



**Systematic review report:
A systematic review of effective methods and strategies for
improving argumentation skills in undergraduate students in
higher education**

University of York

Carole Torgerson

Richard Andrews

Alison Robinson

Beng Huat See

October 2006

Contents	
	Page
Acknowledgements	3
List of Tables	4
Abstract	5
Glossary	6
Executive summary	7
Background	13
Aims	16
Research question	16
Methods	17
Results	21
Conclusions	45
Recommendations	47
References	48
Appendices	51

Acknowledgements

The authors of the report would like to acknowledge the support of the Higher Education Academy in funding this systematic review of strategies and methods for improving argumentation skills in undergraduates in Higher Education. They would also like to acknowledge Sally Mitchell and Paul Prior for their comments on the protocol for the review.

This report should be cited as:

Torgerson, C.J., Andrews, R.J., Robinson, A.M., and See, B.H. (2006) A systematic review of effective methods and strategies for improving argumentation skills in undergraduate students in Higher Education, York: The Higher Education Academy.

© the authors, 2006

Though the report is self-standing, it should be read alongside a pilot empirical study undertaken as part of the same funded Higher Education Academy research project in 2005/06, the reference for which is:

Andrews, R.J., Bilbro, R., Mitchell, S., Peake, K., Prior, P., Robinson, A.M., See, B.H. and Torgerson, C. (2006) Argumentative skills in undergraduate students in Higher Education: a pilot study, York: The Higher Education Academy.

List of tables

Table	
	Page
Table 1: Origin of studies included in review	21
Table 2: Characteristics of studies included in the descriptive map	23
Table 3: Quality appraisal of four studies included in the in-depth review	35

Abstract

Aims and methods: The aims of the systematic review were: to systematically search for, locate, quality appraise and synthesize all the empirical studies employing an experimental, quasi-experimental or case control design, which explored the effectiveness of methods and strategies for improving argumentation skills in undergraduate students in Higher Education settings. We searched the electronic databases PsycINFO, ERIC, SSCI, BEI, and C2-SPECTR for the period 1990 to 2005. We screened for inclusion using pre-established inclusion criteria, and independently double data extracted and quality appraised all the studies which met the inclusion criteria for the overview and in-depth review.

Results: We included 10 studies in the overview, and four studies in the in-depth review. The studies varied in quality and were not homogeneous in terms of strategies and methods evaluated. The most important finding from this systematic review was that, despite the well-developed theoretical literature in argumentation, there is a lack of retrievable large well-designed, well-reported randomized controlled trials or controlled trials in the field. A number of interventions look promising in helping to improve argumentation skills: training in critical thinking; the use of constraint-based argumentation scaffolds and goal instruction.

Conclusions and recommendations: Although the studies in the in-depth review found that some strategies and methods evaluated using controlled research designs were effective in improving argumentation skills in undergraduate students, strong substantive recommendations for practice cannot be made. A large rigorously designed, conducted and reported randomized controlled trial in a UK HE setting is recommended. This could investigate the effectiveness of one or more of the strategies identified in this systematic review as looking promising for the improvement of argumentation skills in undergraduate students.

Glossary

Case control study: Comparison groups are formed by matching, after the intervention has been delivered.

Cluster randomized trial: In trials using this form of design the unit of allocation includes more than one individual, e.g. class, school.

CONSORT: Consolidated Standards for Reporting Trials is the methodological standard adopted by many medical journals for publication of randomized controlled trials.

Controlled trial (CT): This usually means a study with a control group that has been formed by means other than randomization. Consequently the validity of the study using this design is potentially threatened by selection bias.

Heterogeneity: When studies have different characteristics, e.g. different populations or different outcome measures.

Homogeneity: When studies have similar characteristics, e.g. similar populations or similar outcome measures.

Randomized Controlled Trial (RCT): This is where two or more groups have been formed through random allocation (or a similar method). This is the only method that ensures that selection bias is eliminated at baseline.

Systematic review: A review where explicit methods have been used to identify, select and include studies fitting pre-specified criteria.

Type II error: Concluding incorrectly that there is no educational significant difference between the two groups.

Executive summary

Background

In their overview of students' reading and writing, Langer and Applebee (1987) highlighted the difficulties that students often have with writing argumentation. Mitchell and Andrews (2001) painted a similar picture, though with a different analytical perspective, based on research projects on argumentation undertaken in the 1990s in the UK in a range of disciplines. There have been a number of other studies since the early 1980s that have indicated that there is a problem with the teaching and learning of argument at undergraduate levels, as well as indications that 'argument' itself may be part of the problem. Despite considerable attention to argument theory, research, and especially pedagogy, it is apparent that much remains to be done in articulating effective tools and strategies for argument at higher education level. It is therefore important to systematically search for and locate all the relevant studies that can inform us about effective strategies and methods for improving argumentation skills in HE, so that we can see what has been attempted in the field and what has succeeded (or not). This is the purpose of this systematic review.

The present systematic review should be read alongside Andrews et al. (2006), which is a pilot empirical study designed to investigate the feasibility of further research in the field. The systematic review depicts what has been attempted in terms of the *improvement* of argumentation skills in undergraduates. The pilot empirical study examines the complexities of argument and how best to research them in three disciplines at undergraduate level: biology, history and electrical engineering.

Aims

The aims of the systematic review were to search for, locate, quality appraise and synthesize all the empirical studies employing an experimental, quasi-experimental or case control design (published or reported between the years 1990 and 2005) that explored the effectiveness of methods and strategies for improving argumentation skills in undergraduate students in Higher Education.

Research question

The research question for the systematic review was:

What can experimental studies tell us about strategies and methods that are effective in improving argumentation skills in undergraduate students in Higher Education?

Methods

Systematic review methods, as outlined in Torgerson (2003), were used throughout the conduct of the review, in order to limit bias, and to enable replication of the review. A protocol was developed which outlined the methods for the review, including methods for identifying studies for inclusion through electronic searching and screening (with inclusion and exclusion criteria established *a priori*).

The systematic review involved an overview of all the experimental research in the field, meeting all the inclusion criteria, and an in-depth review of studies selected on the basis of study design and reporting criteria and matching of interventions and outcomes to the research question.

Rigorous quality assurance procedures were put in place for the screening of titles and abstracts and of full papers, and for the data extraction and quality appraisal of the included studies.

The synthesis took the form of a qualitative overview: a qualitative summary of the main findings of all the trials included in the mapping section in terms of methods and strategies for improving argumentation skills, and a more detailed description of the studies included in the in-depth review, including a description of the methodological strengths and weaknesses of the studies.

Results

A total of 2353 studies were identified as being potentially relevant through the electronic searches (AR). 2287 of these were excluded at first stage, 38 were excluded at second stage and 18 were excluded at third stage, leaving a total of 10 studies to be included in the overview.

Nine of the ten studies were undertaken in the USA, with only one study having been undertaken in the UK (Joiner and Jones, 2003). There were seven randomized controlled trials (three individual and four cluster), 1 controlled trial and two case control studies.

Six studies were excluded from the in-depth review. Four studies contained no raw data (Gass *et al*, 1990; Kozma, 1991; Larson *et al*, 2004; Joiner and Jones, 2003), and two studies were excluded due to the limited nature of the intervention evaluated (Nussbaum *et al*, 2004; Steinkuehler, 2000). This left four studies to be included in the in-depth review (Bensley and Haynes, 1995; Cho and Jonassen, 2002; Nussbaum, 2005; Sanders *et al*, 1994). These studies fulfilled three inclusion criteria: they all included sufficient raw data for replication or calculation of effect sizes; they all evaluated an intervention of a reasonable length; and they all evaluated argumentation interventions using argumentation outcomes. The four studies were quality appraised using a modified version of the CONSORT statement (Begg *et al*).

Conclusions

The most important finding from this systematic review was that, despite the well-developed theoretical literature in argumentation, there is a lack of retrievable large well-designed, well-reported randomized controlled trials or controlled trials in the field. Specifically, in the period 1990 to the present there are only a few studies with comparison groups that evaluated methods and strategies to improve argumentation in undergraduate students in HE. The lack of RCTs or CTs undertaken in the UK is particularly notable.

Substantive findings

Some substantive conclusions can be tentatively drawn from the studies in the in-depth review:

- Training in critical thinking looks promising both for the conventionalization of the language of argumentation in writing, and for the greater use and appropriate positioning of argumentation contents in writing (Bensley and Haynes, 1995).
- A constraint-based argumentation scaffold can improve both group argumentation and individual argumentation (Cho and Jonassen, 2002).
- The use of goal instruction to (a) persuade and (b) generate reasons in argumentation in an interactive context can improve students' argumentation by increasing argumentation claims, contingent claims, opposition and debate (Nussbaum, 2005).

- Argumentation training can enhance specific critical thinking skills: ability to discern weak arguments, improved self-perception of argumentation effectiveness and decreased verbal aggressiveness in argumentation encounters (Sanders *et al*, 1994).

Recommendations

Substantive recommendations

Although the studies in the in-depth review concluded that some strategies and methods evaluated using controlled research designs were effective in improving argumentation skills in undergraduate students, strong substantive recommendations for practice cannot be made. We suggest that such interventions look promising but should be adopted with caution, given the variable quality of the studies involved and the fact that each strategy or method was evaluated in only one study.

