The promotion of critical thinking skills through argument mapping

CHAPTER · JANUARY 2011

3 AUTHORS:

Christopher Dwyer
National University of Ireland, Galway

Michael J Hogan
National University of Ireland, Galway

Ian Stewart
National University of Ireland, Galway

In: Critical Thinking ISBN 978-1-61324-419-7
Editor: Christopher P. Horvath and James M. Forte 2011 Nova Science Publishers, Inc.

Chapter 4

THE PROMOTION OF CRITICAL THINKING SKILLS THROUGH ARGUMENT MAPPING

Christopher Dwyer, Michael Hogan and Ian Stewart
School of Psychology,
National University of Ireland, Galway

ABSTRACT

Argument mapping is a method of visually diagramming arguments using a 'box and arrow' format with the aim of simplifying the reading of an argument structure and facilitating the assimilation of core statements and relations. The current chapter presents the findings of a controlled trial in which argument mapping training was compared with hierarchical outline training as techniques for teaching critical thinking skills. Eighty-one undergraduate psychology students were allocated to one of three groups: an argument mapping group, an outlining group, or a control group and were tested on critical thinking before and after an 8-week intervention period. Results revealed that students in the argument mapping group scored higher than the control group at post-test on the critical thinking skills of evaluation and inductive reasoning. Students in the outlining group scored significantly higher than those in the control group on tests of analysis and inductive reasoning. There were no significant performance differences at post-test between those in the argument mapping group and those in the hierarchical summary group. Results are discussed in light of research and theory on best practice in the cultivation of critical thinking.

INTRODUCTION

Critical thinking is a metacognitive process which is made up of a collection of sub-skills (i.e. analysis, evaluation, and inference) that, when used appropriately, increases the chances of producing a logical solution to a problem or a valid conclusion to an argument. In education reports around the world, the teaching of critical thinking skills has been identified as an area of education to be developed and examined, specifically, in higher education (Association of American Colleges & Universities, 2005; Australian Council for Educational Research, 2002; Higher Education Quality Control, 1996), because it endows students with the capability to reason not only academically, but also in social and interpersonal contexts where adequate problem-solving and decision making are necessary on a daily basis (Ku, 2009).

Though the benefits of critical thinking are not always obvious to many students in third-level education, it is well established that good critical thinking ability predicts both academic and everyday functioning (Quitadamo & Kurtz, 2007). Good critical thinkers are more likely to get better grades; are better equipped and more likely to use the skills of critical thinking on an everyday basis (U.S. Department of Education, 1990); and are often more employable as well (Holmes and Clizbe, 1997; National Academy of Sciences, 2005). The ability to think critically allows those in the workforce to think independently, analyse data in order to make inferences, communicate well and make sound decisions. In addition, critical thinking skills are highly desired by employers for their workforce (El Hassan & Madhum, 2007) and are also essential for good management (MacPherson, 1999).

However, teaching critical thinking (CT) skills to University students is a major educational challenge (Kuhn, 1991; Willingham, 2007). There are many reasons for this, including the broad challenges of embedding CT into an often crowded curriculum, and designing an effective teaching strategy that targets specific CT skills and offers sufficient practice so these skills develop in an orderly and cumulative way. Two related problems discussed in more detail below include: difficulties in defining critical thinking and constructing thinking frameworks and theories that inform the practice of teachers in the area; and difficulties associated with the assimilation of text-based argument and the challenge of teaching students transferable analysis, evaluation, and inference skills using text-based teaching materials. We will consider both of these issues below and then describe how theory-driven argument mapping training might serve to resolve these problems.

WHAT IS CRITICAL THINKING?

There are many definitions and measures of critical thinking. This variety can make it difficult for researchers and teachers to understand or agree on the key components of good critical thinking and these difficulties may impede their ability to construct an integrated theoretical account of how best to train critical thinking skills. In the absence of greater clarity in relation to the components of critical thinking skill and the way these components work together in the context of solving critical thinking problems, it can be difficult to design critical thinking training programs.

In the past century, there has been little agreement on how to conceptualise critical thinking. John Dewey (1933) provided one of the first multi-level models of thinking in his classic book, *How We Think*. Each level of thinking in Dewey's system differs in terms of its adequacy for the purpose of achieving "active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and the further conclusions to which it tends" (Dewey, 1933, p. 8).

In the first level of Dewey's system is the *stream of consciousness* (e.g. day dreaming), an "uncontrolled coursing of ideas through our heads" (p. 6). At the next level, Dewey describes *imagination* as a more orderly and controlled type of thinking, specifically, where "successions of imaginative incidents and episodes that have a certain coherence, hang together on a continuous thread, and thus lie between kaleidoscope flights of fancy and considerations deliberately employed to establish a conclusion" (p. 6). The third level of thinking in Dewey's system "is practically synonymous with belief" - belief that is accepted or rejected as a set of conclusions, but "not conclusions reached as the result of personal mental activity, such as observing, collecting, and examining evidence" (p. 7). Dewey draws a contrast between this form of thinking and *reflective thinking*, the highest level of thinking in his cognitive system:

> "…Columbus did not accept unhesitatingly the current traditional theory…Skeptical of what, from long habit, seemed most certain, and credulous of what seemed impossible, he went on thinking until he could produce evidence for both his confidence and his disbelief. Even if his conclusion had finally turned out wrong, it would have been a different sort of belief from those it antagonized, because it was reached by a different method. *Active, persistent, and careful consideration of any belief or supposed form of knowledge in the light of the grounds that support it and the further conclusions to which it tends constitutes reflective thought*. Any one of the first three kinds of thought may elicit this type; but once begun, it includes a conscious and voluntary effort to establish belief upon a firm basis of evidence and rationality" (Dewey, 1933, p. 8, italics added).

Dewey's conceptualization of reflective thinking helped to inform more recent conceptualizations of critical thinking. Similarly, recognition of the importance of critical thinking in education followed the growth of interest in informal logic, which was initiated in part by the work of Stephen Toulmin in the late 1950's (Allen, Feezel & Kauffeld, 1967). Informal logic is a type of logic that emphasises the justificatory function of argumentation, namely that a good argument requires sufficient support (e.g. reliable and valid empirical evidence). Toulmin's focus on informal reasoning helped to transform argumentation training initiatives in university, which traditionally had focused on training in formal logic (i.e., reasoning using syllogisms). Deliberations in relation to critical thinking skills grew in part from the notion of informal logic, where the central focus is on the analysis and evaluation of claims - claims could only be made (and justified) after a sufficient amount of analysis and evaluation had been conducted on propositions and their logical interdependencies within the arguments used to support these claims. Since then, dozens of definitions for critical thinking have been offered (see Table 1).

Though there have been dozens of attempts at defining critical thinking, many of the definitions are quite vague. Many of the authors present a number of skills necessary for good critical thinking rather than providing an operational definition that informs measurement and analysis of the skills listed. The one thing that all the authors seem to agree on is that critical thinking is in fact a collection of cognitive and metacognitive skills centred on the analysis and evaluation of beliefs and the ability to draw sound inferences. Though it is reasonable to suggest that a straightforward, singular description and operational definition of critical thinking is not possible due to the variety of perspectives on critical thinking, there is a need for some reasonable group consensus in an educational context; as an agreed upon operational definition is necessary to conduct educational research in this area and, more importantly, to compare findings across different groups and intervention studies.

