosophy courses.....	74
Figure 7. CT gain in Anglo-American analytic philosophy vs No Phi No CT .....	75
Figure 8. Anglo-American analytic philosophy vs all CT instruction in philosophy. ....	76
Figure 9. Anglo-American analytic philosophy vs Traditional CT in philosophy. ....	76
Figure 10. Anglo-American analytic philosophy vs CT instruction in philosophy with argument mapping.....	77
Figure 11. Anglo-American analytic philosophy (Group 1) vs Traditional CT instruction outside philosophy. ....	78
Figure 12. Anglo-American analytic philosophy vs all CT instruction in outside philosophy. ....	78
Figure 13. Anglo-American analytic philosophy vs Traditional CT in philosophy and outside philosophy. ....	79
Figure 14. All CT instruction in philosophy (Groups 2 and 3) vs No Philosophy, No CT instruction. ....	80
Figure 15. All CT instruction outside philosophy vs No Philosophy, No CT instruction. ....	80

# 1 Introduction

This thesis investigates the assumption that studying philosophy improves critical thinking. It seems important, from a philosophical standpoint, to closely examine assumptions of this nature. Examining them, since it involves both close argument and a consideration of the impact of philosophy, is an inherently philosophical undertaking. At the same time, given the thesis's structure and approach, the exercise is also partly educational, because of its implications; and partly statistical in its research methodology. It is, therefore, to some extent a multidisciplinary piece of work.

Both the impulse to undertake this project and the significance that it has are rooted in the reasons that, ultimately, justify the study of philosophy itself. When university candidates, wondering what career path to embark upon and seriously pondering formal philosophical study in this context, ask themselves the question *Why study philosophy?*, they will almost invariably come across the following claims for the merits of doing so: The study of philosophy develops one's abilities to distinguish good and bad reasoning, to develop and defend one's own ideas through arguments and to think critically.

Indeed, this belief that the study of philosophy helps to develop these critical thinking abilities appears to be one of the chief bases on which the discipline of philosophy claims a competitive advantage over other disciplines. It is enough to look at the websites of almost any of the philosophy departments at the world's major universities – at least in so far as they teach Anglo-American analytical philosophy – in order to see just how true this is. The common message is that those who study philosophy emerge as better critical thinkers.

Examples of the advertisement of this claim are readily available on the web. The following, are representative examples:

"Everyone thinks that critical thinking is good, but for philosophy alone is it a *modus operandi*."<sup>1</sup>

"By studying philosophy, you acquire reasoning, verbal and writing skills in critical and reflective thinking..."<sup>2</sup>

"The critical thinking skills attained are applicable to any vocation and philosophy teaches these skills more directly and consistently than any other academic discipline."<sup>3</sup>

---

<sup>1</sup> The University of Alabama. Available at: <http://www.as.ua.edu/philos/why.php>

<sup>2</sup> University of NSW. Available at:  
[http://philosophy.arts.unsw.edu.au/futurestudents/why\\_philosophy.html](http://philosophy.arts.unsw.edu.au/futurestudents/why_philosophy.html)

The view that this kind of statement exhibits is not an unreasonable one. We can find its genesis in two basic points of commonality between philosophical thinking and critical thinking. To begin with, the very idea of critical thinking has its roots in philosophy. Until quite recently, most of the standard textbooks in critical thinking were authored by philosophers. Likewise, most of the courses in critical thinking at universities are still taught by philosophers (Walters, 1994). Secondly, the critical nature of philosophy leads us to infer that studying it will automatically improve a student's skills in critical thinking.

Hence, there is a general and understandable belief, among professors and students, that studying philosophy contributes to improving critical thinking skills and that it does so more than any other course of study. What concrete evidence, however, do we have for the belief that philosophy *does*, actually, contribute to improving critical thinking skills? And, even if we had evidence for that belief, what evidence do we have that philosophy makes *more* difference to the development of such skills than can be or is made by other disciplines? A special case among such other disciplines is courses designed to teach critical thinking as a discipline in its own right. Consequently, a third question must be, to what extent do *critical thinking* courses actually develop critical thinking skills more effectively than either philosophy as such or disciplines other than philosophy?

In order to be rigorous in answering these questions, we need to examine or develop the empirical data that would enable us to establish whether or to what extent the assumption about philosophy is actually warranted and how it compares with other disciplines in this respect. A certain amount of such empirical data exists in regard to each of these questions, but it is incomplete and equivocal. As regards the effectiveness of *philosophy*, the empirical evidence is inconsistent; some suggesting that it makes a difference, although to varying degrees, while other evidence suggests that it makes no appreciable difference.

The available data on other disciplines suggests is similarly inconsistent in the claims it makes for the degree of difference they make to the development of critical thinking skills. What has been lacking so far is both a consistent *research* methodology for assessing the impact of any given discipline, such as philosophy, and also a *comparative* study of the relative impact of different disciplines. This suggests that there is a need to conduct research from a comparative perspective, with a view to establishing whether philosophy does better than other disciplines in developing critical thinking skills.

The present study has, therefore, been designed to examine the difference that philosophical studies make to the development of critical thinking skills, compared with the difference made by non-philosophical university education. In order to achieve this comparative perspective, it is

---

<sup>3</sup> Winthrop University, South Carolina. Available at:  
<http://www.winthrop.edu/phlrelg/whystudy.htm>

based on a *meta-analysis* of existing empirical studies of both. Meta-analysis is a methodology that enables us to weigh the significance of a variety of empirical studies of different magnitudes that are based on different research methods. In the present case, it enables us to make a more careful assessment than we have had to date of the real impact of philosophy on the development of critical thinking.

In order to communicate as clearly as possible the findings of the meta-analysis, the project consists of six chapters. Chapter 1 is the introductory chapter. Chapter 2 consists of the clarification of the key terms in the assumption under examination: philosophy, critical thinking and improvement. Chapter 3 contains an account of the grounds on which the assumption is generally held, as well as an argument for the need to conduct an empirical investigation of it. Chapter 4 examines the available empirical evidence and finds it less than conclusive. The meta-analysis in Chapter 5, challenges the assumption itself. Chapter 6 consists of a discussion of the findings of the meta-analysis and their implications.

In brief, the key findings of the meta-analysis are as follows:

- 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.

Figure 1 sets out a somewhat more detailed picture of the findings of the thesis.

The empirical question at the heart of the project – to what extent do philosophy studies and university studies improve critical thinking skills? – ought to be of interest to anyone thinking seriously about higher education. It has never been addressed properly before; which is to say, that although people always say that a primary benefit of university (and philosophy) education is to improve critical thinking skills, nobody has been able to say *how much* these skills are improved by the standard studies or methods. While they may indeed improve, it would surely be useful to establish how much they improve. Only then would we be well placed to assess precisely why and how. This would, in turn, enable us to set new benchmarks and to set about improving such skills more effectively than we already do.

Figure 1. The Findings of the Thesis

## 2 Clarifying the Assumption

As explained in the introduction, the aim of the present study is to investigate the assumption that studying philosophy improves critical thinking. The first step in this process is to clarify what this means.

Let us focus on the assumption in the context in which it is most commonly made, i.e., by philosophers in universities promoting the value of studying philosophy.

### 2.1 Defining the Terms

The assumption, as loosely articulated, has three components:

1. Studying philosophy
2. Improves
3. Critical thinking

In this section, these are each clarified in turn.

#### 2.1.1 “Studying Philosophy”

The first component, *studying philosophy*, can be further broken down into

Philosophy

Studying [philosophy] at University

##### 2.1.1.1 What is philosophy?

We require a definition of philosophy that is clear enough for the purposes of the current research project. Clearly, over time, philosophy has included many lines of speculation and reflection which could be considered in this context. However, the inquiry is almost exclusively concerned with the impact on critical thinking of the central disciplines of philosophy, as these are understood and practiced in the analytical tradition. Those central disciplines consist of a number of core philosophical problems and certain characteristically philosophical methods for tackling these problems.

The word “philosophy”, derives etymologically from the ancient Greek terms for love (*philos*) and wisdom (*sophia*) and thus means the “love of wisdom”. From the pre-Socratics to Plato, the word ‘wisdom’ had a unified notion: “The understanding of the ultimate principles of things that works as guide to truly live an exemplary life” (Audi, 1999) In this spirit, the early Greek

philosophers wondered about such things as the nature of the universe, the purpose of life, the meaning of right and wrong, and the implications that all of this had in their lives. In studying and searching for wisdom, philosophers built systems of knowledge structured by questions and answers filtered by thinking, specifically by reason.

More contemporary conceptions of philosophy show that the “love of wisdom” is still valid today. Philosophers now, as then, continue to ponder questions about the most fundamental things in life. However, philosophical activity nowadays seems to have acquired a different character. Pierre Hadot, (Hadot & Davidson, 1995), captured this by stating: “Ancient Philosophy proposed to mankind an art of living. By contrast, Modern Philosophy appears above all as the construction of a technical jargon reserved for specialists” (p. 272). It is not necessary here to insist that this aphorism applies in all universities and academies that pursue philosophical activities; only to note that philosophy as it was conceived in its beginning has given way to a vast range of disciplines, which have fragmented human understanding and even philosophical discourse. Today philosophy is taught and studied at universities as one particular discipline among others. Indeed, it is itself many disciplines, only notionally unified as a discipline under a single rubric. It has stopped being mainly an art of living and has, instead, become a largely academic discipline, or family of disciplines.

Given that the search for knowledge and understanding is no longer the exclusive domain of philosophy and indeed other disciplines pursue similar objectives, what, then, is philosophy now considered to be? Webster’s dictionary defines philosophy as: “The branch of knowledge or academic study devoted to the systematic examination of basic concepts such as truth, existence, reality, causality and freedom”. This would suggest that, in fact, rather than being just another branch in the tree of knowledge, philosophy lies at the root of other disciplines, in so far as it is concerned with the nature of truth and knowledge as such.

A difference, therefore, between philosophy and other disciplines is chiefly that philosophy has such a markedly epistemological character. Philosophers considering this question have advanced three elements of philosophy that set it apart from other disciplines. These are: a) The problems it raises, b) The methods it utilizes, c) The spirit it cultivates. Of course, all serious disciplines are distinguished by the problems they raise, by the methods they use and even, to some extent, by the spirit they cultivate. However, philosophy is set apart by its singular pre-occupation with the problems of knowledge and truth, the methods for attaining these things and the spirit of scepticism and love of truth, in short the spirit of critical thinking.

#### **2.1.1.1.1 Philosophical Problems**

William James, claimed, that as soon as questions get accurately answered, those answers are called ‘scientific’, and “what men call ‘philosophy’ to-day is but the residuum of questions still unanswered”(James & Kallen, 1911). But what characterizes those kinds of unanswered questions? Bertrand Russell, in *The History of Western Philosophy*, suggested that philosophy

is something intermediate between theology and science. Theological problems consist of speculations on matters regarding which definite knowledge has, so far, been unascertainable. Scientific problems, he thought, were by nature those to which one could find definite answers. Thus, philosophy, according to Russell, exists in a kind of No Man's Land (Russell, 1961), p. 13). But if this is true, why are there such problems, to which there do not seem to be definite answers? What, at the end of the day, are the characteristics of these 'philosophical' problems?

We might describe philosophical problems, with the English philosopher of education R. J. Hirst as "fundamental problems". (Hirst, 1968). Such fundamental problems, he suggested, share certain characteristics. The first of these is their generality or wide scope. Philosophy raises broad questions whose considerations involve the knowledge of many disciplines. For instance, the mind/body debate, which is considered a philosophical issue, may need assistance from neurology, psychology and, perhaps, theology.

Secondly, philosophical problems are abstract. They do not concern specific persons, events, and phenomena. Rather, they are concerned with the universal character of personhood, the nature of events and the reality of phenomena. Thirdly, the main philosophical problems are to do with the most important ideas which govern how we understand ourselves and the world. In other words, philosophy addresses the most enduring problems we encounter in understanding nature, the human world, and the interaction between the two.

#### **2.1.1.1.2 Philosophical Methods**

One might say that reason is the philosopher's main tool. Analytic philosophy is essentially a rational activity that involves the continual testing of ideas and probing away at the foundations of beliefs. But, how do philosophers proceed in doing so? What methods do they use to do philosophy?

Van Gelder argued that "the best way to identify philosophers, is not by the subject matter they are discussing, nor by where they are housed, but by their method" (Van Gelder, 1998). He claimed that philosophers can be identified by their reliance on certain characteristic methods: arguments, conceptual clarification and historical perspective.

Through argumentation, the philosopher attempts to establish strong premises and proceed from them, by clear chains of inference, to show what point or conclusion follows from them. Using the same skills of argument, he or she is able to evaluate and criticise other people's claims and beliefs, by analysing their logic and challenging them to reconsider the bases for the opinions.

One thing the philosopher always keeps in mind is the guiding principle that nothing should be accepted as true without rational justification. In general, philosophers do not undertake demanding empirical investigations; but empiricism is hardly foreign to philosophy and the use of evidence is central to it. It is here that modern science parted ways with the syllogisms of the

medieval scholastics, in the manner famously explained by Francis Bacon. In particular, the philosopher can take advantage of such work by others; using their findings as premises in arguments (Van Gelder, 1998). This is precisely what we shall be doing here.

Conceptual clarification consists of getting clear on the meaning of terms and concepts involved in arguments. The philosopher, as Van Gelder pointed out, not only has to pay attention to the argument structure, but also to the meaning of the concepts that constitute every premise in order to validate it. Anglo-American analytic philosophy, as represented by the work, for instance, of Austin and Quine, has been especially concerned with this kind of conceptual clarification over the past hundred years. As Wittgenstein observed, in the *Tractatus Logico-Philosophicus*: “The result of philosophy is not a number of philosophical propositions, but to make propositions clear. Philosophy should make clear and delimit sharply the thoughts which otherwise are, as it were, opaque and blurred” (Wittgenstein, 1922, p. 4.112).

Van Gelder also mentioned a third method used by philosophers: historical perspective. This requires the revision of past ideas and treatments of the issue under consideration. It means that philosophers need to become familiar with what great minds of the past have had to say on any particular matter. Thus, a great deal of philosophical reflection has been devoted to puzzling over writings by the likes of Plato, Aristotle, Aquinas, Hegel, Wittgenstein and Heidegger.

#### **2.1.1.1.3 *The Critical Spirit***

There is a third thing, however, that might be said to set philosophers apart. Almost a hundred years ago, William James referred to philosophers as those who “find matter for puzzle and astonishment where no one else does” (James & Kallen, 1911), p. 3). Philosophers are characteristically puzzled by and interested in how human beings arrive at their opinions at all. Philosophy is thus a “critical reflection on the justification of basic human beliefs and analysis of basic concepts in terms of which such beliefs are expressed” (Edwards & Pap, 1972)

This puzzlement and *critical spirit* have been said to distinguish philosophy from all other modes of intellectual inquiry (Priest, 2003). In traditional theology, for example, one is explicitly not allowed to question certain things. In physical science one may be expected to be critical of new ideas and findings, but one is not encouraged to question a well-established scientific position. In philosophy, anything is supposedly open to critical challenge and scrutiny.

This spirit has its association with the sceptical nature of the discipline. The philosopher does not accept propositions without a clear and reasonable justification. These words are reminiscent of Descartes' inquisitive attitude, when he set for himself the principle “never to accept anything as true if I did not know clearly that it was so; that is, carefully to avoid prejudice and jumping to conclusions, and to include nothing in my judgment apart from whatever

appeared so clearly and distinctly to my mind that I had the opportunity to cast doubt on it” (Descartes & Clarke, 1999).

Bertrand Russell, on the other hand, argued, while philosophy is more concerned with criticism than evidence, the critical review of claims and knowledge must be a constructive one. Absolute scepticism leads nowhere. Philosophical critique should lead to new ideas, systems of knowledge, paradigms and insights (Russell, 2001). Aristotle famously remarked “I love Plato; but I love truth more” and proceeded both to challenge many of Plato’s epistemological positions and to offer alternative approaches to knowledge in various fields.

In summary, throughout the history of philosophy, the diversity of opinion among philosophers about the nature of the discipline has been wide. Some thinkers have claimed that philosophy is defined by its content: the nature of its problems. Some have emphasised the methods used. Others again have placed emphasis on the critical spirit. However, most would agree that all three of these things characterise philosophy. They simply differ in emphasis. For our present purposes, philosophy will be understood as consisting of a concern with fundamental problems of knowledge, a spirit of scepticism concerning how conclusions are drawn or opinions defended and a commitment to reason in justifying such conclusions or opinions.

### **2.1.1.2 *Studying Philosophy at University***

What is it to *study* philosophy, in the sense relevant for investigating the assumption? Since the assumption most commonly surfaces as part of an attempt to convince undergraduates that they should enrol as philosophy students, the notion of *studying* philosophy must be something close to what these people would actually do if they did enrol. Thus we can say that to *study* philosophy is to do pretty much what a standard undergraduate would do in the context of a philosophy subject at a typical US, English or Australian university.