Recommendations for further research

A large rigorously designed, conducted and reported randomized controlled trial in a UK HE setting is recommended. This could investigate the effectiveness of one or more of the strategies identified in this systematic review as looking promising for the improvement of argumentation skills in undergraduate students. We would see such a trial as forming part of the larger empirical study recommended in Andrews *et al*. (2006)

Background

Traditions of rhetorical instruction and understanding in Europe date back to ancient Greece and beyond. Aristotle's *Rhetoric*, Cicero's *On Oratory* and the *Ad Herennium* discuss and set out guidance and advice, primarily to speakers, in the social and political contexts in which they wrote. Crucially, the ability to argue in the public and private domains was linked with democratic processes. The rhetorical tradition (as described by Vickers (1988) in *In Defence of Rhetoric*) remained largely unbroken – though changed in many ways - through the medieval period and into the 19th century. The emergence of *belles lettres* studies and subsequently 'English Literature', with their focus on the fictional, brought about a crisis in rhetorical studies and practice. Whereas the Scottish tradition (Bain, 1866) exported its commitment to rhetoric to the USA – where it has established a strong place in undergraduate induction in higher education, in Rhetoric and Composition or Freshman Composition programmes – in England the rhetorical tradition and Classical Studies were supplanted by English Literature as the central humanizing discipline. By the 1920s, with the publication of Sampson's influential report, *English for the English*, the demise of rhetoric was virtually complete in school and university education in England.

Twentieth century revivals of an interest in rhetoric and argumentation have taken place via the work of Burke in the USA, Toulmin (especially as taken up in composition and speaking courses in the US), Van Eemeren/Grootendorst in Russia/Europe, and Eagleton in the UK. Most of the applied work in Higher Education, however, has taken place in the States.

In the US, textbooks for first-year college composition (e.g., *Everything's an Argument*, Lunsford, Roszkiewicz & Waters; *The Craft of Argument*, Williams & Colomb) routinely foreground argument, usually a mix of neo-Aristotelian concepts (logos/pathos/ethos; topoi; stases) and the Toulmin model. Nevertheless, work on argument (e.g., *Argument Revisited*; *Argument Redefined*, edited by Emmel, Resch, & Tenney; Lunsford's 2002 *Written Communication* report on situated uptake of argument instruction) has highlighted the theoretical and pedagogical challenges facing instructors. Research on graduate and undergraduate writing across the curriculum (e.g., Casanave; Herrington; Prior) has highlighted the complex, often tacit nature of disciplinary argumentation.

In their overview of students' reading and writing, Langer and Applebee (1987) highlighted the difficulties that students often have with writing argumentation. Mitchell and Andrews (2001) painted a similar picture, though with a different analytical perspective, based on research projects on argumentation undertaken in the 1990s in a range of disciplines. There have been a number of other studies since the early 1980s that have indicated that there is a problem with the teaching and learning of argument at undergraduate levels, as well as indications that 'argument' itself may be part of the problem. Despite considerable attention to argument theory, research, and especially pedagogy, it is apparent that much remains to be done in articulating effective tools and strategies for argument at higher education level. It is therefore important to systematically search for and locate all the relevant studies that can inform us about effective strategies and methods for improving argumentation skills in HE, so that we can see what has been attempted in the field and what has succeeded (or not). This is the purpose of this systematic review.

The present systematic review should be read alongside Andrews et al. (2006), which is a pilot empirical study designed to investigate the feasibility of further research in the field. The systematic review depicts what has been attempted in terms of the *improvement* of argumentation skills in undergraduates. The pilot empirical study examines the

complexities of argument and how best to research them in three disciplines at undergraduate level: biology, history and electrical engineering.

Aims

The aims of the systematic review were to search for, locate, quality appraise and synthesize all the empirical studies employing an experimental or quasi-experimental design (published or reported between the years 1990 and 2005) that explored the effectiveness of methods and strategies for improving argumentation skills in undergraduate students in Higher Education.

Research question

The research question for the systematic review was:

What can experimental studies tell us about strategies and methods that are effective in improving argumentation skills in undergraduate students in Higher Education?

Methods

Systematic review methods, as outlined in Torgerson (2003), were used throughout the conduct of the review, in order to limit bias, and to enable replication of the review. A protocol was developed in order to establish: the research question; the scope and limitations of the review; the methods for conducting the review; the inclusion and exclusion criteria; and the procedures for extraction of data and quality appraisal. The protocol is available by writing to the first author.

Identification of studies

In collaboration with two of the members of the review team (CT and RA), an Information Officer (AR) wrote the search strategies for the electronic databases (see Appendix A) and carried out electronic searches for the review on PsycINFO, ERIC (Educational Resources Information Center), SSCI (Social Sciences Citation Index), BEI (British Education Index), C2-SPECTR (The Campbell Collaboration's Social, Psychological, Educational and Criminological Trials Register). Search terms used included: argumentation, writing and argument, persuasive writing, persuasive communication, persuasive discourse, rhetoric and composition, and expository writing. On completion of the searches, the citations identified were imported into the reference manager EndNote and de-duplicated.

Criteria for including and excluding studies

The systematic review involved an overview of all the experimental research in the field and an in-depth review of the studies meeting more rigorous inclusion criteria.

Papers included in the overview had to be evaluations of interventions aimed at improving argumentation skills in study populations of students in HE. Papers were included if they were one of the following study types: randomized controlled trial (RCT), controlled trial (CT) or review of RCTs and/or CTs, case control studies. Studies of this type were included on the basis that randomized trials, controlled trials and case control studies can be used to help to answer questions about the effectiveness of interventions designed to improve undergraduate argumentation skills. In addition, studies were included if they were undertaken in English speaking countries and were written in the English language. Finally, papers were included if they were published or reported in the years 1990 - 2005. Studies were excluded on the basis of study type if they were interventions of a pre- and post-test design, or if they were non-interventions. Finally, studies were excluded if all or some of the participants were aged younger than 18 years and/or were not undergraduates studying in a HE institution (see Appendix B for details of inclusion and exclusion criteria for screening at first, second and third stages).

Papers were included in the in-depth review if they included sufficient raw data to allow replication of the analyses, if they evaluated interventions of a reasonable length, and if both intervention and outcomes were specifically focused on argumentation (see Appendix C for details of inclusion and exclusion criteria for the in-depth review).

Screening and quality assurance

The titles and abstracts of studies identified through the searches were screened for inclusion in the review by the first author (CT). Double screening of a random sample of studies identified by the electronic searches was undertaken by RA, using the inclusion/exclusion criteria outlined above, in order to quality assure the screening undertaken by the first author. A measure of agreement between the reviewers was calculated using the Cohen's Kappa. Papers identified as being potentially relevant were

obtained either through electronic resources or library inter-lending. The retrieval rate for identified papers was 100%.

Exclusion at second stage

All received papers were double screened on the basis of the full papers by two members of the review team working independently (CT and RA). Any disagreements between reviewers were discussed and resolved, and papers were included or excluded at second stage by agreement between the two reviewers.

Exclusion at third stage

In a small number of cases, studies were excluded during the data extraction process.

Coding

All included papers were coded (CT) using one of the following codes: 'RCT', 'CT', 'case control', 'review'. Although we searched and screened for reviews, none undertaken in the period of our review were included.

Double data extraction and quality appraisal, quality assurance

All included trials were double data extracted and those containing sufficient data for replication were quality appraised by teams of two reviewers (CT and RA; CT and BHS) working independently. We used pre-established standardized data extraction tables and quality appraisal tables based on the Consolidated Standards for Reporting Trials (CONSORT) guidelines (Begg *et al*, 1999). The two reviewers extracted data on the development and content of the intervention evaluated, the design and results of the trial, and data on the methodological quality of the RCT.

For each paper, data were extracted about the identification and aims of the study; study design and content; use of allocation (random or otherwise) to the different groups; participants; pre- and post-intervention data; attrition rate. Full agreement was established through discussion.

Synthesis

The synthesis took the form of a qualitative overview: a qualitative description of the main findings of the trials in terms of methods and strategies for improving argumentation skills; and a description of the methodological strengths and weaknesses of the trials. It was not appropriate to conduct a meta-analysis to synthesize data from two or more studies because there were insufficient trials of high quality that were homogeneous in terms of intervention and setting. It was also not possible to investigate the presence of publication bias in our review, due to the very small number of studies included in the in-depth review (four).

Results

Studies retrieved through searching and screening

A total of 2353 studies were identified as being potentially relevant through the electronic searches (AR). 2287 of these were excluded at first stage, 38 were excluded at second stage and 18 were excluded at third stage, leaving a total of **10** studies to be included in the overview (see Table 1).

Table 1: Origin of studies included in the overview

	Number identified by searches	Excluded at 1 st stage	Excluded at 2 nd stage	Excluded at 3 rd stage	Included in review
ERIC	1734	1703	13	11	7
PsycINFO	455	429	20	4	2
SSCI	105	100	3	2	-
BEI	56	54	1	-	1
C2SPECTR	2	1	1	-	-
Citation	1	-	-	1	-
Total	2353	2287	38	18	10

(Studies were de-duplicated sequentially in the database order shown above.)

Overview

In order to give an overview of the experimental research in the field of argumentation this section describes all ten studies.

Nine of the ten studies were undertaken in the USA, with only one study having been undertaken in the UK (Joiner and Jones, 2003). There were seven randomized controlled trials (three individual and four cluster), 1 controlled trial and two case control studies. Table 2 gives the main characteristics of the ten studies, including a brief indication of the interventions evaluated and the results and conclusions, as stated by the author(s) of the studies, and reviewer comments on the strengths and limitations in the design and reporting of the studies.