Though debate is ongoing over the definition of critical thinking and the core skills necessary to think critically, to date, there has been only one definition and list of skills that stands out as a reasonable consensus conceptualisation of critical thinking. In 1988, a committee of 46 experts in the field of critical thinking, known as the *Delphi Committee*, gathered to discuss a definition of critical thinking. The committee also discussed the skills necessary to think critically. The findings taken from this meeting, known as *The Delphi Report,* written by Peter Facione (1990), defined critical thinking as:

Table 1. Definitions and Descriptions of Critical Thinking

Author	**Definition/Description**
Glaser (1941)	Critical thinking is: an attitude of being disposed to consider, in a thoughtful way, problems and subjects that come within the range of one's experience; knowledge of the methods of logical enquiry and reasoning; and 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.
Ennis (1987)	Critical thinking is reasonable, reflective thinking, focused on deciding what to believe or do.
Kurfiss (1988)	Critical thinking is the ability to detect and avoid fallacious reasoning and to analyse deductive and inductive arguments.
Allegretti & Frederick (1995)	Critical thinking is evaluating the arguments of others, evaluating one's own arguments, resolving conflicts and understanding the source of conflicts in argumentation; thus coming to a resolution in complex problems and gaining confidence in one's own thinking processes (Allegretti & Frederick, 1995).
Paul (1993)	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.
Wilkinson (1996)	Critical thinking is goal-oriented, purposeful thinking that involves a number of mental skills, such as determining what data is relevant, evaluating the credibility of sources and making inferences.

Bensley (1998)	Critical thinking is reflective thinking in which a person evaluates relevant evidence and works to draw a sound or good conclusion.
Halpern (2003)	Critical thinking is purposeful, reasoned and goal-directed thinking – the kind of thinking involved in solving problems, formulating inferences, calculating likelihoods and making decisions.
Thomson (2009)	Critical thinking involves the identification and evaluation of reasons and conclusions within an argument, the ability to draw one's own conclusions and the use of appropriate language in order to communicate and construct one's own arguments.

> "…purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based." (p. 3).

Furthermore, the Delphi panel overwhelmingly agreed (i.e. 95% agreement) that *analysis, evaluation* and *inference* were the core skills necessary for critical thinking (Facione, 1990). These skills (as described by the Delphi Report) are presented in Table 2. The definition of critical thinking provided by the Delphi Report was adopted by the American Philosophical Association (APA) and as a result, has become a widely accepted definition for good critical thinking (Beckie, Lowry & Barnett, 2001). The same definition of critical thinking was also used by the U.S. Department of Education as a framework for setting its educational goals (Facione, Facione, Blohm & Giancarlo, 2002). The Delphi definition of critical thinking also inspired the creation of the California Critical Thinking Skills Test (CCTST; Facione, 1990), a test that is commonly used as part of the evaluation of critical thinking intervention studies. At the same time, the challenges of teaching critical thinking skills remain, and a question remains as to how best to teach critical thinking skills.

Can Critical Thinking Skills Be Taught? A Look at the Previous Research

Critical thinking (CT) courses have been taught at University in varying academic domains including law, philosophy, psychology, sociology and nursing. Importantly, it is often argued that critical thinking is a domain-general skill that can be taught alongside any academic content (Gabbenesch, 2006). At the same time, whether or not CT can be improved via explicit instruction and how it is best improved are issues that continue to be debated in the literature. This debate is fuelled in part by difficulties interpreting and comparing research studies in the area.

Table 2. The Core Critical Thinking Skills According to the Delphi Report

Skill	Description
Analysis	To identify the intended and actual inferential relationships among statements, questions, concepts, descriptions or other forms of representation intended to express beliefs, judgments, experiences, reasons, information, or opinions. Examining ideas: to determine the role various expressions play or are intended to play in the context of argument, reasoning or persuasion; to compare or contrast ideas, concepts, or statements; to identify issues or problems and determine their component parts, and also to identify the conceptual relationships of those parts to each other and to the whole. Detecting arguments given a set of statements, descriptions, questions or graphic representations, to determine whether or not the set expresses, or is intended to express, a reason or reasons in support of or contesting some claim, opinion or point of view. Analysing arguments: given the expression of a reason or reasons intended to support or contest some claim, opinion or point of view, to identify and differentiate: (a) the intended main conclusion, (b) the premises and reasons advanced in support of the main conclusion, (c) further premises and reasons advanced as backup or support for those premises and reasons intended as supporting the main conclusion, (d) additional unexpressed elements of that reasoning, such as intermediary conclusions, non-stated assumptions or presuppositions, (e) the overall structure of the argument or intended chain of reasoning, and (f) any items contained in the body of expressions being examined which are not intended to be taken as part of the reasoning being expressed or its intended background.
Evaluation	To assess the credibility of statements or other representations which are accounts or descriptions of a person's perception, experience, situation, judgment, belief, or opinion; and to assess the logical strength of the actual or intend inferential relationships among statements, descriptions, questions or other forms of representation. Assessing claims: to recognize the factors relevant to assessing the degree of credibility to ascribe to a source of information or opinion; to assess the contextual relevance of questions, information, principles, rules or procedural directions; to assess the acceptability, the level of confidence to place in the probability or truth of any given representation of an experience, situation, judgment, belief or opinion.

	Assessing arguments: to judge whether the assumed acceptability of the premises of a given argument justify one's accepting as true (deductively certain), or very probably true (inductively justified), the expressed conclusion of that argument; to anticipate or to raise questions or objections, and to assess whether these point to significant weakness in the argument being evaluated; to determine whether an argument relies on false or doubtful assumptions or presuppositions and then to determine how crucially these affect its strength; to judge between reasonable and fallacious inferences; to judge the probative strength of an argument's premises and assumptions with a view toward determining the acceptability of the argument; to determine and judge the probative strength of an argument's intended or unintended consequences with a view toward judging the acceptability of the argument; to determine the extent to which possible additional information might strengthen or weaken an argument.
Inference	To identify and secure elements needed to draw reasonable conclusions; to form conjectures and hypotheses; to consider relevant information and to educe the consequences flowing from data, statements, principles, evidence, judgments, beliefs, opinions, concepts, descriptions, questions, or other forms of representation. Querying evidence: in particular, to recognize premises which require support and to formulate a strategy for seeking and gathering information which might supply that support; in general, to judge that information relevant to deciding the acceptability, plausibility or relative merits of a given alternative, question, issue, theory, hypothesis, or statement is required, and to determine plausible investigatory strategies for acquiring that information. Conjecturing alternatives: to formulate multiple alternatives for resolving a problem, to postulate a series of suppositions regarding a question, to project alternative hypotheses regarding an event, to develop a variety of different plans to achieve some goal; to draw out presuppositions and project the range of possible consequences of decisions, positions, policies, theories, or beliefs. Drawing conclusions: to apply appropriate modes of inference in determining what position, opinion or point of view one should take on a given matter or issue; given a set of statements, descriptions, questions or other forms of representation, to educe, with the proper level of logical strength, their inferential relationships and the consequences or the presuppositions which they support, warrant, imply or entail; to employ successfully various sub-species of reasoning, as for example to reason

	analogically, arithmetically, dialectically, scientifically, etc; to determine which of several possible conclusions is most strongly warranted or supported by the evidence at hand, or which should be rejected or regarded as less plausible by the information given.

Adapted from Facione, 1990.