We limit ourselves here to US, English or Australian university courses in philosophy, because we shall be concerned to examine the effects of Anglo-American analytic philosophy, rather than philosophy more broadly considered, on the improvement of critical thinking skills. This said, the study of philosophy is taken to mean the usual methods of study for virtually any academic discipline:

- Attending lectures
- Reading primary and secondary texts
- Writing essays
- Attending tutorials and participating in discussions

These activities would be undertaken at approximately the level of intensity that would be typical of a reasonably conscientious student. And, in the case of philosophy, they would be concerned

to inculcate in the students the three things which, as we have said, set philosophy apart as a discipline: reflection on fundamental problems, the use of arguments as the main methodological tool and the disposition to engage in critical reflection.

Importantly, this means that studying philosophy should not be understood as some kind of quite exceptional activity, whether at a very high level, or a high degree of intensity, or in some quite special educational context. It may be true that studying in Oxford's small tutorial system, or studying as a PhD student in some first-rate US program, improves critical thinking skills, but these types of study are far more intensive than what people generally have in mind when they say or assume that studying philosophy improves critical thinking. If it turned out that studying philosophy only improved critical thinking skills if you study at Oxford, or as a postgraduate student at Princeton or Pittsburgh, then the assumption as it is generally made would be false.

### 2.1.2 “Critical Thinking”

A common expression to be found in the body of literature about critical thinking (CT) is that *there is no established consensus over its definition*. However, we can refer to those general approaches that have generated relevant paradigms, in order to delimit the conceptual framework that will be used in applying the term CT within the present project. Two approaches have attempted to delimit the conceptual framework of critical thinking: the traditional approach of logical analysis of information and a new approach called the *second wave* of critical thinking.

This ‘second wave’ approach adds to the idea of a critical thinker as one who applies methods of logical analysis, the use of imagination, intuition and the active participation of emotions and values (Thayer-Bacon, 2000; Walters, 1994). In doing this, it attempts to bring into consideration what are undoubtedly interesting elements of human cognition. However, its proponents have yet to provide both the pedagogical methods for teaching CT in this manner and, even more crucially, clear empirical evidence that it is a better way to do CT.

Conversely, there are four strong reasons for sticking with the traditional approach. First, this approach reflects the most common practice of CT in schools and universities. Second, the characteristics that identify this approach are closer with those used in the standard tests used to measure CT. Third; it has the consistent methodological principles needed to build a theory in any discipline. Fourth, it provides the basis for being able to transfer the skills in question into other domains. Indeed, it has provided core tools and resources for good thinking in existing successful disciplines of inquiry (Ennis, 1991).

Even within the traditional approach to critical thinking, i.e. logical analysis of information, there are a variety of definitions of what actually constitutes critical thinking (see Table 1). However, these various definitions share enough common characteristics between them to allow us to describe them as a single general approach. These common characteristics, taken together, constitute the prevailing view of the concept of critical thinking, centring on the role of

informal logic in exploring argument. It is this prevailing view which shall be used for the purposes of the present study.

### **2.1.2.1 The Prevailing View of Critical Thinking**

The predominant view, then, is that critical thinking has become almost synonymous with the methods of informal logic (Kurfiss, 1988). Informal logic is defined as the study of arguments as presented in ordinary language, as contrasted with the presentations of arguments in an artificial, formal, or technical language. Reasoning and logical analysis in general have, of course, played a central role in the critical thinking tradition (Norris & Ennis, 1990), Walters, 1994, (A. Fisher, 2001). CT, in this sense, is analytical, abstract, universal and objective (Walters, 1994). Thus, many CT definitions focus on how to review and carefully evaluate judgments (Dewey, 1910; Harris, Hodges, 1981; B. N. Moore & Parker, 1991) and on the analysis and evaluation of arguments (Kurfiss, 1988, A. Fisher & Scriven, 1997). All this clearly sets this general approach apart from the inclinations of 'second wave' theorists.

The use of reasoning as the core of CT entails three constituents: *knowledge, attitudes and skills* (Watson & Glaser, 1980). The critical thinker requires not only knowledge of the methods of logical analysis, but also of the domain (for example, History, Medicine) in which the critical thinking is being done. A long and often polemical debate is still developing around whether good CT requires only the knowledge of methods of logical analysis, or also demands domain specific knowledge (McMillan, 1987; McPeck, 1981; Norris, 1992b; Pascarella, 1989; Siegel, 1988). There appears to be somewhat less controversy among the experts regarding the requisite skills and attitudes.

Various attempts have been made to operationalize CT, making it a set of well-defined and testable skills, rather than simply a somewhat vague set of dispositions and attitudes. Notably, in 1988-89, 46 experts articulated a definition of CT skills in what is known as the Delphi Report (see Table 1). This has become widely known and used for research purposes. It has promoted the design of tests to guide, foster and evaluate CT skills. For instance, it became the template for the design of two commercial tests to evaluate CT skills and dispositions: the California Critical Thinking Skills Test, and the California Critical Thinking Dispositions Inventory.

Other such attempts to design good tests of CT skills have also been made. In general, those tests have focused on the evaluation of CT skills, rather than the attitudes or dispositions underlying them. They seem to agree in identifying the critical thinker as a person able to interpret the meaning and analysis of inferences, observations, judgments and arguments; undertake deductive and inductive reasoning; and evaluate the validity of claims and arguments (Ennis, Millman, & Tomko, 1985; Facione, 2006; A. Fisher, 2001; Paul & Scriven, 2004; Watson & Glaser, 1980). Paul has included the idea that the critical thinker has also the ability to evaluate his or her own thoughts (Paul, Binker, & Willsen, 1994).

In what follows, critical thinking is framed in terms of the paradigm of the traditional approach of logical analysis and reasoning. Conceptually, CT involves the application of knowledge, and certain kinds of general skills and attitudes to the rigorous pursuit of logical analysis. In general, it is expected that critical thinkers are *disposed* to care that their beliefs be true, and that their decisions be justified; care to present a position honestly and clearly; and care about the worth of other people's opinions (Ennis, 1991). Given such a disposition, the critical thinker requires certain *skills* to be able to interpret and analyze claims, engage in reasoning and evaluate the validity of claims and arguments. This is the conceptual framework of CT which underpins the present study.

**Table 1. Some examples of common definitions of the term Critical Thinking, arranged in chronological order.**

Authors	Definitions
John Dewey, 1910	Reflective thinking is: "Active, persistent, and careful consideration of a belief or supposed form of knowledge in the light of the grounds which support it and the further conclusions to which it tends".
Edward Glaser, 1941 (in Fisher 2001)	CT is: "(1) an attitude of being disposed to consider, in a thoughtful way, problems and subjects that come within the range of one's experience; (2) knowledge of the methods of logical enquiry and reasoning; and (3) some skills in applying those methods. Critical Thinking calls for a persistent effort to examine any belief or supposed form of knowledge in the light of the evidence that supports it and the further conclusions to which it tends".
John, McPeck, 1981	"The core meaning of critical thinking is the propensity and skill to engage in an activity with reflective scepticism".
Harris & Hodges, 1981	Critical Evaluation is "the process of arriving at a judgment about the value or impact of a text by examining its quality in terms of form, style, and rhetorical features, the readability of the author and the consistency between ideas it presents and the reader's experience, including internal evaluation and external evaluation".
Robert Ennis, 1985	It is "the reasonable and reflective thinking focused on deciding what to believe or do".
Joanne Kurfiss, 1988	"Introductory courses on critical thinking teach students to detect and avoid fallacious reasoning and to analyze deductive and inductive arguments".
Harvey Siegel, 1988	To be critical thinker is to be appropriately moved by reasons. Critical thinking involves bringing to bear all matters relevant to the rationality of belief and

	action; and education aimed at the promulgation of critical thinking is nothing less than education aimed at the fostering of rationality and the development of rational persons”.
Delphi Report, 1990	We understand critical thinking to be purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual inquiry. The ideal critical thinker is habitually inquisitive, well informed, trustful of reason, open-minded, flexible, fair-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to consider, clear about issues, orderly in complex matters, diligent in seeking relevant information, reasonable in the selection of criteria, focused in inquiry, and persistent in seeking results which are as precise as the subject and the circumstances of inquiry permit.
Moore and Parker, 1992	“Critical thinking is careful and deliberate determination of whether to accept, reject, or suspend judgment”.
Richard Paul, 1995	“A unique kind of purposeful thinking, in which the thinker systematically and habitually imposes criteria and intellectual standards upon the thinking, taking charge of the construction of thinking, guiding the construction of the thinking according to the standards, assessing the effectiveness of the thinking according to the purpose, the criteria, and the standards”.
Fisher & Scriven, 1997	“Critical thinking is skilled and active interpretation and evaluation of observations and communications, information and argumentation”.

### 2.1.3 “Improves”

What does it mean to say that studying philosophy *improves* critical thinking?

Crudely, *improves* means “makes better.” This may seem too obvious to mention, but unpacking the notion in this way is helpful because it gives us two notions to explicate – improvement or gain, on one hand, and “making” or causing, on the other. In other words, for studying philosophy to improve critical thinking, we need the following:

- People studying philosophy get better at critical thinking
- This improvement is due to studying philosophy

In the following sections these are discussed in turn.

#### 2.1.3.1 What is meant by ‘Improvement’

- People must “be better” and better in some substantial sense.
- In particular, they must be better than they would have been had they not studied philosophy.
- This substantial improvement must be detectable
- For it to be detectable, CTS themselves must be measurable. This means that they must be open to what is called “operationalization”. This means that they can be assessed by means of standard tests.
- Finally, the improvement should be substantial with reference to relevant comparisons. Such relevant comparisons should, as far as possible, consist not only of other university subjects in general; but should include the study of critical thinking as such.

##### 2.1.3.1.1 *The Question of ‘Improving’ CTS Performance*

In the context of this study, ‘improving CTS’ is given a cognitive treatment. The attention is focused on the acquisition and development of certain types of cognitive abilities, rather than on the attitudes associated with the critical thinker. CT is, for the purposes of this inquiry, considered a complex, higher-order, domain independent cognitive skill set.

CT skills, like any other skills, are things that may be learned well or poorly; utilized efficiently or inefficiently. Deanne Kuhn, for instance, after conducting empirical research on the topic, concluded that many adults have surprisingly weak CT skills, and only a few can be said to have achieved mastery (Kuhn, 1991). Kuhn interviewed one hundred and sixty people of different ages and education levels and discovered that while most people readily adopt strong convictions, a majority cannot reliably generate any genuine evidence at all for those

convictions. However, it is necessary to be more precise than this. *What is the criterion to distinguish between better or worse CT performance? What does expertise mean in this context?* Implicitly, this criterion sets the basis to distinguish improvement in CT performance.

To supply a theoretical perspective, Fisher & Scriven, can guide us on this matter. In their definition of critical thinking<sup>4</sup>, they claim that CT is a “skilled” set of abilities (A. Fisher & Scriven, 1997). The inclusion of the word “skilled” reflects the requirement that one cannot classify something as CT if it does not meet minimum standards of quality. To qualify as CT at all, it must meet such basic standards of good thinking, especially with regard to the relevance, accuracy, fairness and adequacy of evidence used in drawing conclusions. Of course, these standards are context-dependent, in the sense that what counts as good or acceptable CT will vary somewhat with the age and sophistication of the thinker (p. 21).

To be more explicit, “CT is thinking that has to be done in a reasonably sound way and in a way that experience and analysis has shown tends to produce true conclusions” (p. 23). In other words, performing CT better means drawing more true conclusions than false ones from any given set of problems. Performing CT more efficiently means being able to do this more speedily and reliably than before or than others. We could say, then, the criteria for improvement in CT are the greater quality of conclusions drawn by students and the efficiency with which they are drawn.

#### **2.1.3.1.2 Measuring CTS improvement**

To establish whether and how improvement is in fact occurring in students’ CTS, we need ways of measuring such improvement. In other words, we have to be able to operationalize CTS<sup>5</sup>. Several attempts to devise such measures have been made. They include common written tests (essays), direct classroom observations, individual interviews, and student and teacher journals. Although techniques for gathering information on students’ CTS can come in a variety of forms, the most objective, standardized measures, however, are multiple-choice tests (e.g. Watson-Glaser Critical Thinking Appraisal, Cornell Critical Thinking Test, California Critical Thinking Skills Test, and the College Assessment of Academic Proficiency tests).

The objectivity of these tests is based on the fact they have received, more than other CT measuring techniques, constant evaluation with regard to two major indicators of quality

---

<sup>4</sup> Fisher & Scriven’s CT definition: “A skilled and active interpretation and evaluation of observations and communications, information and argumentation”. (1997, p. 21)

<sup>5</sup> One makes something operational by providing ‘operations’ or procedures by which it may be observed, measured or distinguished from other things; a standard example being, ‘intelligence’ is what intelligence tests measure” (Colman, 2006).

information in educational evaluation: validity and reliability<sup>6</sup>. As regards to this, two experts on CT evaluation have said that “these tests have been carefully developed and are likely to serve more effectively than hastily constructed instruments” (Norris & Ennis, 1990, p.55). In addition, these objective tests assume that CT skills can be made operational and seek to establish the extent to which this has been accomplished. A basic epistemological precept guides the design of such tests. It is the notion that some answers or solutions are more verifiably correct than others (Pascarella & Terenzini, 2005). Tests guided by this basic precept are the best means currently available for measuring and improving CTS operationally.

These tests can be used in different ways to measure CT performance. In the case of the present study, it is necessary to make the distinction that ours is a study of changes in students’ CTS performance over time. Measuring change in CT performance requires assessing each student’s expertise at various points. Ideally, we would make repeated measurements during the course of a training regime. However, in practice this has not been done and generally there are at most two forms of test for the development of skills during a training period. Therefore, we have focused on research studies that have measured CT gain by pre and post-testing of skills, using the standard CTS tests mentioned. This means, studies that have measured students’ CTS at the beginning of a course or academic period and then again at the end of it. In this sense, CTS improvement means *to get better on CTS tests within a given, standard time interval*.

### 2.1.3.2 ‘Causing’ Improvement

- For improvements in CT to be attributed to philosophy, we must be able to demonstrate that studying philosophy is *causally responsible* for the improvements in question.
- We must, in other words, be able to apply the general concept of a causal factor/hypothesis to the matter, as this is commonly understood in the social sciences (Giere, 1997).
- To do this, we must be able to specify and isolate the *key variables* – which are the amount of study of philosophy, and amount of gain in critical thinking.
- We *also* need to know how much the students in question would have improved *anyway*.

---

<sup>6</sup> The reliability of a test is the consistency of student performance, using the same individual from one administration of the test to another. The validity of a test is the degree to which scores on the test can be trusted to mean what they are purported to mean. (Norris & Ennis, 1990)

### 2.1.3.3 Improving CTS ‘through the study of philosophy’

By the phrase ‘through the study of philosophy’ I mean, for the purposes of the present study, the direct effect on CTS improvement that the study of Anglo-American analytic philosophy can produce, given the considerable cognitive discipline it demands. The aim is to determine whether the amount of practice students receive when studying Anglo-American analytic philosophy generates greater CTS growth on CT tests from beginning to end (in an academic semester or any other academic period) than is generated by the kinds of practice received by students who study any other discipline, or those who do not attend college at all. When referring to ‘the study of philosophy’, I mean the specific contribution of such study to the net improvement on CTS.

Obviously, what we have here is a causal relationship between two variables (the amount of philosophy study and the amount of CTS gain). Ronald Giere (1997) has determined that when this type of relationship is set between the variables in a study, a causal hypothesis has been formulated. A causal hypothesis is a statement (claim, assertion, conjecture) about a relationship between a causal factor and an effect. In simple words it means that “causes, as we shall say, produce their effects or at least contribute to their production” (p.199).

The phrase, “Improving CTS through philosophy” implies, then, that we are dealing with a causal hypothesis. The *causal factor* is the study of Anglo-American analytic philosophy. The *effect* we refer to is the students’ CTS improvement. In fact, however, we are dealing not only with one causal hypothesis, but two. A standard hypothesis would be “the study of philosophy improves CTS”. With this we would only be deciding whether a causal link exists between the two variables at all. However, the present study has also been framed to provide an answer to a stronger hypothesis: “The study of philosophy improves CTS *better than other types of study do*”. It is not merely whether it causes some gain, but the extent to which studying philosophy causes a demonstrably greater gain in the students’ CTS than does the study of other disciplines, or than does studying of CT itself.

Even though the key terms at work in the assumption have been clarified, the evaluation of the causal relationship described above faces a number of difficulties. Such difficulties are not uncommon in the social sciences, when their practitioners undertake research on cause-effect relationships. Philosophically, the whole problem of causation is complex even in the physical sciences, and there are many possible sources of error and illusion. Such problems are exacerbated in the social or human sciences, because the intentional nature of human behaviour and interaction complicates the relationships between variables beyond what we see in the physical sciences.

The study of critical thinking skills is an exercise in social science research. It involves monitoring many variables in human cognition and behaviour. The difficulty of isolating specific variables and determining their associations is further complicated by other problems: the

difficulty of obtaining accurate and well selected observations, the danger of overgeneralization from any given set of data; ambiguity in conceptual and operational definitions; the challenge of measuring variables in a quantitative way, or determining their relationships in a qualitative way. We will have occasion, in this study, to be aware of these problems at two levels: first, in assessing the extent to which various studies have been flawed; second, in seeking to avoid such problems in conducting the actual meta-analysis of the available data.