Table 2: Characteristics of studies included in the overview

Author, date, country, publication status	Aim; Study design	Intervention(s) ; No. of participants	Control(s)	Outcome measure(s)	Results (as stated by authors)	Conclusions (as stated by authors)	Reviewer comments
Bensley and Haynes, 1995, USA, pub.	To test a programme designed to develop critical reading and writing skills in psychology students through the acquisition of a general knowledge structure for argumentation. Controlled Trial (CT)	Critical thinking skills instruction (details given); n = 16	No critical thinking skills instruction; n = 10	Outline generation task in which students were asked to produce a generic outline for a paper in which they were to argue persuasively; they were asked to identify claims and list the evidence for one side of the argument separately from the other side and classify the evidence as to type and draw an appropriate conclusion and identify incorrect assumptions or other problems in the passage all from a critical reading passage.	The critical thinking skills instruction group showed greater use of language similar to that used in instruction and significantly greater use of argumentative language than the group not getting the instruction. For more detailed results see Appendix D.	The results supported the three hypotheses. Critical thinking participants used appropriate argumentative language in the correct order in their outlines.	Small sample size. Possibility of selection bias (how two groups formed NS). Some details about critical thinking in argumentation intervention given (p. 42 and p.43). The authors acknowledge that alternative interpretations of their data are possible. Appropriate analysis taking into account clustered nature of data.
Cho and Jonassen, 2002, USA, pub.	'To examine the use of online argumentation scaffolds to engage and support coherent argumentation' (p.5). Randomized Controlled Trial (RCT), cluster	Content-based scaffolding: Intervention i: well-structured problems and constraint-based tool Intervention ii: ill-structured problems and constraint-based tool	Control i: well-structured problems and threaded discussion Control ii: ill-structured problems and threaded discussion	Assessment of quantity of student arguments during group discussion using a coding scheme adapted from Toulmin's model (which identifies five major components of argument including claims, grounds, warrants, backings and rebuttals P.10). Assessment of problem-solving process during group discussion for their problem-solving function by classifying each message based on a coding system adapted from Poole and Holmes (1995) (which	'Providing a constraint-based argumentation scaffold during group problem-solving activities increased the generation of coherent arguments. The same scaffold further resulted in significantly more problem-solving actions during collaborative discussions' (p.5). The effects of the scaffold varied for problem type.	'The results of this study showed that using a constraint-based argumentation scaffold positively affected the ability of groups to collaboratively construct arguments in an on-line environment. Although only	

				<p>consists of seven categories including problem definition, orientation, criteria development, solution approval, solution critique and non-task statement (p.10).</p> <p>Problem-solving activity (an essay describing a solution and justification) to measure transfer of argumentation and problem-solving skills developed during group problem-solving activities using coding based on Toulmin's model of argument to determine the quality of argumentation. Each report was also analysed for problem-solving performance (p.11).</p>		<p>claims and grounds (evidence) were affected, these are two of the most important components in argument construction and also components in which students are weak.' (p.17)</p>	
<p>Gass et al, 1990, USA, unpub. Conference paper</p>	<p>To determine the extent to which training in critical thinking or argumentation skills combined with initial attitudes and type of warrant employed influenced the participants' evaluation of overall argument potency (p.2). Case control study</p>	<p>Critical thinking skills or argumentation skills instruction (no details given); n = 393</p>	<p>Communication instruction (no details given) (no instruction in critical thinking skills or argumentation skills instruction); n = 151</p>	<p>3 bipolar responses in 3 dimensions of argument strength (measured through responses to a questionnaire: logical; emotional; ethical and global measure of argument effect.</p>	<p>Perceptions of weak arguments: Logical: 'Course instruction in critical thinking improved the subjects' ability to detect logic weaknesses' (p.11), stat. sig. Argument effect: 'Students who had taken a beginning argumentation course were more accurate in detecting argument weaknesses' (p.11), stat. sig. Perceptions of strong arguments: Logical: 'Students who had taken</p>	<p>Critical thinking skills or argumentation skills instruction 'significantly alters how individuals assess arguments. Argumentation instruction enhances the students' ability to detect argument weaknesses while at the same time</p>	<p>No details given about critical thinking skills/argumentation skills instruction. No raw data given. Attrition not given. Possibility of selection bias in the composition of the two comparison groups not acknowledged. Excluded from in-depth review (Exclude 1).</p>

					<p>a critical thinking course perceived the logic of strong arguments as less compelling' (p.17), stat.sig.</p> <p>Ethical: 'Students who had instruction in critical thinking perceived less ethicality in the strong arguments than did students without critical thinking instruction' (p.17), stat. sig.</p> <p>For interaction effect see Appendix D.</p>	<p>causing the student to rely more on the logical dimension and less on the ethical and emotional dimensions in assessing argument strength' (pp.20-1).</p>	
<p>Joiner and Jones, 2003, UK, pub.</p>	<p>'...to compare the effects of different communication media on the quality of arguments and the development of argumentative reasoning' (p.861). RCT, cluster</p>	<p>On-line discussion: the 'on-line group' participated in a seminar discussion over a period of two weeks using Blackboard a commercial virtual learning environment, n = 39.</p>	<p>The 'face-to-face group' participated in a seminar discussion for one hour, n = 34.</p>	<p>Analysis of argument: Arguments in the students' mini essays in the pre-test and post-test were classified under 3 categories: Type I, Type II and Type III, using a framework by Kuhn (1991). Within each category, arguments were further classified according to whether they provided alternatives to CP, whether they offer reasons for or against CPs (Types 1A, B, C & D). Analysis of arguments was based on a scheme developed by Felton and Kuhn (2001). It comprised 3 broad categories: transactive questions, transactive statements and non-transactive statements. A transactive utterance is one that engages the partner in</p>	<p>No statistically significant differences were found between the two conditions in terms of improvement in the quality of students' arguments or in terms of change of opinion about corporal punishment. Argumentative reasoning was not facilitated by either face-to-face or computer-mediated discussion. Participants in the face-to-face condition were more likely to make transactive questions (utterances that request a response from a partner), Table 1, and 'meta-?' transactive statements. The participants in the face-to-face condition made proportionally more 'add', 'agree', 'disagree' and</p>	<p>The quality of argumentation was significantly higher in the face-to-face discussions than the quality of argumentation in the computer-mediated discussions. Argumentative reasoning was not facilitated by either the intervention or the control condition (p.870). Possible reasons for these conclusions are</p>	<p>Limited nature of intervention (very short), fully acknowledged by authors. Exclude from in-depth review (Exclude 2).</p>

				discourse.	'refuse' transactive utterances, Table 3, and proportionally fewer 'unconnected non-transactive utterances. (Non-transactive utterances fail to connect to the partner's preceding utterance.) Finally, participants in the face-to-face condition showed twice as many strategic sequences as the participants in the CMC condition. Therefore, the quality of the argumentation used in face-to-face was higher than that used in the online discussions' (p.861).	short length of intervention/control and insufficiently sensitive outcome measure (possibility of ceiling effect).	
Kozma, 1991, USA, pub.	To examine the impact of computer-based tools and embedded and topical and rhetorical prompts on college writers; specifically, to 'examine the interaction between the cognitive skills of learners and two software packages designed to support these components (higher level structural considerations and rhetorical planning) of the writing process' (p.4) RCT, individual	Computer-based writing software: Intervention i: practice assignment (argumentative writing) using MacWrite plus Acta (n = 13) with or without rhetorical prompts Intervention ii: practice assignment (argumentative writing) using	Practice assignment using MacWrite (n = 13) with or without rhetorical prompts.	Post-tests: measures of cognitive strategies used during the writing process using think-aloud protocol methodology and scored for the proportion of time that subjects engaged in conceptual planning (process planning, rhetorical planning and structure planning); argumentative tasks scored for overall quality, organisation, effectiveness of argument, and the extent to which the paper exhibited an awareness of an audience, using a 6-point primary-traits scale to rate the overall quality of a	Significant effects for both software and prompts. 'Writers did the least conceptual planning with the word processor (6.32%); more with the ideal organiser, Learning Tool (10.02%) and the most with the outliner, Acta (12.58%). Writers did more conceptual planning with embedded prompts (11.53%) than without them (8.08%)' (p.9). There was a significant interaction: 'the effects of software and prompts were compounded' (p.9). There was no interaction	The use of organizational software and embedded prompts increased the proportion of conceptual planning but this did not generally result in better compositions (p.20). The authors concluded that 'planning is necessary but insufficient for writing a good	Sample size small. Exclude from in-depth review due to lack of data (standard deviations) (Exclude 1); very short intervention – one practice composition (Exclude 2); and Exclude 3 on basis that although outcome measures were argumentative writing, interventions were peripheral to this.