CT courses are taught in a variety of different academic domains and are informed by varying conceptualisations of CT. Different intervention studies also use different measures of CT performance that are not directly comparable - the California Critical Thinking Skills Test (Facione, 1990), the Cornell Critical Thinking Test (Ennis, Millman & Tomko, 1985) and the Watson-Glaser Critical Thinking Assessment (Watson & Glaser, 1980). The reported reliability and validity of different measures also varies, which has led Abrami and colleagues to ask: "How will we know if one intervention is more beneficial than another if we are uncertain about the validity and reliability of the outcome measures?" (Abrami et al., 2008, p. 1104). Abrami and colleagues add that, even when researchers declare that they are assessing CT using reliable and valid assessment tools, there still remains the major challenge of ensuring that measured outcomes are related in some meaningful way to the conceptualisation and operational definition of CT that informs their teaching practice. Often, the relationship between the concepts of CT that are taught and those that are assessed is unclear and a large number of studies in this area specify no theory to help elucidate these relationships.

Nevertheless, researchers have attempted to group intervention studies in an effort to examine whether or not critical thinking can be improved via explicit instruction. For example, a recent meta-analysis by Alvarez-Ortiz (2007) examined 52 studies which investigated a wide range of teaching strategies designed to improve CT. The meta-analysis was specifically conducted in order to answer the questions as to whether or not participation in philosophy courses improved CT ability. Results of the meta-analysis revealed that participation in a philosophy course yielded a mean effect size of .26 SD, CI [.12 - .39], with little evidence to suggest that participation in a philosophy course had any greater effect on CT performance than any other academic course (mean effect size = .12 SD, CI [.11, .21]). However, this meta-analysis also suggested that all courses (regardless of academic content) that *directly* taught CT (effect size of .40, CI [.08, .71]) or had CT *infused* into the curriculum (effect size of .26, CI [.09, .43]) yielded better CT performance than courses that did not teach CT in some form (effect size of .12 SD, CI [.08, .17]). These findings lend some support to Gabbenesch's (2006) claim that CT is domain-general, as the course content was not the key factor in improving CT, whereas involvement of some form of explicit CT instruction was fundamental.

Another meta-analysis, conducted by Abrami and colleagues (2008) included 161 CT intervention studies and examined the efficacy of different types of CT training course. They used Ennis' (1989) typology of four CT course types (i.e. general, infusion, immersion and mixed) to differentiate CT intervention methods. In the *general* approach to CT training, CT skills, dispositions and processes "are learning objectives, without specific subject matter content" (Abrami et al., 2008, p. 1105). Conversely, the *infusion* method requires specific course content upon which CT skills are practiced. In the infusion approach, the objective of teaching CT alongside course content is made explicit. In the *immersion* method, like the infusion method, specific course content is required; however, while CT skills are practiced, CT objectives are *not* made explicit in the immersion approach. Finally, in the *mixed* approach, critical thinking is taught independently of the specific subject matter content of the course.

Abrami and colleagues (2008) reported a significant effect on CT performance (g+ = .34) of all CT courses included in the meta-analysis. However, only 91 of the studies assessed critical thinking ability using standardised tests (i.e. as opposed to using an assessment devised by a teacher or researcher), and these 91 studies yielded an average effect size of (g+) .24. Comparing the four CT course types, results of the meta-analysis revealed that courses using the mixed approach had the largest effect on CT performance (g+ = .94), followed by the infusion approach (g+ = .54), the general approach (g+ = .38) and the immersion approach (g+ = .09), respectively. It is important to note that the immersion typology (which had the smallest effect) is the only approach that does not make CT objectives explicit to students. Thus, making CT objectives clear to students may be an important part of any course design aimed at increasing CT ability (Abrami et al., 2008). More generally, the authors concluded that the enhancement of CT ability is greatly dependent upon how CT is taught and that the *mixed* approach to teaching CT worked best as students were required to learn CT skills separate from other course material and then apply them to the material later on in the course.

Abrami and colleagues' (2008) meta-analysis was conducted in light of some very broad distinctions between different types of CT training courses. However, less is known about how different instructional methods impact overall training benefits. In this chapter, we propose that teaching strategies that facilitate the assimilation of argument structures (i.e. analysis of argument structures), and an assessment of the quality of evidence and the logical relationships between propositions in moderately complex arguments (i.e. evaluation of argument structures) may in turn facilitate significant growth in analysis and evaluation skills. With a guiding theory and suitable experimental controls it is possible to compare different teaching strategies in this context.

Beyond Text-Based Learning: The Use of Thought Structuring Tools in CT Education

Central to our theory of CT enhancement is a focus on the problem of working memory demands associated with the assimilation and simultaneous analysis and evaluation of arguments. According to various frameworks for thinking (e.g. Dewey, 1933; Bloom, 1956; Anderson & Krathwohl, 2001; Moseley et al., 2005), there are a number of cognitive and metacognitive skills that are necessary for good thinking. For example, researcher and theorists often point to the ability to build understanding through the *organisation* of ideas (Moseley at el., 2005); and the ability to recognise, appraise and analyse both a *chain* of arguments and the *justification* of claims through reasoning (Allen, Feezel & Kauffeld, 1967). While these kinds of organizational, analytical, and evaluative skills may be fundamental components of good CT, Harrell (2005) notes that students often fail to understand the 'gist' (Kintsch & van Dijk, 1978) of text-based information presented to them; and more often, students cannot adequately 'follow' the argument of a text (i.e. the chain of reasoning and the justification of claims in the chain), as most students do not even acknowledge that information within a text presents an argument and instead read it as if it were a story. Conversely, authors who do understand the nature of argumentation often construct verbose 'maze-like' arguments that consist of massive amounts of text (Monk, 2001). Students who are presented with these texts may thus find it very difficult to capture anything more than the 'gist' of the argument. For example, because text-based arguments contain many more sentences than just the propositions that are part of the argument, these sentences may obscure the intention of the piece and the inferential structure of the argument (Harrell, 2004).

More specifically, as arguments are not sequential in nature, the linear nature of text sometimes makes it difficult to assimilate the information within a text-based argument (van Gelder, 2003). For example, when reading text, a person may read a statement on page three and not read any relevant support (or objection) to this claim until they reach page 16. Between pages 3 and 16, it could be that a variety of other propositions are presented, which places cognitive load on the reader. Cognitive load is the burden put upon an individual in using and distributing working memory resources during cognitive activities such as learning and problem-solving (Sweller, 1988, 1999). This additional load comes from the need to, for example, switch attention from one page to another and back and forth, in order to create some structure for a 'non-user friendly' text. While reading text-based materials, students must figure out the relationship between propositions for themselves, using whatever cues they can, regardless of the ambiguity of the text. Thus, in attempting to mentally structure arguments when reading text, the reader faces cognitive load.

Tindall-Ford, Chandler and Sweller (1997) found that learning is impeded when instructional materials require a high degree of attention switching. They concluded that encoding environments that increase the cognitive load placed on the reader tend not only to slow the learning process, but also reduce overall levels of learning. Presenting information in a way that reduces the level of attention switching may minimize the cognitive load and improve learning.

Having available the structure of an argument is crucial for many reasons: it facilitates logical reasoning, the answering of specific questions about the relation between one proposition and others, and the ready construction of a 'mental image' of the whole argument. Argument mapping is a learning aid which may facilitate thinking in this regard. For example, in the argument map, both the propositions and the relationships among them are explicitly stated and the information is presented in an integrated, organised fashion. Notably, it is generally the case that integrated, organised representations facilitate learning (Sweller, 1999). In previous research, argument mapping has been identified as a technique that might circumvent the many obstacles related to reading text and visualizing the argument simultaneously; and may also enhance overall levels of learning and CT (van Gelder, 2001). Thus, argument mapping is hypothesised as a tool that may support the cultivation of CT skills by helping to resolve the problem of working memory demands associated with the assimilation and simultaneous analysis and evaluation of arguments.