## 2.2 The Nature of the Assumption that Studying Philosophy improves Critical Thinking

### 2.2.1 The Nature of the Assumption

Having established working definitions of our key terms, we are clear, then, that the object of the present study is to explore the proposition that:

Studying Anglo-American analytic philosophy, in the standard manner in which that is done by an undergraduate at a typical Anglo-American university is causally responsible for substantially improving the critical thinking skills of students

Let us call this the weak thesis.

There is, however, a stronger thesis at stake: that such study is practically speaking, the *best* way to improve critical thinking. This stronger thesis can be formulated as follows:

Students studying Anglo-American analytic philosophy improve more than they would have improved had they, instead, engaged in a comparable amount of any other activity they might reasonably have undertaken.

Invoking this notion of “any other activity they might reasonably have undertaken” invites a large, even indefinite number of comparisons. What if those students had, instead, been

- studying anything
- studying science
- studying psychology
- studying journalism
- studying mathematics
- playing chess
- debating
- playing video games
- working at a trade.

It was not feasible, in this context, to compare studying philosophy with such a wide range of alternatives. A more general and somewhat narrower focus was chosen, instead: on the most immediate and “telling” comparisons.

Strong (a):

Students studying Anglo-American analytic philosophy improve more than they would have improved had they, instead, been studying subjects other than philosophy.

Strong (b):

Students studying Anglo-American analytic philosophy improve more than they would have improved had they, instead, been studying critical thinking in its own right.

There is, however, some data, gathered by Pascarella, on the improvement to be expected in CTS among those who do not study at all. This also serves as an interesting point of reference.

### 2.2.2 Key Questions

Such preliminary analysis suggests that we ought to be able to address three broad research questions:

by **how much** do philosophy students improve their critical thinking skills in a single semester?

how much do **non**-philosophy students improve their critical thinking skills in a single semester?

how much do **CT** students improve their skills over a single semester?

## 3 Questioning the Assumption

### 3.1 Why the Assumption Should be Questioned

Philosophy is in the assumption-challenging business. It would, therefore, be rather incongruous were it to fail to examine one of its own most cherished assumptions. This is more than usually true of the present assumption, since this goes to the heart of the public *justification* for teaching philosophy. It is surely, therefore, of more than merely academic interest. It is of *practical* interest, in so far as philosophers wish to know how *well* they are doing. It is of *intrinsic* interest, since, were it to turn out that the assumption is not warranted, philosophers would be compelled to re-examine not only *how* they teach, but *why* they teach at all.

### 3.2 Arguments for the Assumption

Almost a century ago, Bertrand Russell claimed that “it is exclusively among the goods of the mind that the value of philosophy is to be found; and only those who are not indifferent to these goods can be persuaded that the study of philosophy is not a waste of time.” (Russell, 2001). In the same spirit, Simon Blackburn asked recently: “How is philosophy learned? A better question is: how can thinking skills be acquired?” (Blackburn, 1999). What was implicit in Russell’s remarks is explicit in those of Blackburn: to learn philosophy well is to acquire better thinking skills.

The field of thinking skills is wide. However, philosophy seems to concentrate on two central skills: reasoning and arguing. The refinement of both these abilities is indispensable to thinking in a sophisticated manner and, one might expect, that, in particular, the discipline of philosophy contributes to the improvement of such abilities. Indeed, reflection on and development of such skills stands, after all, at the very origin of Western philosophy, with the path-breaking thinking of Pythagoras and Parmenides, Plato and Aristotle.

Partly for this reason, and partly because they are in growing demand now, one of these skills tends to be especially utilized in promoting the value of studying philosophy, or even philosophy as a career. That skill is critical thinking skills. It is enough to look at the website of almost any of the philosophy departments at the major universities – at least in so far as they teach Anglo-American analytical philosophy.

The belief that philosophy improves critical thinking skills appears to be a common assumption among philosophers. Given the origin of philosophy and the mental disciplines it seeks to impart, it seems a natural enough assumption to make. What is surprising is that,

natural though it may be, this assumption has seldom been questioned. Consequently, the degree to which philosophy, or any particular way of teaching philosophy, enhances critical thinking skills (CTS) remains uncertain.

While the central purpose of the present study is precisely to question the veracity of the philosophers' assumption in this regard, it can readily enough be demonstrated that the assumption of a relationship between the two concepts – philosophy and critical thinking skills – is, at least, not arbitrary. The assumption is grounded in a number of intuitively plausible beliefs and it is important to explore these.

### **3.2.1 Conventional Wisdom**

It is the conventional wisdom among philosophers themselves that philosophy improves CTS. In this case, 'conventional wisdom' implies that the opinion of the experts in the field is such as to be widely and unquestioningly accepted by others. The testimony of the experts on this point is clear, not only in informal observations or claims, but in the way they tend to define the very *term* philosophy, as well as in their reflections on the relationship between philosophy and CTS.

American philosopher Matthew Lipman not only claimed that philosophy was a unique discipline, which prepares us to think well across disciplines; but went so far as to assert that philosophy is a better way to teach critical thinking than are dedicated CTS courses taught by non-philosophers. He wrote, to be precise: "It is only within the humanistic context of philosophy that students can experience the cultural relevance and methodological rigor that can be lacking, when, for example, "critical thinking" courses are offered by non-philosophers, or when thinking skills are taught in isolation"(Lipman, 1988), p.34).

The claim that philosophy improves CTS has been advanced not only by the independent voices of particular philosophers, but as the collective opinion of the academic associations which speak in the name of the philosophical profession. The APA (American Philosophical Association), for example, has stated that: 'Typically, philosophy teachers encourage students to be critical' (Audi, 1981). From the belief that this should be the case, or is perceived to be the case, it does not, of course, follow that it is what happens. Nor is it clear, even if philosophers do encourage their students to be critical, that the critical thinking *skills* of such students improve as much as they might by other means. What is clear, however, is that there is a common perception among professional philosophers that what they teach and how they teach it must improve CTS in students.

### 3.2.2 Anecdotal Evidence

Secondly, there is anecdotal evidence that the teaching of philosophy does, in fact, improve CTS in students. This evidence is based on direct experience. In many cases, professors of philosophy have dedicated years to the teaching of CTS and the dispositions that enhance it. This depth of teaching experience may or may not consist in explicitly seeking to teach CTS, but it gives the practitioners, all the same, at the point of evaluation, considerable confidence in their own capacity to judge the CTS of their students and whether they improve, based simply on their own observations.

Doubtless, over the course of a career, the professors are able to observe improvements in various students with respect to certain skills and dispositions relevant to critical thinking, as exhibited, for instance, in response to or the asking of tutorial questions and the writing of essays and assignments. They are, also, able at times to observe or teach students – and even other academics – who are not trained in philosophy and to notice a significant contrast in the CTS of these groups compared with those of the philosophy majors. However, as we shall see, such anecdotal evidence or overall impressions can be quite misleading and need more rigorous examination if our confidence in them is to be vindicated.

### 3.2.3 A Priori Grounds

Thirdly, the acceptance of the statement that philosophy improves CTS has been based on a *priori* grounds. In this case, the assertion is not grounded in the kinds of experience or observations we have just outlined. Rather, it is accepted because it is believed to be analytically true; that is to say, true by the virtue of the meanings of the words contained in the proposition and gives no information about the empirical world. A typical case of what is meant here by analytical or *a priori* truth is the proposition: “Bachelors are unmarried”; that is, by definition, “being unmarried” is entailed in the concept of “bachelor” (Norris, 1992a).

There are contexts in which the assertion that philosophy improves CTS is considered as analytically true. Philosophy and critical thinking are, for example, considered to be essentially the same thing. For instance, the claim might be put as follows: philosophy is critical reflection, and CT involves both a critical disposition and the skills needed to act on it. A well-known example of this is the claim by Richard Paul that philosophical thinking and a strong sense of critical thinking are essentially the same.<sup>7</sup> Of course, once such a premise is accepted, it is

---

<sup>7</sup> A strong sense critical thinker is a concept coined by Richard Paul to design the thinker who is able to explicate, understand and critique their own deepest prejudices, biases, and misconceptions, thereby discovering and contesting their own egocentric and sociocentric tendencies. According to Paul, it is through a philosophical teaching approach that such a thinker can be generated.

natural to conclude that the teaching of philosophy must, in the nature of the case, improve CTS – provided, always, that such teaching is done well.

### 3.2.4 Philosophy Involves Practicing CT Skills

Fourthly, the assumption that philosophy has an impact on students' CTS has also been defended on the following grounds: philosophy and CT are *not* exactly the same; but, in the study of philosophy, students extensively *practice* certain core CT skills under the guidance of experts. The skills in question, of course, are those of reasoning and argumentation. Both of these abilities are indispensable to learning how to think and, one might expect, that, in particular, philosophy contributes to the improvement of such abilities in the course of a student's university studies. Philosophers, starting in ancient times, initiated the practice of arguing for their conclusions and against other philosophers' conclusions and demanding reasons for the claims of others, while giving reasons for their own. The importance in philosophy of reasoned argument lies in the crucial role played by it in understanding.

To achieve an understanding of something, one has to be able to give an account and ultimately a *rational* account, of the thing or belief in question (Annas, 2000). If reason is crucial for advancing the understanding, it is equally fundamental to the acquisition of knowledge and, in short, to the quest to grasp and master the ultimate truth of things. It is because it has this indispensable role in the quest for understanding and knowledge, that one observes the great emphasis on reasoning and argument in the schools of philosophy throughout history. At the present time, when CTS is ever more important across domains, it is not hard to see the overlap between what it requires and the traditional preoccupations of the discipline of philosophy.

Of course, what CTS requires, above all, is advanced facility in reasoning and engaging in argument. Perhaps, the overlap here is due precisely to the contribution of philosophy to critical thinking (A. Fisher, 2004; R. Fisher, 2003; Lipman, 2003; Milkov, 1992; Talaska, 1992; Walters, 1994). This contribution can be specified with reference to the fact that the movement to promote "Critical Thinking" or "Informal Logic" grew out of the efforts of professional philosophers to coach students in applying formal methods of reasoning to real world reasoning problems (Fisher, 2004). It is, predictably enough, for this reason that most courses in CT typically have been taught by academic philosophers and many of the standard textbooks in CT are authored by philosophers (Walters 1994). Taxonomically, CT seems to be best placed under the rubric of 'applied philosophy' (Lipman, 2003).

The crucial link, however, is that, when well taught in philosophy, students are required to extensively *practice* the core skills of reasoning and making or analysing arguments – the central CT skills. Such practice is highly likely to lead to improvement in reasoning and argument skills, and thus to gains in CT abilities. This is a vital consideration, as we shall see; since it is the nature of such practice, its quality and quantity, rather than the study of philosophy

per se, which turns out to be the indispensable condition for the improvement of CTS. There is also an additional element in support for the claim that philosophy tends to help improve CTS. It is the fact that philosophy students receive the guidance of experts in reasoning – that is what professors of philosophy are. This element requires special attention.

### 3.2.5 Instruction by Experts

As a rule, a professor of philosophy is a philosopher himself and has received a rigorous training in the nature and practice of arguing and is imbued with the value and purpose of good reasoning. This quality sets them apart from the general population. Various studies have shown that people in general do *not* possess good reasoning skills (Kuhn, 1991). Furthermore, there seems to be good evidence that many people not only lack good reasoning skills, but simply do not understand what good reasoning actually *is*. Some people, for instance seem to think that being ‘reasonable’ means seeing both sides of an issue, without necessarily understanding either of them; while others appear to believe something is ‘reasonable’ if it conforms to accepted rules, traditions and procedures, without requiring that these things themselves have any kind of rational justification. There are also cases in which students and professors may have very different ideas about what it means to exhibit good reasoning (Garlikov, 2002).

In general, we are surely justified in thinking that a professor of philosophy, who has received a rigorous training in the analysis of the great problems of philosophy and the traditional methods for exploring them, knows what reasoning and arguing are. This training should give such a professor the ability to determine whether a student is reasoning well or badly. By contrast, as Garlikov observes, “it is crucial to understand [that] it is often not helpful to try to improve the performance of someone who is not even doing [at all] what you are trying to get him to do better.”(p.1). The same author presents an instructive analogy to bring out the meaning of his point:

“If little children are playing with chess pieces and a chess board, but are making arbitrary moves in what they think is emulation of adults they have seen playing chess, it is not just that they are playing chess badly. It is that they are not playing chess at all, regardless of what they think they are doing or what they call it”. (Garlikov, 2002).

For this reason, a professor who is unable to distinguish between good and bad reasoning, or to specify and teach what is involved in reasoning and arguing, may well be able to generate *changes* in the cognitive behaviour of his students, but will not necessarily be able to *improve* their reasoning. From all of which, it should follow that philosophy professors are better placed than other academics to teach students the science of good reasoning – that is the nature of their craft and expertise.

### 3.2.6 Adequacy of Framing the Problem in These Terms

There are those, of course, who would argue that both philosophy and critical thinking are too narrowly defined in what I have just said; and that the relationship between the two is not necessarily what I have sketched out. Indeed, there are some who claim that there is no such relationship at all. Whatever the merits of that claim, it can certainly be allowed that the two terms –philosophy and critical thinking – have generated a series of distinct or competing definitions, at least some of which do not necessarily have a positive cause and effect relationship. However, for the purposes of the present project, in which we have defined the term philosophy as Anglo-American analytical philosophy and the term critical thinking as the practice of informal logic, the assumption of such a relationship, as we have shown, seems entirely reasonable.

What is more important is that, just to the extent that there is a positive relationship between the study of philosophy and the development of critical thinking skills, it ought to be *most* in evidence where the two are defined as I have suggested. If it is not the case under these definitions, there are surely *prima facie* grounds for doubting that it is very likely to hold under other definitions of the two terms. What the present study is chiefly concerned with then, is taking these definitions as a kind of limiting case and then examining closely the extent to which the assumed relationship is in fact operative. For to agree that an assumption is natural is not to demonstrate that it is warranted. To satisfy ourselves that the assumption is in fact warranted, we surely must subject it to the tests of rational belief to which both philosophers and critical thinkers, at least as I have defined them here, are committed. What follows is precisely that: a critical reflection on the argument that there is a positive relationship between the study of philosophy and the development of critical thinking.

### 3.3 Problems with the Arguments

The argument that has been set out in the preceding section of this chapter shows that it is natural and not unreasonable to assume that studying philosophy improves critical thinking skills. However, the grounds on which that assumption is being made are not beyond question.

I do not wish to give the impression that there is any intention here to try to demonstrate that the assumption in question is false. Nor is there any intention, on the other hand, to demonstrate that this assumption is neither natural nor reasonable. After all, I have just shown that it is based on plausible grounds. Plausibility, however, should not satisfy us in such a matter; for plausibility is not the same as truth. It is my intention in this section to closely examine the grounds on which this ever so natural assumption is based.

Until we conduct such an investigation, we cannot claim to know whether the assumption is true or not. As it happens, if we reflect critically on the underpinnings of the assumption (that the study of philosophy improves critical thinking skills) several problems arise. These problems will

be presented in this section in the same order in which the arguments in favour of the assumption were presented in the preceding section – as objections to these very arguments.

This is simply with a view to enabling the reader to follow the exposition of these objections with a minimum of difficulty.

### **3.3.1 The Argument from Conventional Wisdom**

The first such belief is based on the conventional wisdom among philosophers that philosophy improves CTS. Although this conventional wisdom is the opinion of experts, there are two objections to it. Firstly, merely because something is the conventional wisdom it does not follow that it is true. Moreover, even if it should prove to be true in some sense, it makes a lot of difference precisely how true it is. It could, when closely examined, turn out to be exaggerated or misleading in significant ways. This objection is well known and was long ago expressed in Latin as the logical fallacy of “argumentum ad populum” (appeal to the people).

Manifestly, it is fallacious to assert that a proposition is true simply because many or all people believe it. In the case of philosophers believing as true the statement that philosophy improves CTS, although they could happen to be correct, it could turn out to be the case that they merely believe it to be true because it has been repeated over and over without ever having been verified. What if it turned out that, never having been rigorously examined, this belief was actually a myth, albeit one which has become a fixed idea among philosophers at large?

There is a second objection to the conventional wisdom that philosophy improves CTS. It is the fact that philosophers, though experts on philosophical issues, are not themselves experts on technical issues such as cognitive gain. Critical thinking, as it has been defined, is a set of cognitive skills and dispositions that can be made operational and are, therefore, measurable. This being so, the assumption that philosophers teach CTS better than anyone else, or that their discipline is better suited to teaching it than any others, is not something the truth of which can be determined merely on the basis of the testimony of philosophers themselves. It must and can be measured by means that are, in themselves, independent of the opinions of philosophers or of philosophy as a discipline. Even if it was the case that the testimony of experts was sufficient grounds for believing their claims about the value of what they do, the testimony of philosophers would be no more compelling than that of scientists, doctors or lawyers that their particular approach was the best way to improve CTS. In short, we need some means other than such testimony to justify a belief of this nature.

### 3.3.2 The Argument from Anecdotal Evidence

Reliance on anecdotal evidence to support the idea that the teaching of philosophy improves CTS in students is just as dubious as reliance on the testimony of philosophers. Anecdotal evidence is based on personal experience used to illustrate a point. Of course, there is nothing wrong with presenting cases or examples to illustrate a conclusion, mainly if they are based on direct experience. But, they are insufficient to prove that a general conclusion is true. They are, at best, partial and incomplete evidence, which needs to be carefully examined and supplemented by more systematic data.