		MacWrite plus Learning Tool (n = 15) with or without rhetorical prompts		composition, the effectiveness of an argument, and the extent to which the composition reflected an awareness of the intended audience.	between writing level and treatments: 'There were no more benefits for novice writers than there were for advanced writers' (p.10). However, 'compositions were no better when organisational software or rhetorical prompts were used' (p.11).	composition' (p.22). The use of graphic interfaces can facilitate the use of plans during composition for some writers (p.22). Finally, the authors argue for caution in the use of computer-based writing tools (p.23)	
Larson et al, 2004, Experiment 2, USA, pub.	To test the effectiveness of a tutorial to aid students in argument comprehension, specifically to aid evaluation of argument during reading (p.208). 'The effect of reading goal was also examined as a factor to improve readers' argument comprehension' (p.216). RCT, individual	Argument tutorial: the short tutorial 'defined key argument terms, challenged common misconceptions about arguments, and explained a series of steps to comprehend written arguments. Participants were provided an example and an opportunity to practice the presented	No feedback practice: participants received no instruction but received an opportunity to practice the task for the same amount of time. n = 35	Participants were given four texts (from Exp. 1 – see p.217 for details) and asked to identify the main claims and reasons. They had to write down the author's main claim and list any reasons mentioned to support the claim (p.219). 'The Comprehension participants were asked to list an unmentioned reason to support the author's main claim while Rebuttal participants were asked to list an unmentioned rebuttal to the author's main argument. After completing the packet, participants rated the strength of each of the author's reasons' (10-point scale from very weak to very strong with justification).	'there was a significant Training x Reading goal interaction' (p.219)	The short tutorial helped argument identification but only if participants focused on the single goal of comprehension . A rebuttal goal did not lead to a better argument	Insufficient raw data (numbers), therefore not possible for reviewer to re-calculate effect sizes. Not stated how many participants in each of the four arms of the experiment. Therefore exclude from in-depth review (Exclude 1 – insufficient data). Also length of intervention was very short (10 mins.).

		steps' (p.216). See pp.217-8 for a full description of the intervention. n = 35 In addition, there were two conditions for the Reading Goal factor: comprehension or rebuttal (numbers not given)					
Nussbaum et al, 2004, USA, pub.	To evaluate the effectiveness of a tool ('note starters') designed to scaffold and improve the level of reflection and argumentation in on-line discussions on students' counter argumentation, and the interaction of this tool with students' personality characteristics (p.114). RCT, individual	Scaffolding: 'Note starters': a menu of phrases, from which students choose a phrase to begin the first sentence of a discussion note. They are a form of scaffolding intended to encourage students to think more deeply through the construction of explanations	No 'note starters', with or without elaborated cases. 'Elaborated cases' are 'detailed scenarios about a classroom situation contained in the initial problem statement. Cases may make the problem situation more meaningful to students	On-line discussion in groups of four (each student posted four messages, one initial response to the question and three responses to the other group members' initial posting) and on-line survey (personality inventory) completed at the beginning of the study to assess individual students' personalities.	There was a higher percentage of disagreement in the note starter condition than in the control condition for both subcategories of disagreement as well as overall. Using computed average disagreement scores for each topic from 0 (consistent agreement) to 1 (consistent disagreement) there was a significant main effect for note starters (F = 4.6, (1,37) p<0.05). In addition note starters were shown to be most beneficial for students who had low levels of assertiveness, openness to ideas and anxiety. 'Note starters appeared to encourage students	'Note starters are a promising innovation for encouraging constructive forms of argument but – like many innovations – some individuals appear to benefit from them more than others' (131).	A small-scale study. The intervention is minimal: phrases used to stimulate counter-argument. The results are positive, but the size of the sample and the minimalism of the intervention suggest that claims are suitably modest, and that further research is called for both in note starter technique, between types of personality/learner, and within or without online discussion. Insufficient raw data given (means, standard deviations, numbers), therefore

		and arguments (p.116).	and, by increasing engagement, increase the number of problem-centred moves' (p.117).		who were not naturally curious and inquisitive (i.e. open to ideas) or assertive to consider opposing viewpoints. However, students who were anxious benefited less from note starters' (p.125).		exclude from in-depth review (Exclude 1). In addition intervention is too short (Exclude 2).
Nussbaum, 2005, USA, pub.	To examine the effect of goal instructions (making explicit the goals of an activity) on students' argumentative writing in an interactive discussion context (Web-based discussions); to explore the effect of the need for cognition (NFC) and to examine the interactions between cognition and goal instruction. RCT, cluster	All participants used the discussion board to discuss the following question: 'Does watching TV cause children to become more violent?' The question was posted on the notice board along with additional instructions. One person in the three-person group was randomly assigned to post the first note, after which the other two could post their notes.	This was provided by the 3x3 design in which the third group acted as a control with no interventions . The design was 3x3 factorial with general goal (Explore, Persuade, None) crossed with a specific goal (Reasons, Counterarguments, None).	Two surveys administered by the Web-CT (electronic discussion board): need for cognition and prior attitudes towards television violence. Both surveys on 5-point Likert scale with responses ranging from 1 (strongly disagree) to 5 (strongly agree). NFC form: 18 items. Prior attitudes towards TV violence: 12 items (p.295). A conceptual framework for analysing arguments was developed. 'In this framework, an argument consisted of a series of claims that bear logical relations to one another, specifically relations of support or opposition. The framework was designed to apply to interactive arguments, where one student may initially make a claim that frames the discussion and other students (as well as the first student) provide additional claims in response to the	'The goals to (a) persuade and (b) generate reasons had the strongest effects on students' argumentation. Both resulted in more argumentation claims; the reason condition also increased the number of contingent claims by encouraging students to consider multiple factors. As hypothesised, persuasion resulted in more opposition and debate, as did the counterargument goal condition' (p.306).	Both goal instruction and need for cognition can have 'substantial effects on the nature of student argumentation in an interactive context' (p. 306).	Course instructors not informed of the nature of the interventions but the author taught one of the sections (possibility of bias, teacher effect). Scoring of notes was undertaken blind to condition. Groups where not all 3 students posted notes or which contained less than 80% of the requested notes were eliminated as were outlying cases – therefore possibility of bias introduced here. Multilevel modelling used to control for clustering effects, therefore two levels of analysis – student level and group level. Significance level set at 0.01 because of number of tests

		The additional instructions varied by goal condition which was one of the following: explore, persuade, reasons, counterarguments		first claim' (p291). Each discussion note was scored for type of argumentation claims made (supporting, contingent, divergent) and level of claim (up to 3).			performed.
Sanders, 1994, USA, pub.	To examine whether 'training in argumentation facilitated critical thinking by enhancing the ability to discern strong and weak arguments and whether such training increased the need for cognition, argumentativeness, perceived arguing effectiveness and decreased verbal aggressiveness' (p.27). Case control study	Argumentation instruction: 299 experimental group members who were receiving instruction in argumentation . No details given.	58 comparison group members receiving instruction in introductory interpersonal communication. No details given.	Questionnaires given out in the first and last weeks of term assessing argument perception, perceived arguing effectiveness, argumentativeness, need for cognition and verbal aggressiveness. Three different orders of the arguments and scales were used. The measure of argument effectiveness consisted of three items: convincing-unconvincing, persuasive-unpersuasive, and effective-ineffective. Need for cognition scale: 18 items measuring respondents' predispositions toward handling complex vs. simple problems. Argumentativeness scale: 20 items designed to assess respondents' approach and avoidance feelings toward arguing in conversational settings.	Results revealed that argument instruction enhanced the ability to discern weak example and causal arguments, increased perceived arguing effectiveness and decreased verbal aggressiveness. Further details in Appendix D.	Argumentation training enhances specific critical thinking skills: ability to discern weak arguments, improved self-perception of argumentation effectiveness and decreased verbal aggressiveness in argumentation encounters	The authors note that 'The choice of this sample was made to minimize the effects of self selection bias' (p.30). Use of pre-test scores as covariate in analysis of covariance. No details given about intervention and control treatment. Some limitations of the research acknowledged, e.g., length of intervention only a single term.

				Verbal aggressiveness scale: 20 items reflecting respondents' tendencies to attack the self-concepts of other people. Self-assessment of argumentation skills: 5 items.			
Steinkuehler, 2000, USA, unpub.	To compare the effects of three forms of online discussion on memory, belief change and argumentation skill (p.1). RCT, cluster	On-line discussion: Reading of a pro/con text followed by online discussion in pairs n = 30	Control i: reading of a pro/con text followed by a form of individualised study technique derived from cognitive memory literature: self-explanation n = 30. Control ii: reading of a pro/con text followed by a form of individualised study technique derived from cognitive memory literature: repeated summarisation and study n = 30	Argument change from pre-test to post-test, transfer of argument skills, text recall, reported and directly assessed opinion change, and perceptions of productivity and participation. The two pre-test measures were: brief written essay and an opinion scale to assess opinion and opinion change. The six post-test measures were: short-answer questionnaire designed to assess recall for the given pro/con text; scale designed to measure opinion change; participants wrote a second essay; scale designed to assess opinions; participants wrote another essay to assess transfer of improvements in argumentative reasoning to a topic beyond the activity; scale to measure productivity and participation. Segmentation: topic and transfer essays were coded by two coders.	Across conditions, there was a statistical decrease from pre-test to post-test in the proportion of nonjustificatory statements made, with no statistical differences among the three conditions. There was no decrease in the proportion of non-functional statements made, either across or among conditions. Across conditions, there was a statistical increase in the number of metacognitive statements made, with no difference among conditions. In summary: across conditions, participants' argumentative reasoning improved from pre-test to post-test. However, there were no statistical differences among conditions ('Of the six variables used to investigate improvement in argumentative reasoning, only one variable differentiated		15 participants excluded from analysis – not clear at which point they were excluded but probably after randomization, but not clear how these 15 students had been 'paired'. Excluded from in-depth review due to short length of intervention (45 minutes) (Exclude 2).

					among conditions – the number of statements made that contained or acknowledged the relevance of evidence. In particular, summarisation produced greater use of evidence than did self-explanation and discussion' p.21). For recall test the difference among the three conditions was statistically significant, $p < 0.001$.		
--	--	--	--	--	--	--	--

(Data extraction tables for each of the ten reviews are included in Appendix D)

Interventions evaluated and limitations in the design of the studies

In order to investigate the effectiveness of different strategies and methods for improving argumentation skills in undergraduate students the following interventions were investigated in the ten studies in the overview:

- Four of the ten studies evaluated critical thinking skills or argumentation skills instruction, and in both cases this intervention was compared with a no critical thinking skills instruction control group.
- Three studies evaluated on-line discussion group interventions, compared with face-to-face discussion groups.
- One study investigated the use of a computer-based writing software package, with two different forms of writing tools and prompts compared with a word-processing package.
- Two studies evaluated the use of online argumentation scaffolds.