Argument Mapping as a Tool for Critical Thinking Instruction

In an argument map, a text-based argument is visually represented using a 'box-and-arrow' style flow-chart that makes the structure of the argument explicit to the reader by organising the propositions within the argument and by displaying all the connections amongst propositions within the argument (van Gelder, 2001). For an example of an argument map, see Figure 1.

Though computer-based argument mapping is a relatively new technique (van Gelder, 2000), some research has examined the efficacy of teaching CT skills using argument mapping as a tool of instruction. For instance, in her meta-analysis, Alvarez-Ortiz (2007) found that students who participated in critical thinking courses that used at least some argument mapping within the course achieved gains in CT ability with an effect size of .68 SD, CI [.51, .86]. In courses where there was "lots of argument mapping practice" there was also a significant gain in students' CT performance, with an effect size of .78 SD, CI [.67, .89]. These findings compare favourably to the effect sizes observed for participation in philosophy courses (average effect size = .26 SD, CI [.12 - .39]), any other academic course (effect size = .12 SD, CI [.11, .21]), and courses that *directly* taught CT or had CT *infused* into the curriculum (average effect size = .49 SD, CI [.39, .59].

Thus, there are a number of studies that have previously used argument mapping as a tool of CT instruction. For example, Tim van Gelder (2001) and van Gelder and Rizzo (2001) provided undergraduate philosophy students with a semester-long CT course, in which students were trained in CT through the use of argument mapping (AM). Students' CT ability was tested both before and after one college semester using alternate forms of the California Critical Thinking Skills Test (CCTST; Facione, 1990). Results revealed an improvement with an effect size of .84, which implied an impressive gain of almost one standard deviation in CT ability over the course of the semester. It is important to note that, though the authors credit much of this gain to AM training, they also admit that this gain could also be due to other aspects of the course, such as the practice regime. Notably, this study did not include a control group or an alternative CT training regime with which to compare AM training.

Similarly, van Gelder, Bissett and Cumming (2004) provided undergraduate philosophy students with a 12-week CT course taught through the use of AM. Students were pre-tested using the CCTST. During the course, students were provided with homework exercises and were free to complete as many practice exercises as they wished. Students also attended one tutorial per week in which they had access to both argument mapping software and to direct personal guidance from their tutors. After completion of the course, students were post-tested using the CCTST. Results revealed that CT scores increased significantly from pre- to post-testing with a large effect size of .8 SD, CI [.66, .94]. There was also a significant correlation between performance and AM practice hours (r = .31).

Butchart et al. (2009) compared two groups of students who attended AM-infused CT modules (i.e. a module with online automated feedback for AM exercises and a module that contained AM exercises only, with no automated feedback). These modules were compared in turn with a 'standard' CT module (i.e. no AM). CT training was heavily concentrated on two CT skills: analysis and evaluation, described by the *Delphi Report* as core critical thinking skills. Prior to commencement of the course, students completed the CCTST Form A as a pre-test. During the course, students were provided with eight homework assignments and 10 sets of exercises. Automated feedback was provided to students in the automated feedback AM group. After completion of the course, students were post-tested using Form B of the CCTST.

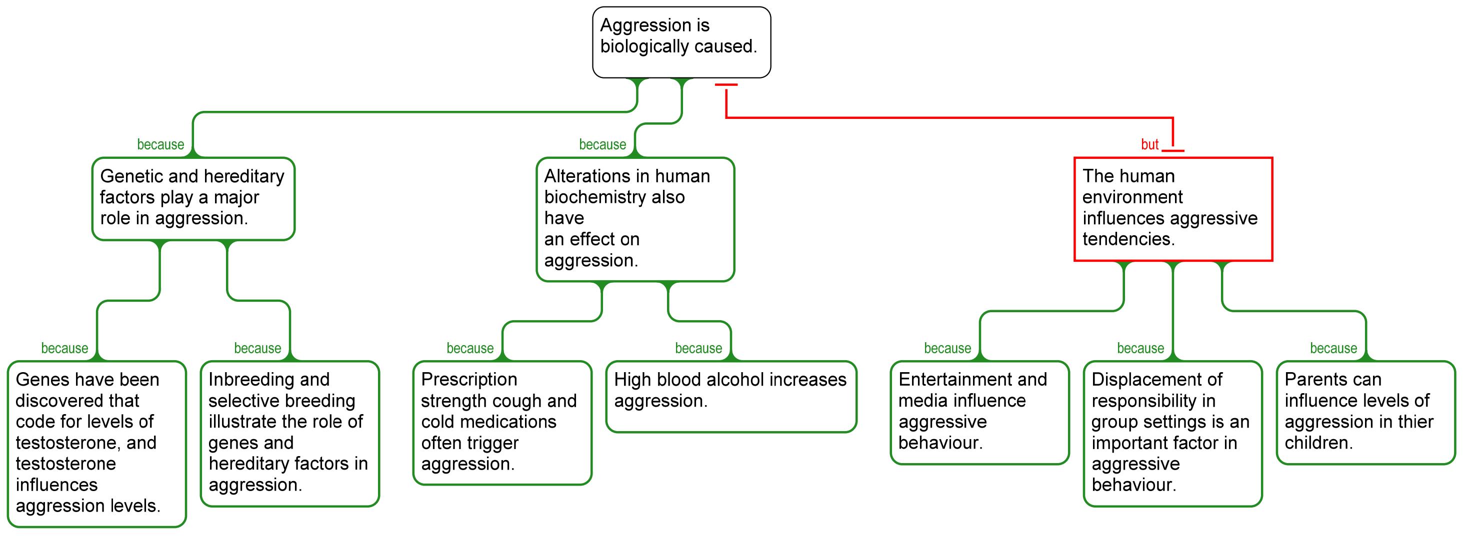

Figure 1. An example of an Argument Map created through Rationale™.

Butchart and colleagues found that those who received automated feedback for their AM exercises showed a significant gain in CT ability with a medium effect size of .45. Students who completed the AM exercises without automated feedback showed a gain with an effect size .22. Those who participated in a standard CT module showed a gain with an effect size of .19. Unfortunately, statistical differences among the three groups in this study were not reported. Furthermore, as admitted by the authors, participants in the automated feedback group could have been provided with more informative feedback, as opposed to simply receiving automated notice of a 'correct' or 'incorrect' response for placement of a proposition. One could argue that this automatic 'correction' of argument mapping exercises is not an ideal form of feedback, in the sense that an explanation as to *why* a response is incorrect would likely have been more informative to the student. Thus, although students who received automated feedback for their AM exercises showed the largest gain in CT ability, it is unclear how feedback worked to improve performance in this context.

Therefore, while research suggests that CT courses taught through the use of AM improve CT ability, there have been a number of problems with the research conducted to date. For example, two of the three studies described above (i.e. van Gelder & Rizzo, 2001; van Gelder, Bissett and Cumming, 2004) did not compare AM-infused CT training with a no-intervention control group or a comparable active intervention control group. In addition, although Alvarez-Ortiz's (2007) meta-analysis suggests that semester-long training courses in AM produce greater gains in CT skills (when compared with standard semester-long courses in introductory philosophy), AM training has not been directly compared with other methods of teaching CT skills, apart from one study where a standard CT course was used for comparison purposes (Butchart et al., 2009). Butchart and colleagues reported using the same course structure and teaching the same content "as far as possible" (Butchart et al., 2009, p. 278), but it is unclear how this worked in practice. Furthermore, though Butchart and colleagues compared three groups in their study, these groups were not adequately matched or randomly assigned to experimental conditions. Participants were assigned to experimental conditions based on the semester in which they registered for the CT course. Those in the automated feedback AM group participated in the first semester of the study; those in the standard CT training condition participated in the second semester; and those in the AM exercises only group participated in the third semester of the study. Also, the pre-and-post-test scores (and the resultant gains) of the three groups were not statistically compared, so it difficult to assess whether or not the groups possessed similar or different CT abilities prior to their participation in the course, and whether or not gains across conditions are statistically different from one another.