There are various forms of cognitive bias that may affect the conclusions drawn from anecdotal evidence. It is plausible that an anecdote can illustrate a desired conclusion rather than a logical conclusion. The misuse of an anecdote in this manner constitutes a common logical fallacy, i.e. a flawed kind of argument. This is not to say that such logically flawed arguments are intended by those who use anecdotes; only that they are entailed in the reliance on anecdotes. Such reliance tends to be a manifestation of the universal human tendency to pick up evidence that supports our own preconceptions, as pointed out by Francis Bacon, in his famous book, *The New Organon* (Bacon, 1905).

It is perfectly possible that those who assert a positive relationship between philosophy and CTS based on their own direct experience, have formed a false impression, albeit with the best of intentions, because they are relying on anecdotal evidence. In other words, it is at least conceivable that, quite unselfconsciously, just because they want to see a positive relationship between philosophy and CTS, they tend to perceive or give undue weight to only that evidence which suggests such a relationship. Naturally, once they have come upon such evidence, they have little difficulty in drawing the desired conclusion from cases specifically chosen to support it.

The problem lies in the corresponding tendency to ignore evidence that might undermine the preferred conclusion. Fairly clearly, as Bacon pointed out, this neglect of what would disappoint our expectations, or the tendency to believe only that which fits our preconceptions, could be attributed to our pride or our prejudices. Needless to say, if we are committed to genuine knowledge, we must be prepared to put aside our pride and overcome our prejudices. Even if we take precautions to do these things, there remains a margin for error. There is an irreducible problem, in that no given piece or arbitrary cluster of anecdotal evidence can in itself be regarded as sufficiently representative, from a statistical point of view, to constitute a compelling argument.

### 3.3.3 The Argument by Definition

There remains a third problem with the argument advanced in the previous section, namely a question mark against the assertion that it is necessarily the case that philosophy improves

critical thinking; in other words, that the assumption we are discussing is analytically true, or true by definition. The question mark is this: how could something which requires evidence to substantiate it be true by definition? One cannot have it both ways: if evidence is adduced to support the assumption in question, that can only be because it is *not* true by definition. Conversely, if it was true by definition, there could not be relevant evidence in the case; but there clearly is.

Moreover, suppose one accepted the tautologous proposition that the study of philosophy necessarily improves critical thinking because it *is* critical thinking, one would still require evidence as to what *kinds* of philosophy or philosophical study improve philosophical (critical thinking) skills. In other words, tautology does not do the work required here. But, in any case, it is untrue to assert that philosophy is, by definition, the study of critical thinking. There is, of course, a relationship between the two, a certain evident affinity; but precisely what that relationship is remains to be clarified. Proponents of the view that the study of philosophy improves critical thinking would not make this claim if were true by definition, since it would be merely stating the obvious; nor would they advance evidence, however incomplete, since this would be equally pointless. In short, there is a case that needs to be made. It remains to examine closely such evidence as has been gathered to date.

There is a further reason why we cannot simply accept a claim by philosophers that their discipline is critical thinking. It is that the practitioners of other disciplines might make the same claim, i.e. that their own discipline is critical thinking. It is surely plausible, for example, that teachers of the pure or applied sciences might make such a claim, or that statisticians might do so, or perhaps historians. Indeed, such claims are made regularly. They run along the lines that scientific (or statistical, or historical) reasoning is the same as critical thinking, given that critical thinking consists of the ability to reason, to evaluate and to construct argument. These competing claims surely cannot be adjudicated by mere presumption or assertion, but need to be closely examined and tested against carefully evaluated evidence as to the effects of studying one discipline compared with another.

### **3.3.4 The Argument that Philosophy Provides the Right *Practice***

Coming back to the core argument, let's consider the fourth reason that is offered in support of the assumption that the study of philosophy improves the development of critical thinking skills. This is that, while philosophy and critical thinking may not be the same thing, nonetheless, in the study of philosophy, students extensively *practice* certain core CT skills. Given the ancillary assumption that the practice of a skill improves it, we would naturally infer that the study of philosophy does, indeed, improve critical thinking skills. However, if we subject even this reason to the tests of rational belief there are some objections to consider.

First of all, while it may be true that those studying philosophy practice certain CT skills, the reality is that such skills are practiced in many different ways, formal and informal, all the time. Philosophy is one of the more formal approaches to such practice, but what we do not know with sufficient confidence is whether its particular formal approach really makes the difference claimed. One would hope that such formal practice would make a distinctive impression, but there is simply no way to be certain that it does unless we submit it to empirical scrutiny. The same kind of question might be asked with regard to any other skills. Consider the example of a taxi driver's skills. They clearly improve relative to the driving skills of others, but they do so almost entirely through informal practice. It may be that formal practice would improve them further, but it is not self-evident which kind of formal practice would do so. So it is with critical thinking skills: they can improve through informal practice; they might well improve more through formal practice, but it is not self-evident that philosophy provides just the right kind of formal practice. It needs to be demonstrated.

Consideration of what kind of formal practice is most likely to improve critical thinking skills raises the subsidiary questions as to both the quality of the practice in question and the quantity. If what is crucial to demonstrating that philosophy improves CTS, is that, in philosophy, students extensively practice the core skills of reasoning and making or analysing arguments, it remains to be shown just how extensive such practice is and how much it improves CTS. This last consideration (how much CTS improves) will depend not simply on how much practice students get, but on the quality of that practice. We must assume that the quality of practice that students receive will determine the extent to which their skills develop. Consequently, we can say that unless the conditions of an extensive quality practice are fulfilled, teachers of philosophy (or for that matter of any other CT course at university) cannot reasonably claim that their course is especially good at producing gains in CTS.

### **3.3.5 The Argument from the Expertise of Philosophers**

In the light of this last consideration, we might also challenge the final component of the claim that philosophy tends to help improve CTS. This component is the belief that philosophy professors are better placed than other academics to teach students the science of reasoning. First of all, we can question the belief that philosophy professors necessarily become experts on reasoning, as distinct from becoming experts on various arcane theories of metaphysics or morals, and are thus better placed than other academics to teach CTS. Secondly, even should it be the case that philosophy professors do become experts in reasoning skills, we might still require evidence that this makes them good at *teaching* critical thinking skills.

The nature and quality of teaching is important because it is precisely the professor, in his or her role as a teacher, who is responsible for providing that quality practice without which the development of the skills in question would not occur. The empirical evidence clearly shows that attending lectures on CT is not sufficient to develop CT skills. (Butchard, 2006; Donohue, Van

Gelder, Cumming, & Bissett, 2002; Harrell, 2004; Hitchcock, 2003; Rainbolt & Rieber, 1997; Ross & Semb, 1981; Spurrett, 2005; Twardy, 2004) The empirical evidence on this matter suggests, also, that there are other aspects of teaching that are important to the development of CTS. However, these need not detain us here. Rather, the idea to which I wish to draw attention here is simply the importance of teaching as such in the development of skills of this nature.

Let me underline this point: we cannot make a convincing case for the claim that philosophy improves CTS based merely on the assertion that it is taught by experts in reasoning. Teaching abilities in and of themselves are important to delivering both the central ideas and the kind of quality practice that the development of CT skills demands. There is no guarantee that a content expert in any given field will be able to teach that content well, much less the cognitive skills that underpin it. This must be held to apply as much to philosophy as to any other subject and to reasoning skills as much as to any other cognitive skills.

### 3.3.6 Summation

In short, there are serious grounds for questioning every one of the reasons offered so far in support of the claim that philosophy improves CTS. This is represented in a compact argument map in Figure 2 (see opposite). Conventional wisdom and anecdotal evidence are of their nature unreliable reasons for accepting anything as true. As we explained, they can both often be wrong, so that if we depend on them, we leave ourselves with beliefs that lack any solid basis. In the case of the claims that the study of philosophy improves CTS because it provides both practice and instruction by experts, we have seen that there are serious questions that require substantive answers. Are the students getting *enough* practice? Is it the right *kind* of practice? Is the practice they get in *philosophy* better than *other* types of formal practice? Or informal practice?

Further, with regard to the idea that philosophy professors are necessarily experts in teaching critical thinking skills, we must ask both: are they, truly, the best experts and are they, in any case, the best teachers? We are left, finally, only with the bald assertion that the claim that philosophy improves CTS is simply analytically true, because CT is philosophy and vice versa. This assertion is directly at odds with the assumption behind all of the other claims, which is that the impact of philosophy on the development of CTS is an empirical issue, i.e. one for which evidence is adduced for and against. There is no getting away from the fact that evidence does bear on this matter in a number of ways. For this reason, I shall argue that the proposition “philosophy improves CTS” is an empirical claim, which requires special attention.

**Figure 2. Questioning the assumption that the study of philosophy improves critical thinking skills.**

## 3.4 The Need for Empirical Scrutiny

### 3.4.1 This is an Empirical Question

There is one fundamental and compound reason why we need an empirical scrutiny to provide stronger support for the claim that philosophy improves CTS. The claim in question is an empirical claim and empirical claims require empirical scrutiny. They require such scrutiny in order to be verified and they need to be verified because they are open to possible empirical refutation, unlike claims which are simply true in the nature of the case. Let us examine this reasoning a little more closely.

An empirical claim is one which can be verified or refuted through the accumulation of appropriate and sufficient evidence in the world of experience. Such claims are also called a *posteriori* propositions or statements (Audi, 1999). In contrast, a belief or claim is said to be justified *a priori* if it has an epistemic justification, i.e. the reason or warrant for thinking it to be true does not depend on experience. The main distinction of empirical claims is the necessity to base their justification on empirical data, what Ronald Giere (1997, p. 27) called “data from the real world”; rather than definitions or deduction. In other words, empirical claims depend upon data derived from observations, or experiments, not generated by pure theory.

We say the claim that ‘philosophy improves CTS’ is an *a posteriori*, or empirical claim based on the following reasoning. The claim mentions something called ‘improvement’ and improvement is something which requires measurement. To measure improvement in the skills in question we must obtain empirical data. Improvement, as it was defined in section 2.1.3.1 means *to get better on CTS tests* within a given, time interval.

### 3.4.2 Two problems for the investigator

This poses two problems for the investigator. First, how to measure the *growth or change* in students’ skills; second, how to establish a cause and effect relationship between the method (philosophy) and the outcome (improved CTS). Relations of change cannot be established by means of static descriptions and associations. To be able to analyse change, we must first be able to measure it and we cannot measure a change of this kind if we lack empirical data (Plewis, 1985).

But we do not lack empirical data in this case, as it happens. The fact that CT skills can be operationalized and that there exist tests which enable us to measure them gives us the option of establishing ‘investigable facts’ in the case. These facts alone show that the claim is empirical, which is to say that there is evidence concerning it. We have, therefore, the option of looking at empirical data, such as the results of experiments that have systematically measured

students' critical thinking skills and have at least attempted to determine the extent to which philosophy has improved them, or at any rate caused changes in them. It is in this sense that the claim is empirical and, being empirical, requires empirical scrutiny.

It is important to underscore the fact that, when we measure change or growth, we are implicitly seeking to measure a cause and effect relationship between two variables. In this case, the two variables are the amount of philosophy study that students get and the amount of CT gains those results from such study. Causal relationships require experimental research designs in order to be tested (Giere, 1997, p.214). Experimental research designs measure the properties of variables in a *systematic and objective* way. We are then able to statistically determine the extent to which the variables are related (Hessler, 1992).

Given, then, that the claim 'philosophy improves CTS' is an empirical one, it requires empirical scrutiny, because it stands in need of verification and empirical scrutiny is the means to do this. Empirical scrutiny requires the accumulation of evidence *for and against* any particular hypothesis or claim. It involves planned methodological research designs for the collection of data. There are, moreover, rigorous standards for ensuring that such research design and data collection are objective<sup>8</sup> and systematic<sup>9</sup>.

These standards are meant to ensure that data is collected and utilized in a precise and reliable manner. If empirical scrutiny were only a matter of collecting evidence of some kind, then anecdotes and testimony derived from personal experience would be sufficient. However, they are not sufficient. Empirical scrutiny, properly so-called, must be objective and systematic, if it is to verify any given claim and enable us to assert it with sound justification.

For all these reasons, the next task in our investigation is to assess the empirical evidence presented so far in support of the claim, in order to determine how much support such evidence really provides for it.

---

<sup>8</sup> Objectivity in this context means the collection of observed facts not influenced by the emotions or prejudices.

<sup>9</sup> Systematic means according to a plan, not casual or sporadic or unintended.

## 4 Review of the Existing Evidence

We shall turn, shortly, to a careful review of the existing evidence, but because the findings of this review are so central to the present thesis, let us summarize them briefly first. With the exception of the Philosophy for Children program (P4C), there has been strikingly little interest in empirical evaluation by philosophers of the extent to which their discipline actually enhances critical thinking. At present, the volume of published work in connection with P4C pedagogy shows a positive impact on children's reasoning skills (Garcia-Morison, Rebollo, & Colom, 2005). Two recent meta-analyses support this claim.

However, I have not been able to find a correlation between the reasoning tests used to evaluate the program for children and the current tests used to evaluate CT. This makes it difficult to diagnose to what extent the evaluation includes all the abilities that are attributed to CT. Nonetheless, given that reasoning is a fundamental component of CT, the findings obtained in this area will be presented as an indication of the possible effect that P4C could be having on CT.

With regard to undergraduate and graduate students, only a few studies of the impact of philosophy on critical thinking were found. This is despite the fact that philosophy departments in Western universities commonly claim that philosophy has this effect. There is also a striking contrast between the abundant research regarding the impact of social sciences, especially psychology and education, or nursing, for example, on critical thinking and the paucity of research done on philosophy (and, for that matter, the hard sciences) in this regard.

This paucity of empirical studies of the impact of philosophy on the development of CTS may well be an indication of just how readily it is *assumed* that it has such an impact. The assumption is so deeply held that there is no felt need to check it against evidence. The few studies that exist may, in any case, have been taken as adequate to make the case, since they were not closely compared with evidence from other disciplines. They may not, however, be at all adequate to make the case.

With respect to such studies as have been made of the impact of philosophy on the development of CTS in undergraduate and graduate students, the evidence is more open to question than it is in regard to Philosophy for Children. This is, firstly, because the methodologies used in these studies are more divergent. Secondly, and not surprisingly, the findings of these divergent studies are themselves divergent. Some studies present positive evidence with regard to our question, but some others report negative or inconclusive findings.

Let me set out in some detail the empirical review of the literature from which the above observations are derived.

## 4.1 A Review of the Literature

This literature review focuses on the most relevant empirical investigations designed to test the assumption that philosophy improves critical thinking skills. Two questions have guided the preparation of the review. Firstly, what *evidence* have these recent investigations found regarding the relationship between the study of philosophy and the improvement of critical thinking skills? Secondly, what *conclusions* can we draw from the various findings of such investigations? In other words, the aim is to establish what kind of evidence we have, and what it is possible to infer from it.

The best recent studies make clear that various researchers have approached the matter of the impact of the study of philosophy on CTS from somewhat different angles. In some studies, the research question has focused on determining the *direct* effect that the study of philosophy appeared to have on CTS. In other studies, the effect was measured in an *indirect* way, because it was not the primary objective of these studies to explore the connection between philosophy and CTS. In these cases, the research findings nonetheless helped to clarify the causal relationship between philosophy and CTS, since they were concerned in some manner or other with the influence on, for example, general reasoning and critical thinking skills of such specific studies as formal logic.

It is not, however, this distinction between kinds of investigation which I have used to structure the present review. Rather, the key organizing criterion has been the academic level at which the investigations were targeted. The natural levels in this regard are: (1) primary and secondary school, (2) college or under-graduate university, and (3) graduate school. The reason for organizing the materials in this manner is that it enables us to focus on the extent to which philosophy impacts on CTS at different stages in the cognitive development of young minds.

## 4.2 Evidence from Primary and Secondary Students

Matthew Lipman's well-established program Philosophy for Children (P4C) is, surely, a natural place from which to start assessing the impact of philosophy on the development of CTS in the very young. But, before presenting the empirical data concerning the effectiveness of the program, it will be useful to review how it works.

The aim of Lipman's program is to teach children to reason through guided discussion of familiar topics and concepts (Lipman & Bynum, 1976). The procedure is clearly very much in the Socratic tradition and, consistent with that tradition, sees the development of reasoning ability as its central aim. P4C is delivered through the reading and discussion of novels by children and

adolescents. The curriculum varies, depending on the students' age. However, many stories have been purposefully written for teaching philosophy.

The students get together to read, discuss and ask questions; and the class acts as a 'community of inquiry'<sup>10</sup>. There are three assumptions underlying the program. First, that philosophy begins in wonder and so does childhood (Lipman, Sharp, & Oscanyan, 1980), suggesting that children are natural philosophers. Second, that children are capable of critical thinking, even at an early age. Third, that active discussion promotes critical thinking.

#### 4.2.1 The Effectiveness of Philosophy for Children

There have been many investigations into the implementation of the P4C program. The results tend to suggest a positive effect on the development of CTS from doing philosophy with children. The strongest support comes from the outcomes of two very recent meta-analyses, conducted in 2004 and 2005.