There was little homogeneity between the interventions even within the four subgroups described above. Although many of these interventions were shown to be effective in improving argumentation skills, caution is urged in interpretation of six of these studies due to limitations in their design and reporting. Most of these studies included a sound rationale, based on previous theoretical and empirical research in argumentation, and a well-developed, theory-based argumentation intervention. However, a number of problems in the research design were noted which precludes any reviewer recommendation for policy or practice. These methodological limitations (which could have led in some cases to poor internal validity) include both reporting and design limitations:

- Very small sample size.
- No raw data given or insufficient raw data given.
- Attrition not given.
- Lack of random assignment to create equivalent groups, leading to possibility of selection bias. Possibility of selection bias in the composition of the two comparison groups not acknowledged (in case control studies and controlled trials).
- Possibility of bias introduced through teacher effect.
- Elimination of outliers in analyses.
- No details or insufficient details given about interventions and control group activities.
- Limited nature of intervention (very short timescale)

Moving from the overview to the in-depth review

Six studies were excluded from the in-depth review. Four studies contained no raw data (Gass *et al*, 1990; Kozma, 1991; Larson *et al*, 2004; Joiner and Jones, 2003), and two studies were excluded due to the limited nature of the intervention evaluated (Nussbaum *et al*, 2004; Steinkuehler, 2000). This left four studies to be included in the in-depth review (Bensley and Haynes, 1995; Cho and Jonassen, 2002; Nussbaum, 2005; Sanders *et al*, 1994).

In-depth review of four studies

Four studies were included in the in-depth review, and therefore analysed in more detail. These studies fulfilled three basic inclusion criteria: they all included sufficient raw data for replication or calculation of effect sizes; they all evaluated an intervention of a reasonable length; and they all evaluated argumentation interventions using argumentation outcomes. The four studies were quality appraised using a modified version of the CONSORT statement (Begg *et al*). The results of the moderated judgements are given in Table 3.

Table 3: Quality appraisal of the four studies in the in-depth review

	Bensley and Haynes 1995	Cho and Jonassen, 2002	Nussbaum, 2005	Sanders et al, 1994
Was the study population adequately described? (i.e. were the important characteristics of the participants described e.g. age, gender?)	NS	NS	Y	Y
Was the minimum important difference described? (i.e. was the smallest educationally important effect size described?)	NS	NS	NS	NS
Was the target sample size adequately determined?	NS	NS	NS	NS
Was intention to teach analysis used? (i.e. were all participants who were allocated included in the follow-up and analysis?)	Y (some analyses) ; N (some analyses)	N	N	NS
Was the unit of randomisation described (i.e. individual students or groups of students)?	NA (not RCT)	Y (cluster)	Y	NA (not RCT)
Were the participants allocated using random number tables, coin flip, computer generation?	NA (not RCT)	NS	NS	NA (not RCT)
Was the randomisation process concealed from the investigators? (i.e. were the researchers who were recruiting students to the trial blind to the student's allocation until after that student had been included in the trial?)	NA (not RCT)	Y	NS	NA (not RCT)
Were follow-up measures administered blind? (i.e. were the researchers who administered the outcome measures blind to treatment allocation?)	Y	Y	Y	NS
Was estimated effect on primary and secondary outcome measures stated?	NS	Y	Y	Y
Was precision of effect size estimated (confidence intervals)?	NS; effect sizes not reported	NS	NS	NS
Were summary data presented in sufficient detail to permit alternative analyses or replication?	Y	Y	Y	Y
Was the discussion of the study findings consistent with the data?	Y	Y	Y	Y

(Y=Yes; N=No; NS=Not stated; NA=Not applicable)

As can be seen from Table 3, a more in-depth analysis of the design and reporting of the four studies in the in-depth review reveals some shortcomings that need to be taken into account when interpreting the results of the studies. For example, none of the four studies explained how the target sample size was determined and yet most of them were underpowered. In only two out of the four studies intention to teach analysis was used, and none of the studies reported confidence intervals. However, in some respects the studies had methodological strengths: three out of the four studies administered follow-up measures blind; all four studies presented summary data in sufficient details to permit replication of the analyses; and in all four studies the discussion of the study findings was consistent with the data.

Bensley and Haynes, 1995

This is a controlled trial investigating a critical thinking intervention designed to improve perception of argumentation potency, with both pre- and post-test data. Findings are 'encouraging but tentative' (p.45). One class received critical thinking skills instruction: the 'critical thinking (CT) program'. These students were provided with booklets which

explained critical thinking within the context of a thinking frame that incorporated three steps: '(a) identify the claim used in an argument; (b) evaluate the evidence relevant to the claim, comparing and weighing evidence both for and against the claim; and (c) draw a reasonable conclusion about the truth of the claim (taking into account the quality and quantity of the evidence)' (second page). The booklet also contained guidelines for writing a persuasive paper in psychology. The CT students were instructed to read the booklet before each class. They also: received instructions about writing a four-page persuasive paper; completed two pages of focused free writing; and submitted an outline. They were then given feedback and further instruction focusing on the model outline and the inclusion of quality evidence.

The outcome measurement was an outline generation task in which students were asked to produce a generic outline for a paper in which they were to argue persuasively. They were asked to identify claims, list the evidence for one side of the argument separately from the other side, classify the evidence as to type, draw an appropriate conclusion, and identify incorrect assumptions or other problems in the passage all from a critical reading passage.

The authors concluded that the results were 'encouraging but tentative' (p.45) in showing that training in critical thinking (as defined in their study) led to significantly greater use of the language of argumentation prescribed in the thinking frame and outlined structure (e.g. greater use of the words 'claim', 'problem', 'purpose', 'question', 'evidence', 'conclusion' and more appropriate content according to the outcome measures. However, they acknowledged that alternative interpretations were possible.

There are a number of threats to the internal validity of this study, some of which were addressed and some of which were acknowledged by the authors. These include:

- Selection bias: use of intact groups, with no random assignment to groups; authors adjust for pre-test in their analyses, and check for baseline equivalence, but the groups could be unbalanced in unknown variables that affect outcome.
- Use of small sample size: the study was underpowered and could have been susceptible to a Type II error. Small sample size also increases the possibility of chance bias.
- Regression to the mean effects: the results of the study could be due to this alternative interpretation because of lack of random assignment.

Therefore the reviewers agree with the authors of the study that the results of the intervention look promising, but tentative.

Cho and Jonassen, 2002

This is a cluster randomized controlled trial which evaluated the use of online argumentation scaffolds 'to engage and support coherent argumentation' (p.5) among undergraduate economics students at a major US university. Specifically, there were four research questions designed to investigate: the relative effectiveness of 'constraint-based scaffolds' and 'unconstrained' discussion boards in improving argumentation and problem-solving during a problem-based activity; whether the quality of the problems (well-structured or ill-structured) affected the nature of argumentation in the two interventions; and whether or not any improvements in argumentation transferred to individual argumentation performance. Therefore random allocation was to one of four conditions: two interventions and two control groups:

- Intervention i: well-structured problems and constraint-based tool; Intervention ii: ill-structured problems and constraint-based tool.
- Control i: well-structured problems and threaded discussion;

Control ii: ill-structured problems and threaded discussion.

The procedures for all intervention and control conditions are described in detail (pp. 9-10; p.11), as are the procedures for outcome assessment. The effectiveness of the interventions was measured by the assessment of the quantity of student arguments during group discussion using a coding scheme adapted from Toulmin's model (which identifies five major components of argument including claims, grounds, warrants, backings and rebuttals p.10); assessment of problem-solving process during group discussion for their problem-solving function by classifying each message based on a coding system adapted from Poole and Holmes (1995) (which consists of seven categories including problem definition, orientation, criteria development, solution approval, solution critique and non-task statement (p.10); assessment of problem-solving activity (an essay describing a solution and justification) to measure transfer of argumentation and problem-solving skills developed during group problem-solving activities using coding based on Toulmin's model of argument to determine the quality of argumentation. Each report was also analysed for problem-solving performance (p.11).

The results of the trial suggested that providing a constraint-based argumentation scaffold during group problem-solving activities increased the generation of coherent arguments. Groups using the scaffold produced more claims and grounds than the control groups. The same scaffold further resulted in 'significantly more problem-solving actions during collaborative discussions' (p.5). The effects of the scaffold varied for problem type, and transferred to individual argumentation but not to individual problem-solving.

There are a number of strengths in the design and reporting of this trial. These include:

- Detailed information about intervention/control procedures and outcome measurement: replication would be possible.
- Blinded follow-up assessment.
- Estimated effect on primary and secondary outcome measures reported (but not confidence intervals).

There was some attrition ($n = 9$), and the number of drop-outs from each condition(s) was not stated. However, the reviewers conclude that the trial is of sufficiently good quality in order to have some confidence in the results and conclusions of the study.

Nussbaum, 2005

This is a cluster randomized controlled trial, which examines: the effect of goal instructions (making explicit the goals of an activity) on students' argumentative writing in an interactive discussion context (Web-based discussions).