In summary, though evidence suggests that critical thinking can be taught and enhanced, research studies in this area are difficult to compare because of differences across studies in the conceptualisations of critical thinking that inform teaching practices and the selection of measures used to assess performance. Some research studies have examined the efficacy of AM-infused CT training; however, this research is also difficult to interpret due to the absence in some studies of a control condition for comparison purposes and the failure to randomly assign groups to experimental conditions in other studies. Therefore, further research is needed to provide more conclusive evidence in favour of the claim that argument mapping is a tool that facilitates critical thinking ability.

Rationale for the Current Research

Argument maps and argument mapping may be a useful pedagogical aid, particularly in situations where students are working to analyse and evaluate complex arguments. The current research is part of a larger set of studies designed to examine the effects of argument mapping on memory for arguments (Dwyer, Hogan, & Stewart, 2010) and growth in critical thinking skill and reflective judgment. The following study examined the effect of AM training on CT skill. Critical thinking performance of those who attended an AM-infused CT seminar series was compared with the performance of those who attended a CT seminar series using identical content but taught using more traditional, hierarchical outlines (HO; see below). The performance of students in both of these active critical thinking training courses was compared to the performance of students who received no explicit CT training. A further aim of the study was to examine the effect of both AM and HO training on students' disposition towards thinking.

Based on previous research (Butchart et al., 2009; van Gelder, Bissett and Cumming, 2004; van Gelder & Rizzo, 2001), we hypothesised that AM training would result in larger gains in CT ability over the course of the semester when compared with both HO training and the control condition. More specifically, though HO organises information for the reader, the structure of an argument is represented as a linear flow of text and it does not make use of a box-and-arrow format, colour cues to represent reasons, objections, and rebuttals, or relational cues (i.e. *but, because* and *however*) that link propositions. However, because information within an HO is hierarchically organised, we hypothesise that training in HO (like AM training) would result in larger gains in CT ability over the course of the semester when compared with the control condition.

METHOD

Design

A series of six one-way ANCOVAs were used to assess the effect of the three experimental conditions (AM, HO, & Control) on six ability outcomes: overall CT, analysis, evaluation, inference, deductive reasoning and inductive reasoning, while controlling for baseline CT skill ability. Similarly, a series of seven ANOVAs were also used to assess the effects of the experimental conditions on students' disposition towards thinking.

Participants

Participants were first year psychology students (N = 81; 57 females, 24 males), aged between 18 and 25 years, from the National University of Ireland, Galway. In return for their participation, students were awarded academic course credits. To ensure confidentiality, participants were identified by ID number only.

Materials and Measures

The CT intervention materials used in this study were the exercise handouts and CT recordings, and a laptop, a projector and DVDs which were used to present the pre-recorded seminar series. These materials are available upon request.

The California Critical Thinking Skills Test (CCTST; Forms B) was administered as a baseline measure at the pre-test session. The CCTST was developed by Peter Facione and colleagues (1990; 2002). The CCTST consists of 34 multiple choice questions, which examine overall CT ability as well as five sub-skills: analysis, evaluation, inference, inductive reasoning, and deductive reasoning. Results are presented as raw scores and are additionally presented as U.S. national percentile equivalents of approximately 2,000 university students. Test reliability ranges from 0.78– 0.84 (Facione, 1991).

The CCTST (Form 2000) was administered at the post-testing session. Gain was not measured from pre- to post-testing (i.e. from Form B to Form 2000). Rather, in accordance with Jacobs (1995), the CCTST (Form B) was administered as a baseline measure and analysed as a covariate, whereas the CCTST (Form 2000) was administered as an outcome measure and analysed as the dependent variable.

The California Critical Thinking Disposition Inventory (CCTDI; Facione & Facione, 1992) was administered at post-testing. Seven subscales of the CCTDI include: truth seeking, open mindedness, analyticity, systematicity, confidence, inquisitiveness and maturity.

Finally, a questionnaire was administered at the end of the course which asked students to rate various facets of the course, such as their ability to understand the course, the quality of the materials, and the quality of the instruction.

Procedure

The study took place over eight weeks. The two experimental groups attended a 16 hour CT seminar series over the course of eight weeks, differing only in method of presentation (i.e. AM-infused or HO -infused CT training). The seminar series was designed to teach CT according to the framework provided by the Delphi Report and the American Philosophical Association. The control group did not attend any CT seminars.

In Week 1, the California Critical Thinking Skills Test (Form B) was administered. The seminar series began in Week 2. Seminars were given to four different groups per week: two of which were AM groups and two of which were HO groups. Both AM and HO lectures were identical in content and pre-recorded voice-over, which was dubbed over a PowerPoint™ slideshow using the Echo360™ system recording. Only the slideshows and in-class handouts varied (for purposes of presenting either AM or HO strategies for organizing arguments). The voice-over was performed by the same person (male: research supervisor) and this person did not facilitate the delivery of these recordings to students in the seminar. Independent evaluators rated the quality of the voice-over and judged whether or not there were any substantial differences in the quality of AM and HO delivery. Quality of voice-overs were rated highly and no differences in quality between AM and HO conditions were noted.

Table 3. Critical Thinking Course Outline

Class No.	Title	What Was Taught
1	Pre-Testing	• Students completed the CCTST (Form B) pre-test.
2	Session 1: *"Introduction to Critical Thinking"*	1. We think in order to decide what to do and what to believe. 2. We ultimately decide what to believe by adding supports or rebuttals to our own arguments (i.e. questioning our own beliefs). 3. Arguments are hierarchical structures. We can continue to add more levels if we like.
3	Session 2: *"Unpacking (analysing and evaluating) a persons' belief"*	1. In order to analyse an argument, we must extract the structure of the argument from dialogue or prose. 2. Identifying types (sources) of arguments and considering the strength of each type is another form of analysis. 3. The evaluation of the overall strengths and weaknesses of an argument can be completed after adequate analysis.
4	Session 3*: "Analysis & Evaluation"*	1. Evaluation includes the recognition of imbalances, omissions and bias within an argument. 2. Evaluative techniques can aid recall. 3. Examining whether or not the arguments used are *relevant or logically connected* to the central claim is also an important factor in evaluation.
5	Session 4: *"Evaluation"*	We must evaluate: 1. Types (sources) of arguments based on credibility 2. The relevance of propositions to the central claim or intermediate conclusions within the argument 3. The logical strength of an argument structure 4. The balance of evidence within an argument structure
6	Session 5: *"Inference"*	1. Evaluation and inference are intimately related. 2. Inference differs from evaluation in that the process of inference involves *generating* a conclusion from previously evaluated propositions. 3. In larger informal argument structures, intermediate conclusions must be inferred prior to the inference of a central claim.
7	Session 6: *"Making Another's Argument Your Own"*	• Review of all the previous 5 sessions
8	Post-Testing	• Students completed the CCTST (Form 2000) post-test and the CCTDI

In the seminars, students were taught skills and then shown how to use them via worked examples. During the seminars, the recordings were often paused and restarted in order to allow time for the completion of exercises. Students were given enough time so that they could actively learn by applying the skills they had just learned. On average, approximately 75% of the time allotted to each class was dedicated to this active learning. The course outline and what was taught in each class is presented in Table 3.