Despite these encouraging results, it is significant that a majority of individual studies of the Philosophy for Children program have been and are open to serious methodological criticisms. It cannot be said that these criticisms outweigh the results obtained from the two meta-analyses. The methodology of meta-analysis, after all, is what enables us to guarantee that the empirical studies selected meet a rigorous criterion related to research design, data analysis, and reporting. The problem is, however, that, in order to conduct these two meta-analyses of P4C, their authors had to exclude the vast majority of empirical studies that had been done, because they did not meet even the *minimum* criteria for methodological soundness (Reznitskaya, 2005). For the purposes of this thesis, I think it is important to consider both these criticisms and their implications.

It is not among the objectives of this thesis to present a detailed analysis of the various criticisms to which evaluations of the P4C program have been subjected. It might have been considered sufficient to merely cite the summary judgment by one of the major authorities on the subject that: "most reports do not follow the basic rules established by the scientific community for the presentation of research reports." (Garcia-Moriyon et al., 2005). However, I think it is important to make clear here at least the main criticisms that have been made of such studies. They have been of three general kinds: that there have been flaws in the methodological design of evaluative studies; a lack of information in the reports of such studies

---

<sup>10</sup> The 'community of inquiry' is a term coined by Matthew Lipman himself. By it he means that the classroom becomes a reflective community in which students use tools of inquiry (questioning) so that they can competently assess evidence, detect inconsistencies and incompatibilities, draw valid conclusions, construct hypotheses, and employ criteria until they realized the possibilities of objectivity with regard to value as well as to fact (Lipman, 1988).

regarding how they have interpreted their data; and problems in the application of statistical procedures in such studies.

It is my belief that reviewing these methodological flaws will serve several good purposes. First, it throws light on areas in need of further attention or investigation in future research studies. Second, and very importantly, it underscores the need in any such future studies to avoid the kinds of flaws and omissions that mar so many of the studies conducted to date. Thirdly, if more rigorous methods are used in future studies, we will benefit from research studies of a higher methodological quality, enabling researchers to draw stronger and more compelling conclusions. I shall, therefore, review the results of the most relevant empirical studies of P4C and then point out the criticisms that the majority of these studies have faced.

#### **4.2.1.1 The Evidence and its Characteristics**

The first experiment to measure the effectiveness of the Lipman program was designed and implemented by Lipman himself, in 1970, as a pilot study. He applied the treatment to two fifth grade classes. One was an experimental group, using Lipman's materials, and the other was a control group, working with social studies materials. The California Test of Mental Maturity (CTMM) was used. The experimental group showed significant gains<sup>11</sup> over the control group in the area of logical reasoning (Lipman & Bynum, 1976).

Since then, the program has been repeatedly evaluated. Philosophers and educators have tested its effectiveness in three general areas: reading, reasoning (or logical skills) and mathematics. Many in this long series of studies, measuring the impact of the Lipman program on the development of cognitive skills in children, are listed on the website of Montclair State University<sup>12</sup>. There are 56 studies in the Montclair list, spanning approximately 30 years of investigations.

Of these 56 studies, only three showed either negative or inconclusive results as regards the development of thinking skills. Of the remaining 53 studies, ten showed positive results in at least one of the following four areas: skills in reading, mathematics, and communication, or the progress made by students with learning disabilities or other special education needs. Seven

---

<sup>11</sup> Adjectives such as: 'significant gain', 'positive results', 'greater gain' used across the literature, are ambiguous and can lead to misinterpretations. This will be clarified in the section Criticisms of Philosophy for Children's Empirical Research.

<sup>12</sup> Montclair State University sponsors the Institution for the Advancement of Philosophy for Children (IAPC). The Institute conducts sponsors and advises theoretical scholarship and empirical research in teaching pre-college philosophy. In addition, IAPC is the head office of the Philosophy for Children Program.

were qualitative studies, and another six studies, showed no data or comments. The remaining thirty studies are the ones of greatest interest for the purposes of the present project. They consistently showed significant gains or positive results in the area of reasoning or logical skills.

Two meta-analyses have also been conducted to evaluate the program (Garcia-Moriyon et al., 2005; Trickey & Topping, 2004). In the Trickey & Topping meta-analysis, similar “positive outcomes” were found. The review included ten studies (three of which are included in the Montclair list) measuring outcomes in reading, reasoning, cognitive ability, and other curriculum-related abilities. The study showed a effect size of 0.43<sup>13</sup>, with low variance, indicating a consistent moderate positive effect for P4C on a wide range of outcomes (p.365).

In Garcia-Moriyon’s meta-analysis, 18 studies fit the following criteria: (1) studies testing the effectiveness of P4C in improving reasoning skills or mental abilities and (2) studies including enough statistical information to calculate the magnitude of effects sizes. All effect sizes except one are positive and the estimation of the mean effect size yielded a value of 0.58 ( $p < .01$ ; CI [ .53 , .64), indicating that P4C has a positive effect on reasoning skills<sup>14</sup>. (p.21)

#### 4.2.2 Criticisms of Philosophy for Children’s Empirical Research

Cebas et al. (2003), in their article “What we know about the research in philosophy for children”, highlighted a consideration that raises doubts about the quality of the evaluative processes used in regard to the program: *important information* is consistently missing from the published research reports that the two subsequent meta-analyses chose to exclude.

---

<sup>13</sup> An effect size describes the magnitude and strength (positive or negative direction) of a relationship between two variables.

<sup>14</sup> It is necessary to provide a benchmark to be able to interpret the effect sizes (ES) obtained in these meta-analyses. Cohen, 1988 attempted to address this issue. He labelled an effect size small if  $d = .20$ , medium-sized (moderate) if  $d = .50$ , and large if  $d = .80$ . According to Cohen’s taxonomy, both meta-analyses reported medium (moderate) effect sizes. However, Cohen chose these numbers to reflect the typical effect sizes encountered in the behavioural sciences as a whole. He warned against using his labels to interpret relationship magnitudes within particular social science disciplines or topic areas (Cohen, 1988). In the case of education, the results of the Michigan Meta-Analysis of Research on Instructional Technology in Higher Education (Witrock, 1986) found a moderate (medium) effect size (.49) for students in Keller plan courses. Despite of having being categorized as a moderate ES, it was found to be greater than those of different teaching strategies (Audio-Tutorials (.20), Computer based-teaching (.25), Program Instruction (.28) and Visual-Based Instruction (.15). This suggests that even though P4C effect sizes could be categorized in general as medium, in the education field, they reflect a considerable achievement.

With regard to this matter of lack of information, a couple of points are particularly worth mentioning. First, the Montclair List, published by the IAPC, the head office of the Philosophy for Children project, offers only brief abstracts of the original papers, with only basic and incomplete information, thus leaving obscure how these papers reached their conclusions. Most of them have been, in any case, been criticized for merely advertising the program and its potential efficacy, instead of critically inquiring about its actual impact. (Garcia-Moriyon et al., 2005). Moreover, the Montclair List seems to follow the tradition of vote-counting<sup>15</sup>, seeking rather to persuade the reader than simply to demonstrate empirical results.

Second, of the 116 studies selected for the meta-analysis by Garcia-Moriyon et al., he and his colleagues had to exclude the vast majority for not meeting the minimum criteria related to research design, data analysis, and reporting. Only 18 of the 116 fitted the criteria. This deficiency in so much in the P4C research has led Reznitskaya, 2005, to draw particular attention to the lack of appropriate statistical procedures in the great majority of these studies. This problem in regard to statistical procedures in assessing the efficacy of P4C calls for special attention.

#### 4.2.2.1 Statistical Considerations:

Many of the evaluative studies of the P4C program lack sufficient statistical information to guarantee that the results obtained have met the standard of rigour established by the scientific community for the presentation of research reports. It is not clear whether this problem is due only to the *presentation* of results without regard to academic standards, or the consequence of incomplete or faulty research *design* (Cebas, 2003).

The Montclair List, for instance, only makes reference to results using phrases such as “*significant gain*”, “*positive effect*”, “*significant improvement*”. These phrases are unacceptably ambiguous chiefly because they present a dichotomy in their meaning. On the one hand, the word ‘*significant*’ means of considerable amount or effect or importance”, suggesting that results are importantly large (Sykes, 1976). On the other hand, it can also mean that the results are “*statistically significant*”<sup>16</sup>. This implies that statistical techniques have been used to determine

---

<sup>15</sup> Vote-counting divides the piles of 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). Vote-counting does not take into account the sample’s sizes of individual studies; every study counts as much as every other. It takes into account neither the varying strengths of results across different studies, nor the size of the effect in the studies. Plainly, for these reasons it is not statistically powerful.

<sup>16</sup> Statistical Significance is a complex concept, but in short, two types of probability are generally associated with the reporting of significance levels in inferential statistics. One refers to the a priori probability you have selected as an acceptable level of falsely rejecting a given null hypothesis. This probability, called the “alpha level” (or significance level), is the probability of a Type I error in hypothesis

to what extent it is possible to draw inferences or generalizations from the studies. It does not, in itself, suggest that the results were *large*. They may merely have been shown to be *predictable* for the reference group.

### 4.2.3 Lipman and Null Hypothesis Testing

In the case of the evaluation by Lipman himself, the problem is different. Lipman used one of the most common tests of statistical significance: the Null Hypothesis Significance Testing (NHST)<sup>17</sup>. “Significant gains”, in NHST, means that at a level of significance of .01 (which is to say, with only a 1% probability of error), the likelihood of obtaining observations as extreme as the ones actually observed in Lipman’s experiment would have been very small, if the Null Hypothesis is true. This result (in 1970) led Lipman to conclude that there is a *cause-effect relationship* between Philosophy for Children and the development of reasoning skills. However, this result is questionable, given the severe criticisms that have been levelled at the NHST itself.

There are two main categories of criticism of NHST. Firstly, that, in its very design, NHST contains logical flaws. Secondly, that it is all too commonly misunderstood or misapplied. Jacob Cohen, the great authority on statistics and probability, in his well-known article “The Earth is Round ( $p < .05$ )” argued that the way in which NHST makes hypothesis testing probabilistic, renders its findings invalid (Cohen, 1994). He warned that researchers frequently misinterpret NHST results. For example, a failure to confute a null hypothesis<sup>7</sup> about the difference between A and B, is sometimes interpreted as indicating that there is no difference. In fact, he pointed out, all it shows is that the direction of the difference is “uncertain”<sup>18</sup> Cohen went on to argue,

---

testing (rejecting the null hypothesis when it is true) and is commonly set at .05 or .01. The other kind of probability, the p value, refers to the *a posteriori* likelihood of obtaining a result that is as extreme as or more extreme than the observed value you obtained, assuming that the null hypothesis is true. Persons unfamiliar with statistical inference might find it more intuitive to interpret the alpha level ( $p < .05$ ) as meaning that the probability that the data should turn out as they did merely by chance are less than 5 in 100.

<sup>17</sup> The hypothesis testing procedure requires the specification of two mutually exclusive hypotheses which the experimenter must, on the basis of the data, choose between. These hypotheses are called: The null hypothesis (the statistical hypothesis stating that there is no effect) and the alternative hypothesis (the research hypothesis indicating some effect between the variables).

<sup>18</sup> It is beyond the aim of this study to present all the criticisms with regard to the NHST; However, two sources of information can help the reader to grasp a greater understanding of them: Gerd Gigerenzer’s article entitled: “The Superego, the Ego, and the Id in Statistical Reasoning” (Gigerenzer, 1993), and the doctoral dissertation: “From Statistical Significance to Effect Estimation: Statistical Reform in Psychology, Medicine and Ecology” by Fiona Fidler (Fidler, 2005).

moreover, that, even if NHST is used and interpreted *correctly*, the results it provides still do not tell us very much.

The problems of “misuse” or “misinterpretation” of NHST include both the quite common error of thinking that obtaining a “significant” result means that there is a cause-effect relationship between the variables involved; and the problem Cohen pointed out: the idea that *no* conclusions can be drawn from a non-significant result. The problem in *Lipman’s* use of the NHST is that, while he appears to have obtained statistically significant results, he drew from them the questionable inference that he had discovered a cause-effect relationship. It is not clear, however, that such a relationship exists.

The above deficiencies in the NHST have prompted thoughtful statisticians to recommend reforms to statistical methodology such that experimenters routinely report effect sizes in the form of confidence intervals, and mention power analysis<sup>19</sup>. Neither Lipman’s experiment nor the reports of the Montclair list make any reference to these statistical disciplines. Consequently, these sets of results are, in themselves, inconclusive

These flaws in the methodological design and presentation of the individual studies in the evaluation of P4C render difficult any close analysis of the variables involved. In this case, the variables involved are the the study of philosophy and its impact on the development of cognitive skills. The fact that two meta-analyses could be conducted indicates that there are at least enough good studies to make possible a reasonable synthesis of the apparent results of the P4C program. However, there is still a good deal of work to do in this regard. We need more and better data and, in particular, the deficiencies mentioned above in the quality of all too many studies should serve as a point of departure for the better design of research methodologies and the presentation of results.

### 4.3 Evidence from Undergraduate Students

The evaluation of the impact of philosophy on thinking skills at this educational level has been less extensive and its results are even more imprecise than those for the P4C program. Given how imprecise and inconsistent this body of evidence it is, it does not permit us to conclude that the study of philosophy as such enhances CTS during college. On the face of it,

---

<sup>19</sup> Effect size reflects the magnitude of an effect or the strength of a relationship. Power is the probability that the test of significance rejects the null hypothesis when the alternative hypothesis (some effect between the variables) is true. High power, for instance, makes it likely that you are making the correct decision: accepting the alternative hypothesis (some effect) when it is actually true. This indicates that if an investigation reports some effect with high power, the effect is statistically significant. There are three major factors involved in the study design that influence the magnitude of the power: effect size in the population, level of significance, and number of observations (D. Moore, McCabe, G., 2003).

of course, this is a startling discovery. We shall examine it more thoroughly in the meta-analysis for precisely this reason. The presentation of this evidence takes place in two sections: (1) Studies Bearing Directly on the Topic, (2) Studies Bearing Indirectly on the Topic.

### **4.3.1 Studies Bearing Directly on the Topic**

I have been able to discover only a small number of studies (five) which have sought to directly measure the relationship between the study of philosophy by undergraduates and the development of CTS. Even within this group, there were distinct variations in research design. What sets them apart, however, is that, albeit in different ways, they were all concerned to measure the link between studying philosophy at college and improving CTS. In all (Annis & Annis, 1979; Harrell, 2004; Reiter, 1994; Ross & Semb, 1981) but one (Facione, 1990) of these cases, the philosophy students belonged to the experimental group. This distinction need not cause any problem for the inclusion of the Facione study in the set as a whole. However, it's slightly different focus does give rise to a number of observations, which will be discussed in the conclusion to this section.

Despite their common aim, the five studies in this set of investigations actually used different design methodologies and asked different questions. However, all the CT tests used in these studies, followed standardized, objective CT testing designs. All five studies were pre-test/post-test designs. All, finally, were conducted within a single college semester.

#### **4.3.1.1 Divergent Findings on Under-Graduates**

The findings from these five studies were divergent and it is difficult, therefore, to draw from them firm conclusions about the relationship between philosophy and CTS. In addition, the studies do not report enough statistical information as to guarantee that their research designs have followed the appropriate methodological rigour. With the exception of Reiter who reported low power in her study, the others, failed to report this information. With this, it is difficult to determine the validity of the findings and therefore the magnitude of the effect that the study of the discipline of philosophy could have on the development of CTS. Nevertheless, it is necessary to present here the results obtained in these investigations.

Annis & Annis 1979, comparing students in four courses (Introduction to Philosophy; Ethics; Logic and Psychology) suggested that which course a student took had no significant effect on the total score on the Watson-Glaser. However 'in general the results indicate that Logic has a consistent impact on certain aspects of critical thinking'(p.150). The students in Logic performed significantly better ( $p < .05$ ) in deduction, inference, and interpretation than students from the other philosophy classes. Thus, it would appear that the study of Logic tends to have a positive effect on the ability of students to do basic reasoning. However, there were two areas in the CTS tests that even the students of Logic did not show significant improvement in: the identification of assumptions and the evaluation of arguments.

Ross & Semb 1981, generated results that were different from those of Annis & Annis. In this case, the students in Introduction to Philosophy (following an educational instruction based on the Keller Plan)<sup>20</sup> gained significantly better ( $p < .01$ ) on the Watson Glaser test than did the control group of non-philosophy students. Thus, the study suggested that “philosophy can have a high degree of effectiveness in improving critical thinking ability” (p.114). The same applies in the case of Harrell 2004. Showing students how to use argument mapping, in Introductory Philosophy, Harrell found that philosophy students using this innovative teaching approach improved CTS ( $p < .01$ ).

Reiter (1994) applied a different innovative teaching technique, called ‘dialogical instruction’, to college students in Philosophy, Ecology and Children’s Literature classes. She then put all of them through a general reasoning test and found that the Ecology and Children’s Literature students had made greater gains in reasoning skills than had the philosophy students. This particular finding would appear to suggest that there is *no clear direct relationship* between the study of philosophy and the development of better critical thinking skills.