In the intervention condition all participants used the discussion board to discuss the following question: 'Does watching TV cause children to become more violent?' The question was posted on the notice board along with additional instructions. One person in the three-person group was randomly assigned to post the first note, after which the other two could post their notes. The additional instructions varied by goal condition, which was one of the following: explore, persuade, reasons, counterarguments. The control condition was provided by the 3x3 design in which the third group acted as a control with no interventions. The design was 3x3 factorial with general goal (Explore, Persuade, None) crossed with a specific goal (Reasons, Counterarguments, None). Outcome measures were two surveys administered by the Web-CT (electronic discussion board): need for cognition and prior attitudes towards television violence. Both surveys used a 5-point Likert scale with responses ranging from 1 (strongly disagree) to 5 (strongly agree). A conceptual framework for analysing arguments was developed. Each discussion note was

scored for type of argumentation claims made (supporting, contingent, divergent) and level of claim (up to 3).

The results indicated that 'the goals to (a) persuade and (b) generate reasons had the strongest effects on students' argumentation. Both resulted in more argumentation claims; the reason condition also increased the number of contingent claims by encouraging students to consider multiple factors. As hypothesised, persuasion resulted in more opposition and debate, as did the counterargument goal condition' (p.306).

The design of this study has a number of strengths including

- Random allocation to condition.
- Blinded assessment of outcome.
- Inclusion of all estimates of effect on outcome measures.
- Details of interventions and analyses are included.
- Multilevel modelling used to control for clustering effects, therefore there were two levels of analysis – student level and group level.
- All the discussion and conclusions are consistent with the data.

It should be noted that the course instructors were not informed of the nature of the interventions, but the author taught one of the sections (introducing the possibility of bias due to teacher effect). Groups where not all 3 students posted notes or which contained less than 80% of the requested notes were eliminated as were outlying cases – again introducing the possibility of bias. However, the reviewers thought that the trial was of sufficiently high quality for them to have confidence in the results and the author's conclusions.

Sanders et al, 1994

This is a case control study, which examined whether 'training in argumentation facilitated critical thinking by enhancing the ability to discern strong and weak arguments and whether such training increased the need for cognition, argumentativeness, perceived arguing effectiveness and decreased verbal aggressiveness' (p.27). Two hundred and ninety nine students receiving argumentation instruction were compared with 58 comparison group members who didn't receive argumentation instruction, but did receive instruction in interpersonal communication. The participants were undergraduates at two major western US universities. The outcome measurement was assessment of argument perception, perceived arguing effectiveness, argumentativeness, need for cognition and verbal aggressiveness through questionnaires.

Results revealed that argument instruction enhanced the ability to discern weak example and causal arguments, increased perceived arguing effectiveness and decreased verbal aggressiveness.

The authors noted that 'the choice of this sample was made to minimize the effects of self selection bias' (p.30). However, the design of the research (case control) could have been susceptible to selection bias. The authors did use pre-test scores as covariates in the analysis of covariance, but this may not have been sufficient to eliminate selection bias. No details were given about the intervention and control conditions, and therefore these could not be replicated. The authors acknowledge some limitations to the research design. For example, the length of intervention was only a single term. In conclusion, the reviewers agree with the authors that argumentation training probably does enhance specific critical thinking skills: ability to discern weak arguments, improved self-perception of argumentation effectiveness and decreased verbal aggressiveness in argumentation encounters.

Quality assurance results: Screening at first stage

An initial sample of 20 studies was independently screened and moderated by all four reviewers. Overall, there was very good agreement. There was disagreement on one problem study, and no further disagreement between CT and RA. There were only two further disagreements for AR and BS. CT and RA tended to be less inclusive than AR and BS, and this was probably due to the relative inexperience of the more inclusive reviewers. We discussed issues arising from the exercise and agreed to move to the screening of the entire database. CT screened the entire database; RA screened a 10% sample. The reliability of the reviewers' screening was high (97%) agreement. A Cohen's Kappa calculation for the inter-rater agreement was 0.54 (moderate). This was due to the extremely high proportion of exclusions relative to the number of title and abstracts screened, which in turn was due to the sensitive, but not specific, search strategy specifically designed to be as inclusive as possible. It was felt by all members of the review team that they could have confidence in the main reviewer's screening decisions, and so the double screening was taken no further.

Quality assurance results: Screening at second stage

CT and RA screened the full papers sent for after screening at first stage. Agreement was high; any disagreements were discussed and resolved and papers were included or excluded by consensus between the two reviewers.

Quality assurance results: Screening at third stage

Where studies were excluded at third stage, i.e., during the data extraction process, this was done by consensus between at least two reviewers.

Quality assurance results: data extraction and quality appraisal

Two teams of reviewers independently undertook data extraction of all the studies in the map (CT and RA; CT and BS). The teams of reviewers then met and moderated all decisions. Agreement was very high; any disagreements were resolved through discussion. Similarly, independent double quality appraisal was undertaken and any disagreements were resolved through discussion.

Conclusions

The most important finding from this systematic review was that, despite the well-developed theoretical literature in argumentation, there is a lack of retrievable large well-designed, well-reported randomized controlled trials or controlled trials in the field. Specifically, in the period 1990 to the present there are only a few studies with comparison groups that evaluated methods and strategies to improve argumentation in undergraduate students in HE. The lack of RCTs or CTs undertaken in the UK is particularly notable.

Substantive findings

Some substantive conclusions can be tentatively drawn from the four studies in the in-depth review, although these should be read in conjunction with the discussion about the strengths and limitations of each study, and it should be noted that each of the conclusions is only supported by one study:

- Training in critical thinking looks promising both for the conventionalization of the language of argumentation in writing, and for the greater use and appropriate positioning of argumentation contents in writing (Bensley and Haynes, 1995).
- A constraint-based argumentation scaffold can improve both group argumentation and individual argumentation (Cho and Jonassen, 2002).
- The use of goal instruction to (a) persuade and (b) generate reasons in argumentation in an interactive context can improve students' argumentation by increasing argumentation claims, contingent claims, opposition and debate (Nussbaum, 2005).
- Argumentation training can enhance specific critical thinking skills: ability to discern weak arguments, improved self-perception of argumentation effectiveness and decreased verbal aggressiveness in argumentation encounters (Sanders *et al*, 1994).

For studies not included in the in-depth review caution is urged in drawing any strong substantive conclusions, due to limitations in both the design and reporting of some of the six studies, and in the limited match between argumentation intervention and outcomes in other identified studies.

Recommendations

Substantive recommendations

Although the studies in the in-depth review concluded that some strategies and methods evaluated using controlled research designs were effective in improving argumentation skills in undergraduate students, strong substantive recommendations for practice cannot be made. We suggest that such interventions look promising but should be adopted with caution, given the variable quality of the studies involved and the fact that each strategy or method was evaluated in only one study. This is not high quality evidence for changes to policy or practice.

Recommendations for further research

A large rigorously designed, conducted and reported randomized controlled trial in a UK HE setting is recommended. This could investigate the effectiveness of one or more of the strategies identified in this systematic review as looking promising for the improvement of argumentation skills in undergraduate students. Such a trial might form part of a larger empirical study of argumentation at undergraduate level in the UK, as recommended in the report of the pilot study that ran alongside the present review (see Andrews et al., 2006).

References

Studies included in the overview (* studies in the in-depth review)

*Bensley, D.A. and Haynes C. (1995) The acquisition of general purpose strategic knowledge for argumentation, *Teaching of Psychology*, 22(1): 41-45.

*Cho, K.-L. and Jonassen, D.H. (2002) The effects of argumentation scaffolds on argumentation and problem solving, *Educational Technology Research and Development*, 50(3): 5-22.

Gass, R.H., Sanders, J.A. and Wiseman, R.L. (1990) *Warrants, argument dimensions, perceived argument effectiveness, and the influence of critical thinking instruction*. Paper presented at the Annual Meeting of the Western Speech Communication Association, Sacramento, CA, February 16-20.

Joiner, R.W. and Jones, S. (2003) The effects of communication medium on argumentation and the development of critical thinking, *International Journal of Educational Research*, 39(8): 861-871.

Kozma, R.B. (1991) The impact of computer-based tools and embedded prompts on writing processes and products of novice and advanced college writers, *Cognition and Instruction*, 8(1): 1-27.

Larson. M., Britt, M.A. and Larson, A.A. (2004) Disfluencies in comprehending argumentative texts, *Reading Psychology*, 25(3):205-224.

Nussbaum, E.M., Hartley, K., Sinatra, G.M., Reynolds, R.E. and Bendixen, L.D. (2004) Personality interactions and scaffolding in on-line discussions, *Journal of Educational Computing Research*, 30(1-2): 13-136.

*Nussbaum, E.M. (2005) The effect of goal instructions and need for cognition on interactive argumentation, *Contemporary Educational Psychology*, 30(3): 86-313.

*Sanders, J.A., Wiseman, R.L., and Gass, R.H. (1994) Does teaching argumentation facilitate critical thinking? *Communication Reports*, 7(1): 27-35.

Steinkuehler, C.A., Derry, S.J., Levin, J.R. and Kim, J.-B. (2000) *Argumentative reasoning in online discussion*. Paper presented at the Annual Meeting of the American Educational Research Association, New Orleans, LA, April 24-28.

Other references

Andrews, R.J., Bilbro, R., Mitchell, S., Peake, K., Prior, P., Robinson, A.M., See, B.H. and Torgerson, C. (2006) *Argumentative skills in undergraduate students in Higher Education: a pilot study* York: The Higher Education Academy

Aristotle (1909) *The Rhetoric* (trans. R.C. Jobb), Cambridge: Cambridge University Press.