In Week 8, after completion of the seminar series, the CT ability of all three groups was again measured using CCTST Form 2000. In Week 8, the CCTDI was administered to all groups in order to examine students' disposition toward thinking. Students also completed a questionnaire which asked them to rate various facets of the course and make suggestions for improving the course. Students who did not complete the course (e.g. those who simply dropped out), were also given the questionnaire and were asked for reasons why they did not complete the course.

RESULTS

Means and standard deviations for the three groups are presented in Table 6. A series of between-subjects ANCOVAs were conducted to examine the effects of experimental conditions on CT outcomes, while also controlling for baseline CT ability. A preliminary analysis evaluating the homogeneity-of-slopes assumption revealed that there was no significant difference amongst groups on pre-test CT or sub-skill performance.

There was a main effect of group on analysis performance, F (2, 77) = 4.74, MSE = .04, p = .011, partial η^2 = .11, with those in the HO group scoring significantly higher on the analysis post- test than those in the control group ($p < .05$). Post-hoc analysis revealed border-line difference between the AM and control group, F (1, 77) = 3.59, MSE = .04, p = .06, with the AM group scoring higher than the control group.

Though there was no main effect of group on evaluation performance, F (2, 77) = 2.35, p = .103, there were some interesting trends in the data. Specifically, post hoc analyses revealed that the AM group scored significantly higher than the control group on post testing, F (1, 77) = 4.29, MSE = .02, p = .042. There was a borderline main effect of group on inductive reasoning, F (2, 77) = 3.08, p = .052. Post-hoc analyses revealed that the AM group scored significantly higher than those in the control group, F (1, 77) = 4.44, MSE = .02, p = .038; and that the HO group also scored significantly higher than those in the control group, F (1,77) = 4.52, MSE = .02, p = .037. No other effects were observed.

A further series of eight ANOVAs were conducted in order to examine the effects of group on disposition and the sub-scales of the CCTDI. There were no main effects of group on overall disposition score, or on the sub-scale scores of truth seeking, open mindedness, analyticity, systematicity, confidence, inquisitiveness or maturity. However, overall disposition score was significantly correlated with post-test CT performance (r = .45, p = .001), but not with pre-test CT performance (r = .16, p = .244).

Table 6. Means and standard deviations (%) and sample size (N) for the three groups at both pre-test and post-test

	Pre-Test			Post-Test	
	N	M	SD	M	SD
Overall CT					
AM	23	.48	.15	.51	.14
HO	28	.44	.16	.50	.13
Control	30	.41	.15	.43	.13
Analysis					
AM	23	.55	.15	.67	.17
HO	28	.49	.18	.72	.16
Control	30	.45	.19	.58	.22
Evaluation					
AM	23	.46	.17	.45	.17
HO	28	.41	.14	.41	.17
Control	30	.36	.14	.34	.13
Inference					
AM	23	.44	.16	.47	.16
HO	28	.40	.15	.46	.15
Control	30	.41	.16	.43	.16
Inductive Reasoning					
AM	23	.49	.16	.59	.13
HO	28	.42	.16	.58	.15
Control	30	.40	.13	.50	.14
Deductive Reasoning					
AM	23	.46	.15	.42	.16
HO	28	.45	.16	.42	.16
Control	30	.40	.13	.36	.16

DISCUSSION

We examined the effects of AM-infused and HO-infused CT training on students' CT performance. Performance on various sub-skills of CT (i.e. analysis, evaluation, inference, inductive reasoning and deductive reasoning) was measured both before and after the intervention, as was students' disposition towards thinking on post-testing.

Results revealed that students in the HO condition performed better than students in the control group on analysis and inductive reasoning at post-testing. Students in the AM condition performed better at post-test than students in the control condition on evaluation and inductive reasoning. Unlike van Gelder and colleagues (2001, 2003, 2004), we did not find a significant effect of AM training on overall CT performance. Although we must evaluate the results of the current study with caution, findings suggest that certain critical thinking skills (i.e. evaluation and inductive reasoning) can potentially be enhanced by argument mapping training. These results also suggest that certain critical thinking skills (i.e. analysis and inductive reasoning) are enhanced by training in hierarchical outlining.

Results also revealed that a positive disposition toward critical thinking was related to better critical thinking performance at post-testing (Ennis, 1987; Facione, 1990, 1992; Facione, Facione, Blohm & Giancarlo, 2000; Halpern, 2003, 2004). Notably, the correlation between pre-test CT performance and dispositions was not significant. This suggests that dispositions such as truth seeking, open mindedness, analyticity, systematicity, confidence, inquisitiveness or maturity may emerge as significant correlates of CT performance only after students have been exposed to some training in CT skills. However, we must interpret these findings with caution because we only measured dispositions at one point in time, that is, at post-test. Furthermore, researchers have identified problems with the measurement of dispositions, including the problematic nature of measuring CT dispositions using self-reports (Ku, 2009).

There were a number of limitations in this study. One limitation was the small sample size, which impacted on the power of our statistical analysis. It was difficult to persuade students to register for this extra-curricular CT training course, and although we managed to recruit a relatively large number of students, there was significant attrition from pre-test to post-test. Reasons for attrition included students having conflicting schedules, being too busy with other subjects, and other personal reasons. From the 129 students who initially completed the CCTST pre-test, only 81 completed the post-test. This reduced the power of our statistical analysis. Some borderline effects may well have been significant if our sample size was larger.

Another limitation of the current study was the randomization of participants to the control condition. From a pool of approximately 1,000 eligible students, it was hoped that roughly 300 would register for the course. Those who did sign up were to be randomly allocated to the AM group, the HO group, or the control condition. Initially, only 101 students signed up for the course. As a result, to ensure adequate statistical power in the comparison of the AM and HO conditions, those 101 students were randomly allocated to one of two conditions. Therefore, it was necessary to recruit an additional group of students for the control condition. Those who were assigned to the control condition were students who had expressed an interest in attending the CT course but who could not attend due to conflicting schedules. Thus, it may be that those who were recruited for the control condition may have been ultimately less motivated to take part and perform well on the CT tests, as many students in the experimental conditions actually rearranged their schedules in order to attend the CT course. Notably, participants in the control condition performed less well than participants in the AM and HO groups on certain post-test CT skills and it is possible that these differences between the experimental groups and the control group may have been a result of differences in motivation. Although we measured disposition toward critical thinking and found a correlation between CCTST and CCTDI performance on post-test, the CCTDI does not provide us with a direct measure of student motivation to perform well on the CCTST. A measure of students' motivation would have been useful in the current study, and it could have been included as a second covariate in the analysis of experimental condition of CT outcomes.

Another potential limitation of the study is the lack of feedback provided to students during the course. In a meta-analysis by Marzano (1998), it was found that by providing feedback to students on the type of strategy they used and how well they were using it to improve a specific type of cognitive process, students showed a significant gain in achievement, with an effect size of 1.13. Provision of feedback could potentially have also motivated students and curbed attrition in our study. Another challenge in the current study was the selection of a critical thinking test that allowed for adequate measurement of gains associated with CT training. Though the CCTST measures CT and the CT sub-skills according to the Delphi definition and framework, the test itself is not necessarily ideal for evaluating gain in intervention studies. For example, according to research by Jacobs (1995), the CCTST Forms A and B are dissimilar in that they possess different levels of difficulty. As a result, Jacobs has recommended that these tests are not used for purposes of measuring individual differences or gains from pre- to post-testing; and instead, one form should be used a covariate measure. In accordance with Jacobs' research findings, we used Form B as a covariate and baseline measure of CT and the Form 2000 of the CCTST as the outcome measure.