Facione (1990) found that, at least as measured by the standard CCTST test, the philosophy students made *no apparent gain* in critical thinking skills ( $p < .05$ ). He phrased his finding delicately, observing that “This suggests that whatever growth in CT skills may have occurred in Introduction to Philosophy, it was not measurable on the CCTST” (p.16). The philosophy students had been selected, however, based on the assumption that they would exhibit improvements in CT skills, given the formal instruction they had received in how to think well.

In summary, the results of the five studies of direct effects on College students are divergent and, partly for that reason, inconclusive. What compounds this problem is that the studies are methodologically and substantively heterogenous. Their contradictory findings also come based on too little statistical information. Nonetheless, four very interesting, if tentative, conclusions may be drawn from these studies, taken as a set.

The overall evidence presents no compelling case that the study of philosophy as such enhances CTS during college. Where an improvement occurred which was correlated with the study of philosophy, it remains difficult to discriminate among sub-disciplines or teaching methods as to what, precisely, generated the improvements. The evidence also presents no compelling case that the study of philosophy is more likely to lead to improvements in CTS than the study of any other serious discipline.

---

<sup>20</sup> The Keller plan is also called the personalized system of instruction (PSI). The primary features of the system are: “study guides with questions to be answered in writing, occasional lectures for enrichment and motivation, more frequent quizzes which can be repeated until mastered, and proctors to provide immediate feedback on quizzes”. (Ross & Semb, 1981, p.118)

The specific study by Ross & Semb 1981 raised a suggestion that is surely worthy of closer and more serious analysis. These authors indicated that the gains in CTS due to studies in philosophy appeared to depend on the method of instruction. The implication here is that *traditional* philosophy (lectures, discussions) does not generate the high degree of effectiveness in improving CT abilities that are achievable with more innovative methods of teaching.

This fourth point needs amplification. Ross and Semb set up a second experiment, in which the experimental group (following the Keller Plan) were encouraged to concentrate on informal argumentation and the comprehension of written texts with personalized instruction (PSI), while the control group were taught by traditional means of lectures and discussion. They found that the experimental group made greater gains in CTS than did the control group.

Harrell, in 2004, established a similar correlation using argument mapping in the experimental group. Her conclusion was: "While, on average, all of the students in each of the sections improved their abilities on these tasks over the course of the semester, the most dramatic improvements were made by the students who learned how to construct arguments." (p. 14). Reiter, also, introduced an innovative technique - dialogical instruction. However, because her intention was not to compare a traditional approach to philosophy with a more innovative one, but to test the utility of the dialogical method, something independent of philosophy altogether, it is not possible to draw clear conclusions from her findings.

In addition, the considerations made by Annis & Annis 1979, and Facione 1990, also seem to support the suggestion that the correlation between the approach to teaching basic philosophy and the results achieved is what merits more attention. In these studies, all sampled students enrolled in traditional philosophy courses and the findings of the studies were inconclusive as to the efficacy of these courses in inducing improvements in CTS.

This fourth point is further underscored by the fact that consistently positive results in CT gain have been provided by critical thinking courses offered by Western Philosophy Departments (Butchard, 2006; Donohue et al., 2002; Hitchcock, 2003; Rainbolt & Rieber, 1997; Spurrett, 2005; Twardy, 2004). These courses are exceptional, in that they are not standard philosophy courses. They are focused on teaching critical thinking skills in their own right. Their positive results reinforce the idea that innovative methods of teaching within philosophy departments might produce better results in the teaching of critical thinking skills. However, we cannot infer from that that philosophy *per se* improves critical thinking skills.

### **4.3.2 Studies Bearing Indirectly on the Topic**

The studies included in this section did not consider in any direct manner the key variables that guide this research project: the study of philosophy as the independent variable; and the change in CTS as the dependent variable. On the contrary, these studies were concerned with

different research questions; but their findings yielded information that happens to be relevant to our project.

In this section, the studies in question were concerned with the teaching of Logic as such, rather than with the teaching of a broader philosophy curriculum. They were interested in the impact of such teaching on the development of either or both of reasoning and CTS. They are of at least indirect relevance to the present inquiry, because the teaching of abstract principles of reasoning through logic courses is, clearly, a variation or sub-component of the teaching of philosophy. Moreover, their common central interest was in establishing the extent to which the teaching of abstract principles helped students to improve their reasoning or CTS. Their significance, paradoxically, lies in the fact that their findings do *not* allow us to draw any definitive conclusions regarding the utility of instruction in Logic for the development of CTS. This, of course, places them at the heart of the ongoing debate about the issue.<sup>21</sup>

In 1989, Ernest Pascarella conducted the first longitudinal study to compare the critical thinking development of college students with individuals of the same age and social cohorts who had *not* gone on to college – whom he referred to as his ‘non-college’ control group (Pascarella, 1989), p.20). The study extended over a single year, the first year of college and the testing means used was the Watson-Glaser Critical Thinking Appraisal (CTA). It was determined that the college students main gains in CTS were, on average, 17% greater than the gains made by the non-college group. Based on this study, Pascarella sought to explore various correlations between college experience and the Watson-Glaser scores of his experimental group. He looked at ten carefully selected variables<sup>22</sup> within the college experience of the group, of which the number of hours spent studying Logic and science was one.

What Pascarella found, interestingly, was that there was *no observable correlation* between any of the variables and the CTS scores. This is particularly noteworthy, from our point of view, with regard to the study of Logic and science. Pascarella compared the number of science and Logic hours the students took with their Watson-Glaser scores. They failed to have significant

---

<sup>21</sup> It is important to mention that with the exception of Solon’s study, who showed high power in his study, the rest did not mention this variable. Therefore, the contradictory findings in this section could be just the predictable result of low statistical power analysis.

<sup>22</sup> In addition to completing the Watson-Glaser, the students also completed a questionnaire that asked about ten specific college experience variables: 1) living on campus versus commuting to college; 2) average number of hours a week spent studying; 3) number of non-classroom discussions with faculty during the year; 4) number of non-classroom discussions with other students; 5) frequency with which the ‘editorial pages’ of a newspaper were read; 6) number of college-sponsored lectures and debates attended; 7) number of unassigned books read on intellectual, academic, or controversial topics; 8) number of science or logic courses taken during the freshman year; 9) number of university-sponsored extracurricular activities, and 10) college attended.

partial association with CTS performance<sup>23</sup> (Pascarella, 1989). What Pascarella deduced from his overall findings was that “critical thinking, at least as measured by the Watson-Glaser, is a broad dimension of student intellectual development that is unlikely to be substantially influenced by any one specific college experience or curricular emphasis.” (p.25). This was consistent with the earlier findings of McMillan, in 1987.

However, some studies seem to suggest that Formal Logic courses produce a positive effect in the development of CTS. The Panowitsch-Balkcum study published in Rest, J. 1979, reported such findings. This study sought to establish two things: what kind of philosophical education was most conducive to the development of moral judgment, a course specifically in Ethics, or a course in Formal Logic? They used the Defining Issues Test (DIT) to measure the development of moral judgment. But they wanted, at the same time, to validate the DIT itself as a means of measuring the development of moral judgment, by comparing the scores it produced to scores produced using the Cornell Critical Thinking Test (CCTT), which measures skills in logical rather than moral reasoning (Rest, 1979)

The pre- and post-testing results showed that the students studying Ethics improved on the DIT but not on the CCTT; whereas, those studying Formal Logic improved on the CCTT but not on the DIT. While Panowitsch and Balkcum were interested in the implications of these results concerning the development of moral judgment and the utility of the DIT, their findings also showed something else, which interests us: that a course in Formal Logic yields greater CT gains than a course in Ethics. In other words, the *kind* of philosophy course one studies makes a difference to the gain in CTS.

Tom Solon’s experiment (Solon, 2003) testing the hypothesis that an experimental group of Logic students (formal and informal) would improve their CTS more than either of two control groups, of psychology and rhetoric students, respectively, also yielded positive results. His findings showed that the experimental group significantly out-scored both control groups<sup>24</sup>.

Hatcher 1999, in a one year longitudinal study, measured the effectiveness of an integrated logic and writing course on CTS (Hatcher, 1999). Using the CCTST, Hatcher suggested that this combined course worked in improving CTS. In a subsequent study, in 2001, he argued that, while the study of philosophy did seem to prepare students better than did other disciplines for doing CT, real improvement in CT within the universities requires both the standardization of the idea of CT and its incorporation across the curriculum, rather than its isolation in brief specialist courses (Hatcher, 2001).

---

<sup>23</sup> Pascarella set the significance level for the partial correlation at  $p < .10$

<sup>24</sup> Note from Solon’s study: The psychology post mean was significantly higher than the rhetoric post mean at the .05 level. Also, the logic group significantly outscored psychology at the .05 level and rhetoric at the .01 level. (p.11). Solon also reported effect sizes and high power for the study.

There are, in addition, some studies which have sought to explore the relationship between Logic and the development of reasoning skills. In so far as reasoning and CT skills are not wholly co-terminous, these studies could not resolve the debate about the relationship between Logic and CTS, even were they definitive in themselves. As it happens, their findings, like those of a number of the studies already considered, are not conclusive even within their prescribed domain. Some of these studies seem to indicate that the study of Logic produces statistically significant gains in reasoning skills (Stenning, Cox, & Oberlander, 1995; Van der Pal & Eysink, 1999). Conversely, others indicate that the study of the abstract principles of logic produces no discernible improvement in reasoning (Cheng, Holyoak, Nisbett, & Oliver, 1986).

#### **4.4 Evidence from Graduate Students**

There appear to have been only three studies that have tried to estimate the growth in reasoning and CTS in students who have completed, or are close to having completed, an undergraduate major in philosophy. Two of these studies (Hoekema, 1987; Nieswiadomy, 1998) were centred on determining the level of competence in reasoning abilities of such students seeking to enter various graduate schools at university. Although these studies suggest that philosophy majors generally do better on graduate tests (GMAT, LSAT, GRE), they do not disentangle the contribution of philosophy from the students' selection effect. People that choose to study philosophy may, in general, be good at reasoning anyway and the evidence these two studies provide does not indicate to what extent the study of philosophy in itself *improves* reasoning skills.

As Harrell (2004) herself said about these tests: "We can cite statistics about which majors generally do better on the LSAT and GRE; but what we have not been able to do in the past is show evidence that our [philosophy] classes improve critical thinking skills" (p.15). The third study (Kuhn, 1991) was an attempt to determine the informal reasoning abilities of subjects who were classified as reasoning experts – post-graduate philosophy students – compared with a wide variety of other people. Although this pool of three studies is very small, it is still worth noting and analysing its findings.

Hoekema (1987), called his study: 'Why not major in something practical like Philosophy?' His aim was to determine how philosophy students perform on graduate admission tests. The evidence was based on the test performance of the students in the following: GRE (Graduate Records Examination), GMAT (Graduate Management Admissions Test) and the LSAT (Law School Admissions Test). In the section of the GRE on verbal reasoning, students of philosophy surpassed all other humanities majors. Only English majors came close. Not only this, but they also surpassed all those majoring in other fields, such as social sciences, natural sciences, business, engineering, and computer science.

In the LSAT and GMAT, philosophy majors performed substantially better than majors in any other humanities field, better than all the social science majors except economics, better than all natural science majors except mathematics, and better than all business and applied fields, including engineering. The findings of these two tests seem to suggest that students of philosophy are more able than most others to transfer their skills in verbal reasoning to domains other than philosophy.

Nieswiadom, by comparison, undertook a study with the intention of determining the reasoning abilities of economics students and the level of preparation in reasoning skills necessary to the pursuit of a career in the law. His findings, however, confirmed those of Hoekema. The data for the study came from applicants that took the LSAT in two academic years, 1991-92 and 1994-95. The sample populations numbered 1,547 and 1,884 students, respectively. Nieswiadom concluded that philosophy majors placed second, behind only physics/math majors, in a group of 29 disciplines for both years in the improvements observed in the development of CTS.

Deanna Kuhn (1991), in her often-cited book, *The Skills of Argument*, presented evidence regarding the abilities of philosophy students. The problem with Kuhn's data is that her sample of philosophers was tiny – five in total – and all of them were advanced PhD students at a “highly reputed” American university, which may just show selection bias. Admittedly, this tiny sample did exhibit excellent general thinking skills, which distinguished them from the other subjects of Kuhn's study. However, this does not by itself show that training in philosophy was responsible for the philosophers' good performance. The five philosophy graduates were, after all, the products of an elite education system, whose skills may well have been developed as a result of multiple influences over many years. That said, Kuhn's central argument is a highly important one and well worth more systematic testing.

The primary objective of her study was to examine how people reason about real and complex issues of general interest. Subjects from adolescence to late adulthood were asked to describe their opinions about social problems, such as crime and unemployment. Her study indicated that, in general, people have difficulty in supplying any evidence in support of their opinions or theories, in thinking about alternative theories or counterarguments and in evaluating arguments. However, the thinking of the philosophers selected for the sample, as experts in reasoning itself, is worthy of attention. They showed a *perfect* performance in generation of genuine evidence, alternatives theories, counterarguments, and rebuttals.

By reading extracts from Kuhn's interviews with the philosophers, one comes to appreciate the critical thinking skills and attitudes of mind that they bring to the process of argument; in particular, the ways in which they clarify concepts before making a judgment, and the manner in which they present and evaluate evidence. Kuhn's findings have implications that extend beyond the quality of philosophical expertise itself to the broader question of the relation

between the content of philosophical education and the general application of reasoning skills across domains. In this regard the author indicated: "The performance of the philosophers shows that it is possible to attain expertise in the reasoning process itself, independent of any particular content to which this reasoning is applied". (p. 262)

## 4.5 General Conclusions

Five core conclusions arise from the literature review:

- There is a serious lack of sound empirical data regarding the impact of studying philosophy on the development of CTS.
- With regard to P4C, there is a clear need for better standards of methodological design and reporting in the gathering of such empirical evidence.
- Nonetheless, two meta-analyses plainly suggest that there P4C has had a positive effect on the development of general reasoning skills in children.
- With regard to under-graduate and graduate students, the lack of evidence is more serious and needs to be supplemented.
- What evidence exists as regards under-graduate and graduate students is contradictory and inconclusive, so that we cannot make a compelling case either way as to the impact of philosophy on the development of CTS.

In the light of these general findings from a review of the literature, it is, surely, clear that we need more studies of the relationship between philosophy and the development of critical thinking skills, at both the undergraduate and graduate levels. There is a clear requirement, also, that such future studies be designed in such a way as to elicit and make explicit the kinds of information vital to rigorous assessment of the relationship between philosophy and CTS. Finally, while such further studies are highly desirable, the *existing* studies – or at least those that are of any substance – need to be subjected to a meta-analysis, in order to determine, as far as possible, how compelling a case they make as things stand. It is this last task that the present thesis attempts.

## 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.<sup>25</sup>

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

---

<sup>25</sup> 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.<sup>26</sup> 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

---

<sup>26</sup> 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:<sup>27</sup>

- 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 design to measure this construct (e.g. California Critical Thinking Skills Test, Watson-Glasler 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)

---

<sup>27</sup> 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.<sup>28</sup> (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

---

<sup>28</sup> 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 findings<sup>29</sup>. 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

---

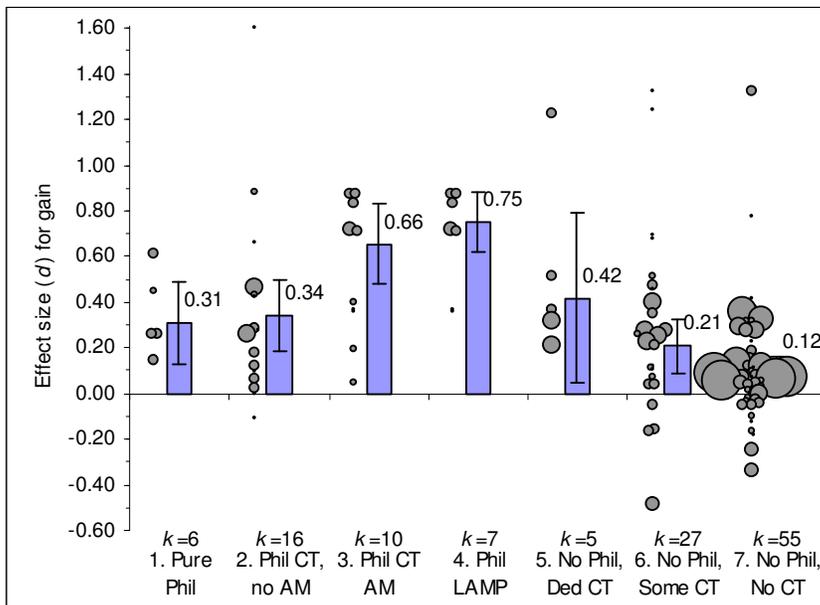
<sup>29</sup> For purposes of meta-analysis, a single *research finding* is a statistical representation of one empirical relationship involving the variables of interest to the meta-analyst, measured *on a single subject sample*. For instance, for those studies that used an experimental-control group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

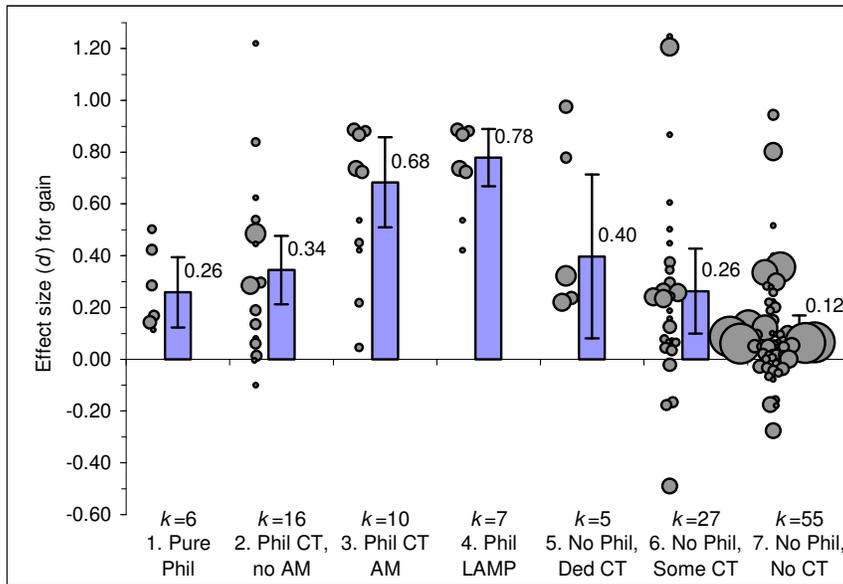
The information in the fifty-two studies was coded, following the coding protocol indicated in section 5.3, A Meta-Analysis of the Field. Table 2, in Appendix A, shows the Master List of studies used in this meta-analysis. This Table sets out the study characteristics, course information, and research 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].