Bain, A. (1866) *English Composition and Rhetoric*, London: Longmans and Co.

Begg, C.B., Cho, M.K., Eastwood, S., Horton, R., Moher, D., Olkin, I., Pitkin, R., Rennie, D., Schulz, K.F., Simel, D. and Stroup, D.F. (1996) Improving the quality of reporting of randomized controlled trials: The CONSORT statement, *Journal of the American Medical Association*, 276(8): 637-639.

Burke, K. (1950) *A Rhetoric of Motives*, Berkeley: University of California Press.

Casanave, C.P. (1995) Local interactions: Constructing contexts for composing in a graduate sociology program. In D. Belcher and G. Braine (Eds.), *Academic writing in a second language: Essays on research and pedagogy* (pp. 83-110) Norwood, NJ: Ablex.

Cicero (1948) *On Oratore* (Ed. H. Rackham), London: Heinemann.

Cicero (1954) *Ad Herennium* (trans. H. Caplan), London: Heinemann.

Eagleton, T. (1983) *Literary theory: An Introduction*, Oxford: Blackwell.

Emmel, B. Resch, P. and Tenney, D. (Eds.) (1996) *Argument Revisited; Argument Redefined: Negotiating Meaning in the Composition Classroom*, Thousand Oaks: CA: Sage.

Herrington, A. (1985) Writing in academic settings: A study of the contexts for writing in two chemical engineering courses, *Research in the Teaching of English*, 19:331-359.

Langer, J.A. and Applebee, A.N. (1987) *How writing shapes thinking: A study of teaching and learning*, Urbana, IL: National Council of Teachers of English.

Lunsford, A., Ruskiewicz, J. and Walter, K. (1001) *Everything's an Argument*, (2nd edition), Boston, MA: Bedford/St. Martins.

Lunsford, K (2002) Contextualizing Toulmin's model in the writing classroom: A case study, *Written Communication*, 19, 109-174.

Prior, P. (1998) *Writing/Disciplinarity: A Sociohistoric Account of Literate Activity in the Academy*, Mahwah: Lawrence Erlbaum.

Mitchell, S. and Andrews, R. (2001) (eds.) *Learning to Argue in Higher Education*, Portsmouth NH: Heinemann/Boynton-Cook

Sampson, G. (1921) *English for the English*, Cambridge: Cambridge University Press.

Torgerson, C. (2003) *Systematic reviews*. London: Continuum.

Toulmin, S. (1958) *The Uses of Argument*, Cambridge: Cambridge University Press

Van Eemeren, F., Grootendorst, A., Jackson, S. and Jacobs, S. (1993) *Reconstructing Argumentative Discourse*, Tuscaloosa: University of Alabama Press.

Vickers, B. (1988) *In Defence of Rhetoric*, Oxford: Clarendon Press.

Williams, J. and Colomb, G (2003) *The Craft of Argument*, (2nd edition), New York: Longman.

Appendices

Appendix A: Search strategy

1. Search history ERIC via Cambridge Scientific Abstracts (n = 1734)

1. KW = argumentation
2. KW = writ* and (persuasive* or argument*)
3. KW = (persuasive* and communicat*)
4. KW = (rhetoric and composition) or (rhetoric* and instruction*)
5. DE = essay* or (persuasive discourse) or (expository writing)
6. DE = (higher education) or (college freshmen) or (undergraduate students) or (undergraduate study)
7. (1 or 2 or 3 or 4 or 5) and 6

PT=(140 reports general) or PT=(141 reports descriptive) or PT=(142 reports evaluative) or PT=(143 reports research) or PT=(042 dissertations/theses masters theses) or PT=(072 book /product reviews) or PT=(080 journal articles) or PT=(010 books) or PT=(040 dissertations/theses) or PT=(041 dissertations/theses doctoral dissertations)

Date range: 1990-2005

Limited to: English only

2. Search history PsycINFO via ARC2 WebSPIRS (n = 495)

1. KW = argumentation or (persuasive discourse)
2. KW = writ* and (argument* or persuasive* or expository or essay*)
3. KW = (rhetoric and composition) or (rhetoric* and instruction*)
4. DE = (persuasive communication)
5. DE = (higher education)
6. KW = undergraduate* or (university student*) or (college student*)
7. (1 or 2 or 3 or 4) and (5 or 6)

(DT:PSYI = authored-book) or (DT:PSYI = chapter) or (DT:PSYI = dissertation-abstract) or (DT:PSYI = edited-book) or (DT:PSYI = journal) or (DT:PSYI = original-chapter) or (DT:PSYI = original-journal-article) or (DT:PSYI = peer-reviewed-journal) or (DT:PSYI = reprinted-chapter) or (DT:PSYI = reprinted-journal-article) or (DT:PSYI = review)

PY:PSYI = 1990-2005

LA:PSYI = English

3. Search history SSCI via ISI Web of Knowledge (n = 196)

1. TS = argumentation or (writ* and (persuasive* or argument* or expository* or essay*))
2. TS = (persuasive* and communicat*) or (persuasive discourse) or (rhetoric and composition) or (rhetoric* and instruction*)
3. TS = (higher education) or (undergraduate*) or (college student*) or (university student*)
4. (1 or 2) and 3

DocType=All document types; Language=English; Database(s)=SCI-EXPANDED, SSCI; Timespan=1990-2005

4. Search history BEI via Dialog Datastar (n = 93)

1. KW = argumentation

2. KW = writ\$ AND (argument\$ OR persuasive\$ OR expository OR essay\$)
3. KW = (persuasive ADJ discourse) OR (rhetoric AND composition) OR (rhetoric\$ AND instruction\$)
4. KW = (higher ADJ education) OR undergraduate\$ OR (university ADJ student\$) OR (college ADJ student\$)
5. (1 or 2 or 3) and 4

PT=conference-papers OR PT=conference-proceedings OR PT=evaluation-studies OR PT=journal-article OR PT=literature-reviews OR PT=research-reports

Limit year > 1989

LG=English

5. Search history C2SPECTR (n = 12)

1. KW = {argumentation} or {persuasive discourse} or {persuasive writing} or {persuasive communication} or {expository writing}
2. KW = {rhetoric} and {instruction}
3. KW = {rhetoric} and {communication}
4. KW = {writ*} and {argument*}
5. KW = {higher education} or {undergraduate*} or {freshman} or {college} or {university}
6. (1 or 2 or 3 or 4) and 5

Date range: 1990 – 2005

Limited to: English only

NB: Numbers shown represent papers retrieved prior to de-duplication.

Appendix B

Inclusion criteria

- (1) Topic focus: Must be about improving argumentation skills
- (2) Setting focus: Must be set in a Higher Education institution with undergraduate participants aged 18 years or over
- (3) Study type: Must be an evaluation of an intervention using one of the following study types – randomized controlled trial, controlled trial, review
- (4) Language focus: Must be undertaken in an English-speaking country and written in the English language
- (5) Year of publication: Must be published or unpublished but in the public domain in the years 1990-2005

Exclusion criteria:

- (1) Topic focus: Not about improving argumentation skills (Exclude 1)
- (2) Setting focus: Not set in a Higher Education institution with undergraduate participants aged 18 years or over (Exclude 2)
- (3) Study type: Not an evaluation of an intervention using one of the following study types – randomized controlled trial, controlled trial, review (Exclude 3)
- (4) Language focus: Not undertaken in an English-speaking country and written in the English language (Exclude 4)
- (5) Year of publication: Not published or unpublished but in the public domain in the years 1990-2005 (Exclude 5)

Appendix C

Inclusion criteria:

- (1) Must contain raw data (means standard deviations, numbers) to enable replication of analyses
- (2) Must evaluate an intervention of a reasonable length
- (3) Must evaluate intervention focusing on argumentation and use argumentation outcome measures

Exclusion criteria:

- (1) No raw data
- (2) Limited intervention
- (3) Intervention or outcome measures not specifically focused on argumentation

Appendix D: data extraction tables for all included studies

Bensley, D.A. and Haynes, C. (1995) The acquisition of general purpose strategic knowledge for argumentation. <i>Teaching of Psychology</i>, 22(1): 41-45.	
Country of origin	USA
Setting	Psychology classes at Texas Wesleyan University
Objective	<p>To test a programme designed to develop critical reading and writing skills in psychology students through the acquisition of a general knowledge structure for argumentation.</p> <p>Critical thinking is defined as ‘the evaluation of evidence relevant to a claim so that a reasonable conclusion about the truth of the claim can be made’ (second page*).</p> <p>It was hypothesized that ‘the general knowledge for argumentation would be reflected in both the content and positioning of the words used.....and that compared to CO students, after training CT students would (a) more frequently use the language of argumentation conventionalized by the thinking frame and outline structure, (b) use this language in positions consistent with the thinking frame and outline structure, and (c) use more contents judged to be appropriate by critical thinking experts’ (second page).</p>
Rationale	‘findings (from a number of studies) suggest the need for structures that can be used strategically to help organise, guide, and support critical thinking and writing’ (second page).
Theoretical and research background	<p>An example of a structure was proposed by Perkins (1987): ‘thinking frames that organize and support thinking’; and he provided the essay formula from rhetoric, ie, thesis, argument, counter-argument, rebuttal, conclusion. Also, ‘Few programs have combined training of critical thinking and argumentative writing in psychology. In one of these programs Rathus (1993) provided suggestions for critical thinking and writing but did not offer product guidelines, a structured thinking frame for structuring critical writing, or direct instruction in the language of argumentation’ (second page). No evaluation of this program’s effectiveness has been published. Pilot study from the authors’ laboratory showed that students’ use of critical/analytical vocabulary in critical-thinking programmes was quite variable. For the study, a thinking frame was designed using the language of argumentation found in a common definition of critical thinking.</p>
Study design	<p>Controlled trial – cluster (2 clusters): ‘Random assignment was not possible due to instructional constraints, so two intact treatment groups were formed’ (second page).</p> <p>The students were all pre-tested and post-tested.</p>
Outcome measures	<p>Use of argumentative language in constructing generic outlines for a persuasive term paper.</p> <p>Pre-test: outline generation task in which students were asked to produce a generic outline for a paper in which they were to argue persuasively; they were asked to identify claims and list the evidence for one side of the argument separately from the other side and classify the evidence as to type and draw an appropriate conclusion and identify incorrect assumptions or other problems in the passage all from a critical reading passage.</p> <p>Post-test: outline generation and critical reading tasks.</p>
Study details	
Participants	10 men and 16 women from two small, general psychology classes. There were 16 in the CT group and 10 in the CO group.
Intervention	One class received critical thinking skills instruction: the ‘critical thinking (CT) program’. These students were provided with booklets which