Further limitations of the CCTST are apparent when one examines the format of the test. Though there are 34 items which measure analysis, evaluation, inference, inductive and deductive reasoning skills, performance is assessed via multiple choice questions (MCQs). More specifically, the CCTST and many other MCQ tests of critical thinking have been criticised for being basically tests of verbal and quantitative knowledge (Halpern, 2003), since the test-takers are not free to determine their own evaluative criteria nor generate their own solutions to the problem (Ku, 2009). The measurement of critical thinking through MCQs is also problematic given the potential incompatibility between conceptualisations of critical thinking and its assessment using MCQs. MCQ tests assess cognitive capacities associated with identifying single right- and- wrong answers in relation to CT problems, and this approach to testing is unable to provide a direct measure of test-takers use of metacognitive processes such as reflective judgment (Halpern, 2003; Ku, 2009).

One solution to this problem would be to use a critical thinking assessment that asks open-ended style questions, which allow for test-takers to demonstrate whether or not they spontaneously use a specific critical thinking skill. However, the only commonly used critical thinking assessment that uses an open-ended format is the Ennis-Weir Critical Thinking Essay Test (Ennis & Weir, 1985), which has been criticised for its domain-specific nature (Taube, 1997), the subjective nature in which the tests are scored and potential biases in favour of test-takers who are more proficient in writing (Adams, Whitlow, Stover & Johnson, 1996). In short, both MCQ and open-ended test formats for assessing critical thinking have their respective limitations. The current trend is to combine the two response formats into one test (Ku, 2009).

One final issue to consider in relation to this study is what amounts to a sufficient amount of argument mapping training. One possible reason why AM did not emerge as a better training method than HO is that, because AM is a relatively novel method that students must first master before it becomes useful in promoting CT, more AM training is needed relative to other CT training methods. Based on van Gelder, Bissett and Cumming's (2004) finding that 'deliberate practice' in argument mapping facilitates growth of CT skills, we have to question whether or not sufficient practice in AM was provided in the current study. Van Gelder and colleagues recommend a semester long course in AM-infused CT training. Although the course in this study was only eight weeks, our study suggests that those who attended CT training scored higher on some CT skills at post-testing when compared with those in the control condition. The problem is that these effects were not very large or significant, and we have to question whether or not this is due to the small sample size or the intensity of the training provided. One improvement on the current design would be to include more argument map practice *outside* the classroom. We initially sought to control for the level of practice in both AM and HO conditions by restricting work to class-time only. While this was done with the good intention of controlling for potential confounds in the comparison between AM and HO training, it may have had a negative impact on the overall efficacy of the course. Therefore, extended training in argument mapping through CT, both inside and outside the classroom, is recommended in future studies.

Further research is needed in order to discover the conditions that most positively affect CT skill development, including research into the effects of extended training in argument mapping. Further research should also measure and control for students' motivation level, as the lack of motivation of students is one possible cause of the high attrition rate in this study. In addition, future research should also use measures of CT ability that allow for an assessment of the meta-cognitive abilities of students. For example, open-ended short answer questions that allow for both quantitatively and qualitatively scoring rubrics would be ideal, as they require test-taker to truly consider and evaluate all possible solutions and alternatives and construct an argument in response to a probe question, rather than simply choosing the correct answer on an MCQ.

Furthermore, according to the survey taken by students who participated in our study, it was recommended that future research should also aim to assess student ability throughout the course and provide students with feedback throughout. Students also suggested that identification of the student's initial critical thinking strengths and weaknesses, and how they are improving over time, may act as an incentive to continued participation and engagement throughout the course. Though incentives were provided to students (i.e. research participation credit and two laptops to be awarded to those who showed best effort), it was also suggested by a handful of students that perhaps too many material incentives and rewards may not be a good thing, as any sort of performance-based feedback could potentially be rewarding enough.

It was further suggested that making the course available online would be an ideal method of increasing participation as it avoids the issue of timetable clashes with other classes, as students would be able to take the course whenever it suited them. It was also suggested that argument mapping software be provided to students outside the class setting (also suggested by van Gelder, Bissett & Cumming [2004]). In the current study, we did not train students in the use of argument mapping software per se, but rather in the method of argument mapping using paper and pencil materials and exercises provided in class.

In summary, thinking is an important aspect of the human experience and consciously thinking about thinking is necessary in academic settings so that students may assess what they have to learn and what they already know. Students often encounter arguments in academic text-books that are difficult to analyse and evaluate due to the way in which the arguments are presented in text format. In order to promote good CT, educators must help students to lessen the cognitive load associated with reading and assimilating text. Argument mapping is a learning tool which may help students in this context. Results from this study suggest that AM training may increase specific CT sub-skills, such as evaluation and inductive reasoning. However, the observed effects were weak and methodological problems in the current study prevent us from drawing any strong conclusions in relation to the value of AM training.

REFERENCES

Abrami, P.C., Bernard, R.M., Borokhovski, E., Wade, A., Surkes, M.A., Tamim, R. & Zhang, D., (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research, 78,* 4, 1102-1134.

Adams, M.H., Whitlow, J.F., Stover, L.M. and Johnson, K.W., (1996). Critical thinking as an educational outcome: an evaluation of current tools of measurement. *Nurse Educator,* 21, 23–32.

Allegretti, C.L. and Frederick, J.N. (1995). A model for thinking critically about ethical issues. *Teaching of Psychology*, Vol. 22, 1, 46-48.

Allen, R., Feezel, J., & Kauffeld, F. (1967). A taxonomy of concepts and critical abilities related to the evaluation of verbal arguments (Occasional Paper No. 9). Madison: Wisconsin Research and Development, Center for Cognitive

Learning.

Alvarez-Ortiz, C. (2007). *Does Philosophy Improve Critical Thinking Skills?* Unpublished thesis.

Anderson, L.W. & Krathwohl, D.R. (eds.) (2001). *A Taxonomy for Learning, Teaching and Assessing: A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.

Association of American Colleges & Universities, (2005). Liberal education outcomes: A preliminary report on student achievement in college. Washington, DC.

Australian Council for Educational Research, (2002). *Graduate skills assessment.* Commonwealth of Australia.

Beckie, Theresa, M., Lowry, Lois, W. & Barnett, Scott (2001). Assessing critical thinking in Baccalaureate nursing students: A longitudinal study. Holistic Nursing Practice, 15(3), 18-26.

Bensley, D.A. (1998). *Critical thinking in psychology: A unified skills approach.* Pacific Grove, CA: Brooks & Cole.

Bloom, B.S., (1956). *Taxonomy of educational objectives: The classification of educational goals. Handbook 1: Cognitive domain.* New York: McKay.

Butchart, S., Bigelow, J., Oppy, G., Korb, K., & Gold, I. (2009). Improving critical thinking using web-based argument mapping exercises with automated feedback. *Australasian Journal of Educational Technology, 25,* 2, 268-291.

Dewey, J. (1933). *How we think: A restatement of the relation of reflective thinking to the educative process*. Lexington, MA: Heath.