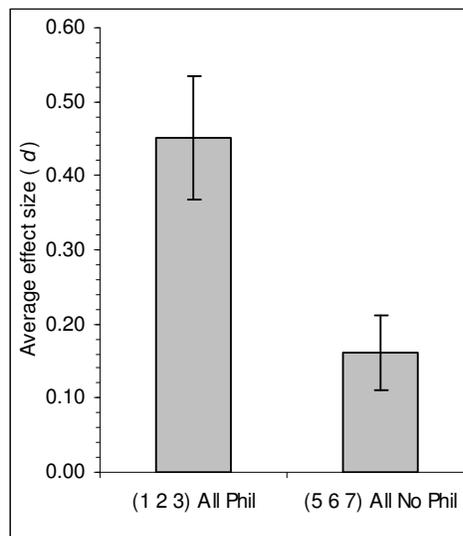
### 5.4.4 Relevant Comparisons:

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  $p < .01$ .

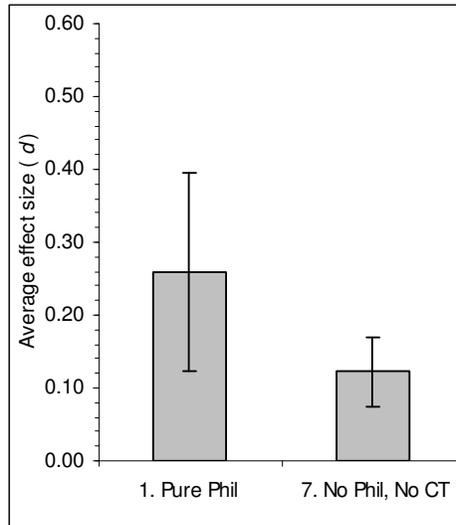


**Figure 6. CT gain in philosophy vs CT gain in Non-philosophy courses.**

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.

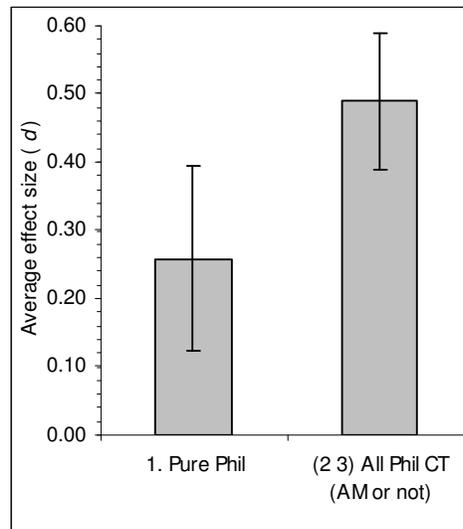


**Figure 7. CT gain in Anglo-American analytic philosophy vs No Phi No CT**

#### 5.4.4.2 Is philosophy better than CT courses?

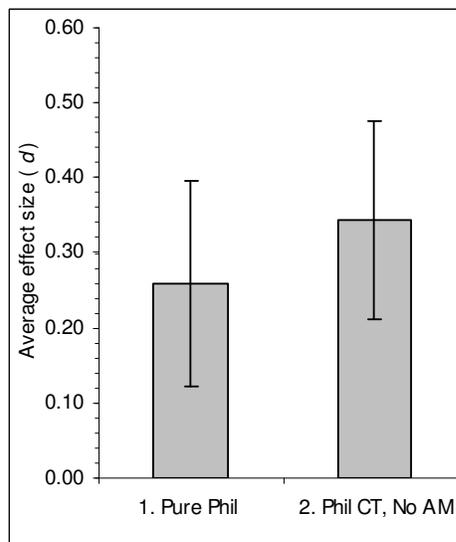
In this comparison, we are concerned with whether pure philosophy instruction (Group 1) makes more difference than CT instruction in its own right. There are various sub-comparisons to make here.

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



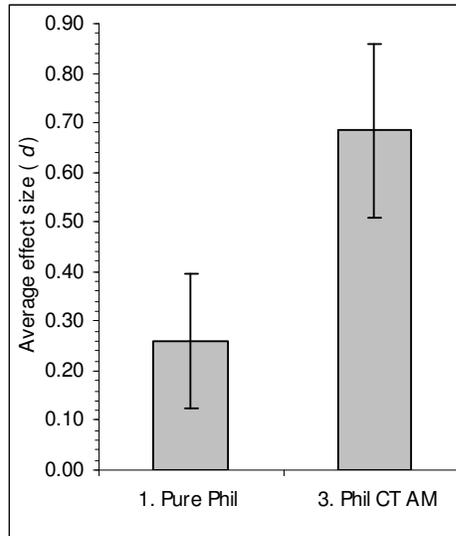
**Figure 8. Anglo-American analytic philosophy vs all CT instruction in philosophy.**

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  $< .05$



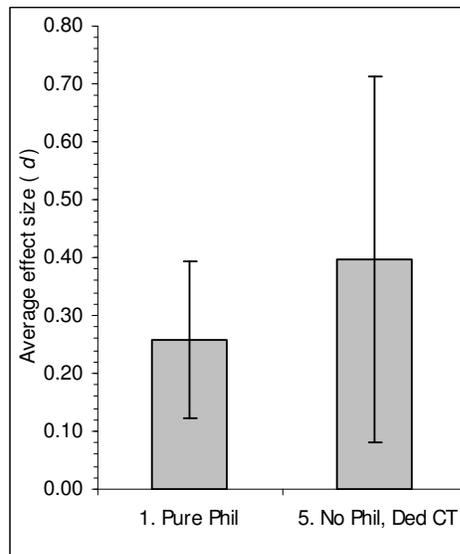
**Figure 9. Anglo-American analytic philosophy vs Traditional CT in philosophy.**

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  $< .01$



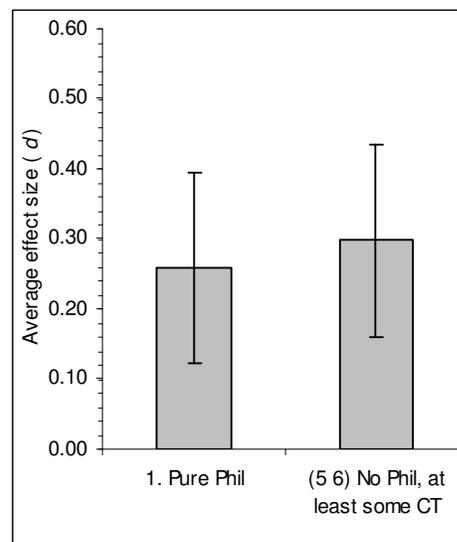
**Figure 10. Anglo-American analytic philosophy vs CT instruction in philosophy with argument mapping.**

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  $< .05$



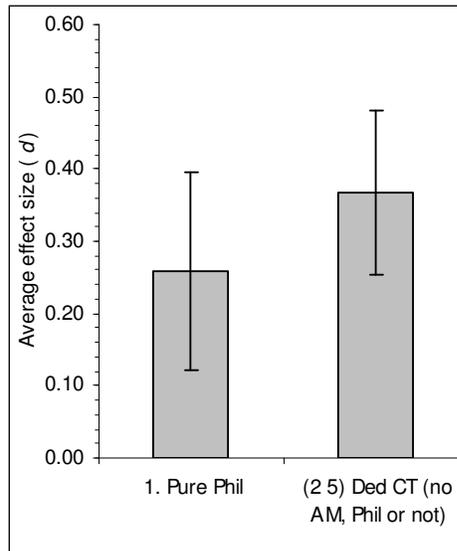
**Figure 11. Anglo-American analytic philosophy (Group 1) vs Traditional CT instruction outside philosophy.**

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  $p < .05$



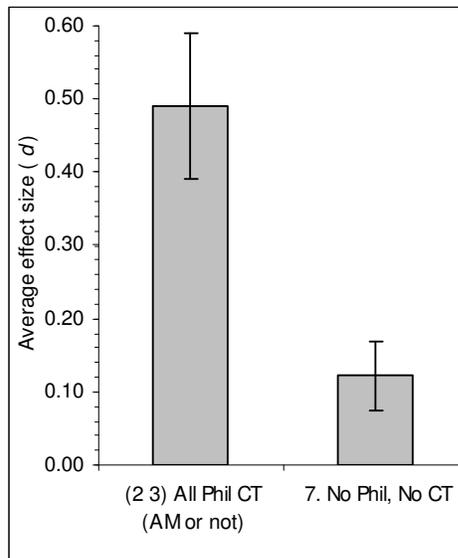
**Figure 12. Anglo-American analytic philosophy vs all CT instruction in outside philosophy.**

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  $p < .05$ .



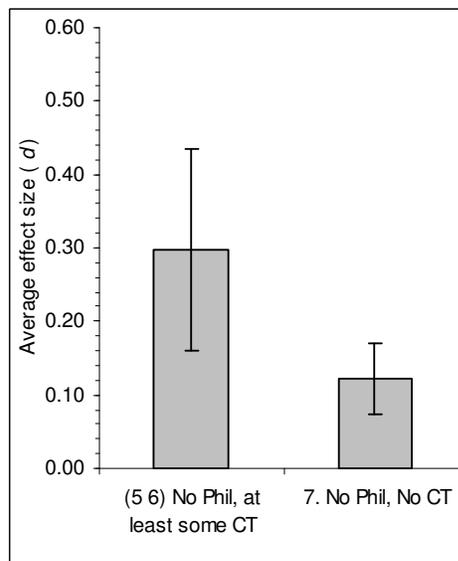
**Figure 13. Anglo-American analytic philosophy vs Traditional CT in philosophy and outside philosophy.**

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.**

## 6 General Discussion

At the beginning of this thesis, we posed three questions:

- 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?

In order to answer any one of these three questions, we needed to be able to answer two subsidiary questions: (1) To what extent is the gain, in any given case, directly attributable to the discipline in question, whether it be philosophy, some other course of study or a critical thinking skills course? (2) How effective is the given course, compared with other possible ways of teaching CTS, in improving CTS?

These subsidiary questions are important, because the answers to the three main questions converge on a common point, which is an effort to determine to what extent the study of Anglo-American analytic philosophy is an *effective way* – to say nothing of being the *most effective way* - to improve CTS. In the course of the discussion that follows, we shall be answering each of the questions separately, but at some points the answers will overlap, as each is placed in the context of the general case being made.

### 6.1 Question one: Does (Anglo-American analytic) philosophy improve critical thinking skills?

Our first question has to do with whether the study of philosophy improves critical thinking skills at all, without immediate regard to whether it does so more effectively than any other discipline. As we saw in the results section (see 5.4 Results of the Meta-Analysis) the mean effect size of the pool of studies from this category yielded a value of 0.26 SD; CI [0.12, 0.39]. This suggests that the study of Anglo-American analytic philosophy *does* actually improve CTS *over one semester for undergraduate students*. However, it is necessary to determine, firstly, to what extent those gains are attributable to the study of *philosophy*, as such, rather than to students simply having undertaken a course of study at all, whether or not it was philosophy?

Secondly, it is necessary to ascertain to what extent the gains attributed to philosophy make the discipline of philosophy, as it is traditionally taught, an *effective* way to improve CT skills?

It is difficult, with the data we have, to determine how much of the gain of .26 SD is attributable to philosophy, chiefly because there are several variables which, plainly, could have affected this result. First of all, the data for philosophy do not, in themselves, enable us to determine whether the gain is due to philosophy, or would have been made by students, in the normal course of events, spending a semester in any serious discipline. We need, therefore, to compare the philosophy data with that from other disciplines. We shall, of course, be looking at this comparison in addressing the second of our main questions.

Two other variables must be taken into consideration in examining the extent to which the gain of 0.26 of a SD is attributable to the study of philosophy. These are the considerations that both the method of teaching philosophy and the substantive content of philosophy courses can vary significantly, with measurable implications for the gains in CTS. It turns out, for example, that students taught philosophy according to the Keller Plan made far greater gains in CTS than those taught by traditional methods, at least according to what limited data are available. And, again, the study of Logic – a traditional philosophy subject, but one which could as readily be taught within a Science or Mathematics or Economics major – generated greater improvements in CTS than other more specifically Philosophical subjects. There is, also, the caveat that philosophy students may turn out to be a self-selecting group of students who are predisposed to develop their critical thinking skills and might have done so just as well, or even better, in other courses of study than philosophy.

The self-selection of students is an interesting variable, which would repay closer examination. It is easy to infer from the fact that students of philosophy do relatively well in critical thinking tests that they do so *because* they studied philosophy. Yet, on the face of it, this is an elementary fallacy of reasoning, *post hoc ergo propter hoc*. How do we know that it is not, rather, a case of students choosing to *study* philosophy because they are good at critical thinking? The fact that the students are pre-tested and post-tested and can be shown to have improved can all too easily beguile us into believing that it *must* have been the philosophy that made the difference.

If, however, the students began the course of study in philosophy with a predisposition to develop their critical thinking skills, then the philosophy may merely have provided an occasion for *demonstrating* that improvement, without necessarily *causing* it. Indeed, as we shall see, much greater improvements are possible than have been observed in the traditional study of philosophy. It is even conceivable that philosophy, far from causing an improvement in critical thinking skills, may actually hold back the natural development of such skills in the better students.

With regard to both the method of teaching philosophy and the substantive content of philosophy courses, it is necessary to draw attention to two studies which produced effect sizes considerably larger than the rest of the philosophy studies in the pool utilized in the meta-analysis. These two studies are the Introduction to Philosophy used by Ross & Semb, 1981, in which the Keller Plan, a special approach to education, was used with notable results; and the course in Logic reported by Rest, 1979. The Keller Plan and Logic courses yielded results so much better than the other Introduction to Philosophy courses that they inflated the results for this pool of studies. In fact, if these two are separated out, the difference made to CT skills by the Introduction to Philosophy courses becomes appreciably smaller ( $ES = 0.19 SD$ ). What is more important, however, is the intriguing implication that *particular approaches to teaching* made a more substantial difference than the subject matter in itself. As we have seen, this turns out to be true in regard to other approaches, also. (See section 4.3, "Evidence from Undergraduate Students", in the Review of the Existing Evidence)

Both courses merit attention, because in both cases there are pedagogical elements which could go far toward explaining the differences between the effect sizes. On the one hand, while Logic is a sub-component of philosophy, it is one in which there is a highly specific commitment to imparting the abstract principles of reasoning. The comparative finding in our study, that Logic has a greater impact on the development of CTS than other introductory courses in the study of philosophy should not, therefore, occasion any particular surprise. A similar finding by Annis & Annis, in 1979, which was not included in our meta-analysis, because it did not provide sufficient statistical data, nonetheless points in this same direction. They found that a course in Logic, compared to courses such as Introduction to Philosophy, Ethics or Psychology yielded better results in some aspects of Critical Thinking. Although these results are not definitive, they surely merit further research.

On the other hand, the Keller Plan technique, which fairly sound research suggests is a highly effective educational tool, could by itself account for much of the effect size. The results of the Michigan Meta-Analysis of Research on Instructional Technology in Higher Education (Wittrock, 1986) found a "moderate" (medium) effect size (.49) for students in Keller Plan courses. Although this effect size was described as 'moderate', it was found to be greater than those for other teaching strategies, such as Audio-Tutorials (.20), Computer based-teaching (.25), Program Instruction (.28) and Visual-Based Instruction (.15). What all this suggests is that, while studying philosophy will make some difference, a greater difference will be made if one does either of two things: concentrate on Logic itself or keep the broader content, but change the way you teach it. These are, surely, both interesting, if provisional, findings.