Dwyer, C. P., Hogan, M.J. & I. Stewart, (2010). The evaluation of argument mapping as a learning tool: Comparing the effects of map reading versus text reading on comprehension and recall of arguments. *Thinking Skills and Creativity, 5,* 1, 16-22.

El Hassan, K. and Madhum, G., (2007). Validating the Watson Glaser Critical Thinking Appraisal. Higher Education, 54, 361-383.

Ennis, R.H., (1987). A taxonomoy of critical thinking dispositions and abilities. In *Teaching Thinking Skills: Theory and Practice,* eds. J.B. Baron, and R.J. Sternberg. New York: W.H. Freeman, 9-26.

Ennis, R.H., (1989). Critical thinking and subject specificity: Clarification and needed research. *Educational Researcher, 18,* 3, 4-10.

Ennis, R.H., Millman, J. & T.N. Tomko (1985). Cornell Critical Thinking Tests CA: Critical Thinking Co.

Ennis, R.H. & Weir, E., (1985). The Ennis-Weir Critical Thinking Essay Test. Pacific Grove, CA: Midwest Publications.

Facione, N.E. (1990). The Delphi report. Committee on pre-college philosophy. *American Philosophical Association*.

Facione, P.A., (1991). The California Critical Thinking Skills Test: College Level Technical Report #1 - Experimental Validation and Content Validity. *Unpublished.*

Facione, P.C. & Facione, N.A. (1992). *California Critical Thinking Skills Test*. California Academic Press, California.

Facione, P.A., Facione, N. C., Blohm, S. W., & Giancarlo, C. A. (2002). *The California Critical Thinking Skills Test: CCTST. Form A, Form B, and Form 2000*. Test manual, 2002 updated edition. Millbrae, CA: Insight Assessment.

Gabennesch, H. (2006). Critical thinking…what is it good for? (In fact, what is it?). *Skeptical Inquirer,* Vol. 30, 2, pp. 36-41.

Glaser, E. M. (1941). *An experiment on the development of critical thinking. Teachers College Contributions to Education, No. 843*. New York: Bureau of

Publications, Teachers College, Columbia University.

Halpern, D.F. (2003). *Thought & Knowledge: An Introduction to Critical Thinking (4th Ed.)*. New Jersey: Laurence Erlbaum Associates.

Halpern, D.F. (2004). A cognitive taxonomy for sex differences in cognitive abilities. *Current Directions in Psychological Science,* 13, 135–139.

Harrell, M., (2004). The improvement of critical thinking skills. In *What Philosophy Is* (Tech. Rep. CMU-PHIL-158). Carnegie Mellon University, Department of Philosophy.

Harrell, M., (2005). Using argument diagramming software in the classroom. *Teaching Philosophy,* 28, 2,

Higher Education Quality Council, Quality Enhancement Group, (1996). *What are graduates? Clarifying the attributes of "graduateness"*. London: HEQC.

Holmes, J. and Clizbe, E., (1997). Facing the 21st century. *Business Education Forum,* 52, 1, 33-35.

Jacobs, S. S. (1995). Technical characteristics and some correlates of the California Critical Thinking Skills Test Forms A and B. *Higher Education Research, 36*, 89-108.

Kintsch, W. & van Dijk, T.A. (1978). Toward a model of text comprehension and production. *Psychological Review, 85*, 363-394.

Ku, K.Y.L., (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking Skills and Creativity,* 4, 1, 70- 76.

Kuhn, D. (1991). The skills of argument. Cambridge: Cambridge University Press.

Kurfiss, J.G., (1988) Critical thinking: Theory, research, practice and possibilities. *Association for the Study of Higher Education, Report No. 2*, Washington D.C.

Marzano, R.J., (1998). A theory-based meta-analysis of research on instruction. Aurora, CO: Mid-Continent Regional Educational Laboratory.

http://www.mcrel.org/pdf/instruction/5982rr_instructionmeta_analysis.pdf, 26, October, 2007.

MacPherson, K., (1999). The development of critical thinking skills in undergraduate supervisory management unites: Efficacy of student peer assessment. *Assessment & Evaluation in Higher Education, 24,* 3, 273-284.

Monk, P. (2001). Mapping the future of argument. *Australian Financial Review.* 16, March, pp.8-9.

Moseley, D., Baumfield, V., Elliot, J., Gregson, M., Higgins, S., Miller, J. & Newton, D., (2005). *Frameworks for thinking: A handbook for teaching and learning.* Cambridge: Cambridge University Press.

National Academy of Sciences, National Academy of Engineering, Institute of Medicine (2005). Rising above the gathering storm: Energising and employing America for a brighter economic future. *Committee on Prospering in the Global Economy for the 21st Century,* Washington, DC.

Paul, R.W. (1993). Critical Thinking – what every person needs to survive in a rapidly changing world (3rd ed.). Santa Rosa, CA: Foundation for Critical Thinking.

Quitadamo, I.J. and Kurtz, M.J., (2007). Learning to improve: Using writing to increase critical thinking performance in general education biology. *CBE –Life Sciences Education,* 6, 140-154.

Sweller, J. (1988). Cognitive load during problem solving: Effects on learning. *Cognitive Science*, *12*, 257-285.

Sweller, J. (1999). Instructional design in technical areas. *Australian Education Review No. 43.* Victoria: Acer Press.

Taube, K.T., (1997). Critical thinking ability and disposition as factors of performance on a written critical thinking test. *Journal of General Education, 46,* 129-164.

Thomson, A., (2009). *Critical Reasoning: A practical introduction (3rd ed.).* UK: Routledge.

Tindall-Ford, S; Chandler, P. & Sweller, J., (1997). When two sensory modes are
better than one. *Journal of Experimental Psychology: Applied,* Vol. 3, No. 4: 257 -287.

U.S. Department of Education, (1990). National goals for education. Washington, DC.

van Gelder, T.J. (2000). Learning to reason: A *Reason-Able!* approach. In C. Davis, T. J. van Gelder, & R. Wales (eds.), *Cognitive Science in Australia, 2000:Proceedings of the Fifth Australasian Cognitive Science Society Conference.* Adelaide: Causal.

van Gelder, T. J. (2001). How to improve critical thinking using educational technology. In G. Kennedy, M. Keppell, C. McNaught, & T. Petrovic (Eds.), *Meeting at the Crossroads: Proceedings of the 18th Annual Conference of the Australian Society for Computers In Learning In Tertiary Education,* 539–

548. Melbourne: Biomedical Multimedia Unit, Ùniversity of Melbourne.

van Gelder, T.J., (2003). Enhancing deliberation through computer supported

argument mapping. In Kirschner, P., Buckingham Shum, S., Carr, C. (eds.), *Visualizing Argumentation: Software Tools for Collaborative and Educational Sense-Making,* 97–115, London: Springer-Verlag.

van Gelder, T.J., Bissett, M., & Cumming, G. (2004). Enhancing expertise in informal reasoning. *Canadian Journal of Experimental Psychology 58*, 142-52.

van Gelder, T.J. & Rizzo, A. (2001). Reason!Able across curriculum, in *Is IT an Odyssey in Learning? Proceedings of the 2001 Conference of ICT in*

Education, Victoria, Australia.

Watson, G. & E.M. Glaser (1980). *Critical thinking appraisal*. New York:

Psychological Corporation.

Wilkinson J.M., (1996). *Nursing process: A critical thinking approach* (2nd ed.).

Menlo Park: Addison Wesley Nursing.

Willingham, D.T., (2007). Critical thinking: Why is it so hard to teach? *American Educator, 2007, 3,* 8-19.