Apart from these questions of causal attribution, it must be noted that this growth (0.26 of a SD) happened during the first year of a College course. The majority of the studies, with the exception of Rest 1979, used freshmen students. Consequently, we don't know what is likely to be the overall impact on the development of CT skills of a full philosophy degree. Does CT

continue to grow each semester at the rate shown for a single semester in freshmen? Or is this a one off gain, followed by a flattening out of subsequent development? This could be a fruitful theme for future investigation. Such an investigation would seek to determine if the major gains in CT skills by philosophy students (or other College students) occur in the first year, or there is a different and more interesting pattern of development? It has in fact been suggested by a number of specialists (Pascarella and Terenzini, 2005, Donohue et al., 2002) that the major gains do, indeed, occur in the first year of general university education. This proposition should, surely, be rigorously tested.

## **6.2 Question Two: Does (Anglo-American analytic) philosophy improve critical thinking skills over and above university education in general?**

The discussion about the net effect of the study of philosophy on the development of CTS also leads us to ask ourselves: To what extent would those philosophy students have made the gains they did had they *not* been studying Anglo-American analytic philosophy? Or to put it another way, is the gain due to *philosophy*, or would it have been made by students in the normal course of events, spending a semester in *any* serious discipline? Whence the question, Does (Anglo-American analytic) philosophy improve critical thinking skills over and above university education in general? In order to answer this question, we must first establish what difference a university education in general actually makes.

We have data for this, but the data on its own is insufficient. What our data shows is that a university education, in general, produces a gain of 0.12 of an SD over any given semester. Superficially, this compares unfavourably with the gain of 0.26 SD in a semester for philosophy. We have noted several caveats with respect to the results for philosophy. We need to ask here what the 0.12 SD for university education in general actually means. As it happens, those not attending university at all appear to improve their CTS by 0.10 SD in the equivalent of the first semester after leaving school (Pascarella, 1989). This would suggest that attending university, at least initially, *makes no appreciable difference*, because CTS improve at that age anyway by much the same amount.

We need, however, to be a little cautious in drawing conclusions here, since Pascarella's own studies suggested an improvement of 0.26 SD in the first semester of university – the improvement our own data from the meta-analysis suggest is achieved by students in philosophy. What does all this mean? In fact, statistically speaking, as explained briefly above (5.4 Results section) once you allow for confidence intervals, there is not much to choose

between gains of 0.1, 0.12 and 0.26. In short, whether we use Pascarella's data alone or the more systematic data deriving from the meta-analysis, it appears that there is little to choose between not going to university, going to university in general and studying (pure) Philosophy at university, as regards improvements in CTS in a given semester period. This is both a counter-intuitive and even a disconcerting conclusion.

What bearing do these considerations have on the answer to our second question? There are, in fact, quite a number of implications. To begin with, the difference attributable to the study of philosophy looks like 0.26 of an SD minus 0.12 of an SD, given that the latter is the gain that students would be expected to make simply by being at university. This cuts the gain to be attributed to philosophy down to 0.14 of an SD. Moreover, this is without taking into consideration the significance of the better results for the Keller Plan and Logic than for philosophy as such. If you allow for this consideration, the gain attributable to philosophy is reduced further, to only 0.07 of an SD.

Superficially, this indicates, at least, that the gains attributable to philosophy are marginally greater than those attributable to university education as such. However, the reality is not quite as simple as it appears. As we saw in the results section, the difference between 0.26 of an SD (or 0.19 of an SD, considering the effects of the Keller Plan and Logic) and 0.12 of an SD, although the first appears to be twice the second, is actually not statistically significant ( $p < .05$ ). This, statistically speaking, is because the confidence intervals overlap (see Figure 7, Meta-Analysis Results). Where the confidence intervals for measurements overlap in this way we cannot with confidence differentiate between them. There is, therefore, less reason than might appear to be the case for believing that a student trying to gain greater CTS will actually do better by studying philosophy than by studying some other subject.

There are two further things to note with regard to our second question. These two things have implications for further research. First, a factor that interferes in the comparison between philosophy and other subjects is that no-one has made a direct comparison between (Anglo-American analytic) Philosophy and some other specific subject, only a broad cluster of them (Group 7). We simply do not know whether the effects of a course such as, for example, law, mathematics or engineering, taken on its own and compared with philosophy, would present a different picture. It was not possible to establish such direct comparisons in the present study, because there is insufficient research data on which to base them. The one possible exception might be nursing, which is over-represented in the data.

The second thing to note has to do with the educational level of the students in the study samples. In Group 1 (Anglo-American analytic philosophy), for instance, the majority of students were freshmen, while in Group 7 (No-Philosophy, No CT) the sample was heterogenous. This matter warrants further attention, since it is possible that the most dramatic gains in CTS are normally made, one way or another, in the first year of university education. This would imply, if

it was the case that the relatively lower gains in subsequent years brought down the average in Group 7. Consequently, there needs to be some investigation as to whether, in fact, the first year of university of study does yield better results than the later years. (Pascarella & Terenzini, 2005, Donohue et al., 2002).

In short, the answer to our second question is that, based on the available research data, the study of philosophy cannot be said to improve critical thinking skills over and above university education in general. While further research might yet show that it does, we are not entitled to presume this without such research being conducted and evaluated..

What is still missing from these conclusions, however, is the comparison with actual CT courses, which, in the nature of the case, constitute an interesting and important element in the equation. The discussion that follows, then, centres on both this element in the picture and the third of our research questions: Do critical thinking courses make a difference to critical thinking skills, whether or not such courses take place within the discipline of philosophy? The comparison between philosophy courses and CT courses, as it happens, yields some interesting findings.

### **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?.**

The discussion of the effect of the study of philosophy on the development of CT skills, leads us also to examine the effectiveness, in this regard, of university courses specifically designed to develop critical thinking skills. There are two underlying questions here. First, is philosophy better at teaching CT skills than are CT courses themselves? Second, are CT skills taught by philosophers, or in philosophy Departments, better at improving CT than CT courses taught elsewhere, or by others? There has been some debate as to whether philosophy Departments or philosophers are better at teaching critical thinking skills than non-philosophers (Hatcher, 2001). Our meta-analysis gets us at least part of the way to a useful answer to these questions. In the discussion that follows, we shall address both of them.

Two conclusions can readily be drawn from analysis of the data (see Figure 5). First, philosophy Departments, in general, seem to be more effective at teaching CT skills than Non-Philosophy Departments. Second, although this could appear to contradict the first claim, the study of Anglo-American analytic philosophy *per se* is not more effective than CT courses, when it comes to improving CT skills. There is not, however, a contradiction, as it happens; for it is the teaching of CT within philosophy Departments, compared with the teaching of CT elsewhere

with which we are concerned. The distinction will be clear so long as we do not conflate the teaching of CT in philosophy Departments with the teaching of Anglo-American analytic philosophy. Let us, then, explain both conclusions with reference to the evidence from the data.

As can be seen in Fig 5 (see 5.4 Meta-Analysis Results) there is no statistically significant difference between Group 1 (Anglo-American analytic philosophy) and Groups 2, 5 and 6 (traditional CT courses whether in or outside of philosophy Departments). This is because, in all these cases, there are substantial overlaps between the confidence intervals for the various groups. In other words, always assuming that we are talking of a single semester of university studies, there is no basis for concluding that philosophy is a more effective way to teach CT than are traditional CT courses. Conversely, however, there is, equally, no sound basis for preferring traditional CT to philosophy.

However, as Fig. 5 also shows very clearly, there is a striking corollary to this judgment. There is a clear statistical difference between Group 1 and Groups 3 (Phil CT with some argument mapping) and 4 (Phil CT with lots of argument mapping practice), which is to say, between Anglo-American analytic philosophy without CT and the teaching of CT within philosophy Departments using argument mapping. Here, the comparison is clearly unfavourable to Anglo-American analytic philosophy. The results show, in this instance, a statistical significance with a p-value at a level of .01. This merits special attention.

The argument mapping approach has distinctive features that are necessary to mention: students practice argumentation skills, involving the use of a software programme (Rationale) that permits the visual representation of arguments. Such representation is encouraged on the basis that the brain can economize on cognitive effort where much of what it seeks to comprehend is visually cued. Use of a good deal of practice in argument mapping is based on the hypothesis that reasoning is a skill and that, as with any skills, 'deliberate practice' (Van Gelder, Bissett, & Cumming, 2004) is crucial to the development of the skill. The effect size of the Group 'Lots of Argument Mapping' (LAMP) is greater (0.78 of an SD) than for the Group 'Some Argument Mapping' (0.68 of an SD), suggesting that the more practice the better. This, however, requires further investigation, since the confidence intervals for the two Groups overlap greatly.

The difference is significant not only between Groups 3 and 4 (argument mapping within philosophy) and Group 1 (philosophy without CT), but between Groups 3 and 4, on the one hand, and Groups 2, 5, 6 and 7 on the other. *This is the single most striking finding of the meta-analysis.* Prima facie, it suggests that argument mapping within philosophy is the best way to improve CT. It is not clear, though, that it needs to be within philosophy – any more than Logic does, or any more than the Keller Plan need be a teaching method confined to philosophy departments. What remains to be explored is how great an improvement in CT would be generated by teaching CT using argument mapping *outside the context of philosophy*. This has

yet to be attempted, perhaps, in part, because argument mapping has largely, and only recently, been invented by philosophers.

It is, of course, just these results, in Groups 3 and 4, which actually explain the finding that, over all, philosophy departments appear to be better at teaching CT skills than Non-philosophy Departments. *What follows from this is that it is argument mapping that is the key variable.* It makes no discernible difference whether you teach traditional CT in or outside of a philosophy Department; the results are much the same. Yet when you teach CT within a philosophy Department using argument mapping, the results seem to be strikingly better. Plainly, this compels the question, is it the context of a philosophy department or the use of argument mapping (or perhaps the combination) that is decisive? The context, clearly, is insufficient on its own; but is argument mapping sufficient on its own? We cannot yet say with confidence. Further research is needed to pin this down.

To summarize all that can be concluded from the data with respect to our third question: Anglo-American analytic philosophy, in itself, is not a particularly effective way to improve CT skills, when it is compared to other subjects, or to CT courses. Equally, traditional CT makes no real difference, compared with Anglo-American analytic philosophy, to the improvement of CT skills. Philosophy departments are more effective than otherwise at teaching CT skills when it is done using argument mapping. Argument Mapping courses are by far the most effective way to improve CT skills. But there are a number of loose ends to tie up here.

There is an apparent anomaly in the data. Whereas Anglo-American analytic philosophy seems to be no more effective than either traditional CT taught within philosophy, or outside it, *or* than other subjects without CT; traditional CT itself appears to be better than other subjects without CT, even though it is *not* better than philosophy. This requires some explanation, because it is not immediately apparent from Fig. 5 and is based on statistical subtleties that will not be readily understood by non-statisticians. It is derived from a careful analysis of Confidence Intervals, as regards actual statistical significance. The results presented in the meta-analysis indicate that there exists a significant difference ( $p\text{-value} < .01$ ) between all the traditional CT courses offered by all departments (philosophy or non-philosophy) (see Fig. 5 in the results sections) and other subjects.

Before we come to the more general conclusions that can be drawn from our interpretation of the results, let us briefly recap on what we have deduced to this point. First, it is difficult to determine, from the data we have, to what extent the gain of 0.26 of an SD is attributable to the study of Anglo-American analytic philosophy. There are several factors which may have contributed to producing the net effect that the discipline appears to have on the development of CTS. These factors include the teaching methods used, most notably the Keller Plan and LAMP. The latter is a special case, but both merit close attention.

Another factor is the specific content of particular courses which, though they may be taught in a philosophy course, are not, by their nature, necessarily philosophy courses. Here, the most notable case is Logic. A further factor is the possible self-selection of philosophy students, as unusually bright students precociously committed to, or with a prior disposition to develop their CTS. Finally, there is an influence of uncertain magnitude which can be attributed simply to the experience of being at university, independently of the specific course of studies undertaken.

Second, even if we accept the effect attributed to philosophy, of roughly a quarter of an SD, this does not make the discipline of philosophy the most effective way to improve CTS. Close analysis of the relevant comparisons, as we have seen, demonstrates that, based on all the available data, philosophy does not, statistically speaking; actually improve CTS over and above what can be expected from university education in general or standard CT courses.

Additional considerations emerge from our interpretation of the data, in this regard. CT courses taught using lots of argument mapping (LAMP) with the support of software tools, and concentrating on exercises that call only on general knowledge, seem to be the most effective of all current methods for improving CTS. Conversely, and perhaps surprisingly, traditional CT courses, whether or not these are offered by Philosophy departments, prove to be more effective than other courses at improving CTS, but only marginally and not to anything like the extent seen with LAMP. These basic findings, important in themselves, serve as the point of departure for the final conclusions and implications for future research, which are set out in the following section of the thesis.

## 7 Conclusions

### 7.1 Summary of the case

A meta-analysis allows us to place many individual studies in comparative perspective and to gain an overview of the whole body of research bearing on a given topic. The analytical evidence it provides, by breaking down and reducing to common measures a mass of empirical evidence, allows for a clear and systematic evaluation of the topic. Such evidence thus provides the ground on which insight and action can proceed with some confidence. To the best of our knowledge, this is the first time that a meta-analysis has been conducted regarding the impact of the discipline of philosophy on the development of Critical Thinking Skills. The utility of this exercise is twofold. It has thrown into relief what was previously only guessed at; and it has indicated where further research would be useful.

We set out to test whether philosophy, specifically, improved CT skills and to establish if it was the *most effective* way to improve them – a widely held assumption, among philosophers. We have found that although it *marginally* improves CT skills, it is *not* the most effective way to do it. Two reasons support this claim. First, there is, simply, insufficient evidence to support the claim that it is any more effective than other standard methods or other subjects. Second, the study of philosophy appears to be less effective than the study of critical thinking skills in their own right, although the evidence is not altogether conclusive. Third, the available evidence strongly suggests that philosophy as such is strikingly *less* effective than LAMP (Lots of Argument Mapping Practice) – an innovative approach to teaching CT; with the caveat that no-one has yet tested LAMP outside the context of philosophy.

While the focus of this thesis is on the assumption that philosophy *is* the best discipline for improving CTS (an assumption that it undermines), its findings point to a more general observation: that there may be a complacent and unwarranted assumption across the tertiary education sector that critical thinking skills are in general taught well, or at least as well as they might reasonably be. For the data examined here suggest that *far greater improvements are in general possible*. They challenge all practitioners to raise the bar and also to become much more self-critical in assessing the impact of both their disciplines, as normally conceived, and their methods for teaching CTS.

There are also some indications in the findings of the thesis that both specifically *what* is taught (Logic, for instance, as compared with philosophy subjects less directly concerned with reasoning skills in themselves) and *how* it is taught (Keller Plan or LAMP) are the crucial considerations. This should not seem at all surprising, of course. Its implications, however, are

that claims for efficacy in teaching CTS should be confined to very specific subject content and teaching methods, not to broad disciplines or, indeed, disciplines as such, independent of the approach to teaching them.

## 7.2 Future Directions for Research

If we wanted to be more confident of these findings, there are some lines of research that it would be useful to undertake:

There have been far too few studies of the impact on CTS of philosophy. The point of departure for this inquiry was, of course, the question whether the assumption that philosophy is an especially effective way to teach CTS is warranted. The meta-analysis makes clear that, *based on existing data*, this assumption cannot be sustained. If, of course, there was a very much greater pool of studies on which to draw, this finding might turn out to be in error; but such studies have not been done or found. Until they are and, pending their findings, no philosophy department is justified in asserting that students should study philosophy, rather than other courses, in order to improve their CTS.

Specifically, the data here presented would suggest that there is a need for further, detailed studies of the impact of Logic on CTS and of the impact of the Keller Plan approach to teaching, when applied to the teaching of CTS. We need to establish more clearly whether there is any difference between general CTS and domain specific CTS. Should it turn out to be the case that there is some domain specific CTS, the implication could be that neither philosophy nor dedicated CT courses would be the most effective way to teach these particular skills. Each of these (philosophy and CT courses) is, in the nature of the case, committed to the proposition that CTS are general, not domain specific skills.

Further studies are clearly needed of the impact on CTS of a number of specific disciplines, such as mathematics, basic physical science, economics, engineering and the humanities. At present, there are numerous studies in Nursing and dedicated CT courses, but altogether too few of other major disciplines. There is a lot of scope here, therefore, for carefully designed and coordinated research, so that we can ascertain with much higher confidence what courses or approaches most contribute to, or inhibit the development of CTS.

Given the striking evidence that LAMP delivers dramatically better results than philosophy, CT courses or other subjects, but that it has, thus far, only been taught within the context of philosophy courses and only to first year under-graduates over a single semester, further studies are called for. Such studies should include further examination of the impact of LAMP in philosophy courses; but they should broaden to include experimental use of LAMP in a range of other disciplines and on its own.

We need studies of the impact of LAMP on various groups of philosophy students, testing different variables. We also need studies of students over more than a single semester; and on students beyond first year university. We need similar differentiation in the studies of students in contexts other than philosophy courses. Finally, all further studies badly need to use a consistent and rigorous research and reporting methodology, if they are to be reliable and useful. There has been an immense wastage in studies conducted to date, owing to the deficiencies in their design and reporting procedures.

## 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](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.
- Rimienne, 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.