	<p>explained critical thinking within the context of a thinking frame that incorporated three steps: '(a) identify the claim used in an argument; (b) evaluate the evidence relevant to the claim, comparing and weighing evidence both for and against the claim; and (c) draw a reasonable conclusion about the truth of the claim (taking into account the quality and quantity of the evidence)' (second page). The booklet also contained guidelines for writing a persuasive paper in psychology:</p> <ul style="list-style-type: none"> I Statement of basic problem or question <ul style="list-style-type: none"> A Purpose in writing the paper B How the problem is to be discussed C Definition of relevant terms II Development of the discussion of the problem <ul style="list-style-type: none"> A Evidence for position or claim B Evidence against position or claim III Conclusion <ul style="list-style-type: none"> A Review of main points of the discussion of the question B Evaluation of main points of the discussion and conclusion about the question <p>The CT students were instructed to read the booklet before each class. They also received instructions about writing a four-page persuasive paper. They also completed two pages of focused free writing and submitted an outline. They were then given feedback and further instruction focusing on the model outline and the inclusion of quality evidence.</p>
Control	One class did not receive the critical thinking skills instruction – the 'control (CO) group'. These students studied course content similar to the CT students using the same general psychology textbook. The text did not contain any explicit instruction in critical thinking.
Outcome	
Results	<p>The critical thinking skills instruction group showed greater use of language similar to that used in instruction and significantly greater use of argumentative language than the group not getting the instruction. The ANCOVA showed that the CT group produced significantly more of the 22 content categories than did the CO group (means, standard deviations and numbers given). Results given in tables 1 and 2: only percentages of frequency given so reviewers not able to calculate any effect sizes. Further ANCOVA of the number of correctly classified examples of relevant evidence identified by subjects in the critical reading task using pre-test as a covariate showed that the CT group performed significantly better than the CO group (again means, sds and numbers given).</p>
Reviewer summary	Controlled cluster trial with only two clusters, therefore design not able to control for selection bias and chance bias. Also, no random assignment, therefore not able to control for any extraneous factors. However, as described on third page, ANCOVA was performed on post-test scores using a procedure in which pre-test scores were adjusted for the reliability of pre-test and post-test scores. However, raw data for pre-test not given.
Comments	Cluster trial with only two clusters – therefore probability of selection bias likely. Extremely small sample size two clusters of 16 and 10 participants – therefore possibility of chance bias. A discussion of these limitations, plus also the possibility of resentful demoralisation and regression to the mean accounting for the observed effects is included in the paper. The authors therefore describe their findings as 'encouraging but tentative' (fifth page*).

*Page numbers not available

Cho, K.-L. and Jonassen, D.H. (2002) The effects of argumentation scaffolds on argumentation and problem solving. <i>Educational Technology Research and Development</i>, 50(3): 5-22.	
Country of origin	USA
Setting	Major university in the eastern US
Objective	<p>'To examine the use of online argumentation scaffolds to engage and support coherent argumentation' (p.5).</p> <p>Four research questions outlined in detail on page 8:</p> <ul style="list-style-type: none"> • How do constraint-based scaffolds affect the nature of argumentation during a problem-solving activity when compared with an unconstrained discussion board? • How do constraint-based scaffolds affect the nature of problem-solving activity during discussions? • Do constraint-based discussions differentially affect the nature of argumentation and problem-solving while solving well-structured versus ill-structured problems? • Do the effects of constraint-based scaffolds transfer to individual argumentation performance and to individual problem-solving performance following treatment?
Rationale	Direct instruction in argumentation has produced inconsistent results. Rationale for 'CSCA' page 6 – theory of constraints in argumentation. Research results into CSCA inconsistent – 'research is needed to clarify these differences' (p.8)
Theoretical and research background	Theoretical and empirical background to purposes of 'argumentation'. Empirical research about students' abilities in argumentation. Theoretical and empirical background to alternative approach - 'scaffolding' of argumentation skills through the use of cognitive tools - 'computer-supported collaborative argumentation' – to scaffold use of warrants and evidence from supporting claims. 'Limited research' has examined the use of constraint-based argumentation scaffolds.
Study design	Cluster RCT. Random assignment to groups of 3; random assignment to one of 4 treatment groups, according to the 2 independent variables: problem type (2 different types of problem – well-structured and ill-structured) and type of discussion forum used by the groups (threaded discussion or constraint-based discussion using Belvedere).
Outcome measures	<p>Page 10: assessment of quantity of student arguments during group discussion using a coding scheme adapted from Toulmin's model (which identifies five major components of argument including claims, grounds, warrants, backings and rebuttals).</p> <p>Page 10: Assessment of problem-solving process during group discussion for their problem-solving function by classifying each message based on a coding system adapted from Poole and Holmes (1995) (which consists of seven categories including problem definition, orientation, criteria development, solution approval, solution critique and non-task statement).</p> <p>Page 11: Problem-solving activity (an essay describing a solution and justification) to measure transfer of argumentation and problem-solving skills developed during group problem-solving activities using coding based on Toulmin's model of argument to determine the quality of argumentation. Each report was also analysed for problem-solving performance.</p>
Study details	
Participants	College students registered for undergraduate introductory economics course at a major university in the eastern united states. n = 69 volunteers. However, missing data reduced the number in the final analyses to 60 participants: 32 male and 28 female. First year to fourth year undergraduates (US equivalent).

Intervention	Intervention i: well-structured problems and constraint-based tool Intervention ii: ill-structured problems and constraint-based tool
Control	Control i: well-structured problems and threaded discussion Control ii: ill-structured problems and threaded discussion Page 11: 'during the study, students in small groups worked together to perform the well-structured or ill-structured problem-solving tasks. Group members met online. Although students were not required to use university laboratories to complete these activities, most did. Participants in the threaded discussion groups used only the BBS, while students in the Belvedere group constructed solutions and arguments that they then corresponded to each other using the BBS. Each week for three weeks, the groups solved one economics problem. During the study, each group solved three problems. The week after the treatment period, all students completed the individual problem-solving activity in class'.
Outcome	
Results	<p>Summary: 'Providing a constraint-based argumentation scaffold during group problem-solving activities increased the generation of coherent arguments. The same scaffold further resulted in significantly more problem-solving actions during collaborative discussions' (p.5). The effects of the scaffold varied for problem type.</p> <p>A MANOVA showed that students using the argumentation scaffold, Belvedere, produced significantly more argument components during group discussion than students in the unscaffolded discussion groups (Table 2 ES 0.231).</p> <p>Groups solving ill-structured problems produced more arguments during group discussion than students solving well-structured problems (ES 0.27). The MANOVA showed a main effect for the constraint-based scaffold. The groups using the scaffold produced more problem-oriented comments during their discussion (Table 3 ES 0.51)</p> <p>The one-way ANOVA showed that students who had used the constraint-based scaffold, Belvedere, created higher quality argumentation in their individual problem-solving exercises than students who had used the threaded discussion. Another one-way ANOVA showed that students faced with ill-structured problems created higher quality arguments in individual problem-solving than students facing well-structured problems. A one-way ANOVA showed that students who had used the argumentation scaffold, Belvedere, did not perform significantly better on their individual problem-solving tasks than students who did not have access to the scaffold. The result indicated that there is no conclusive evidence that argumentation scaffold transfers to individual problem-solving performance as determined by the performance criteria used.</p>
Reviewer summary	
Comments	No pre-tests. Not clear whether there is adjusting for cluster effects.

Gass, R.H., Sanders, J.A. and Wiseman, R.L. (1990) <i>Warrants, argument dimensions, perceived argument effectiveness, and the influence of critical thinking instruction.</i> Paper presented at the Annual Meeting of the Western Speech Communication Association, Sacramento, CA, February 16-20.	
Country of origin	USA
Setting	Two major western universities in the US.
Objective	To investigate the interrelationships among warrants, argument dimensions and perceived argument effectiveness by surveying subjects with varying degrees of training in critical thinking 'to determine the extent to which their degree of training and initial attitudes, coupled with the type of warrant employed, influenced their evaluations of overall argument potency as measured along three dimensions' (p.2).
Rationale	At the time of the study it was not known whether or not a basic course in critical thinking would affect how argument potency is perceived or evaluated (p.1). The existing literature on argumentation was almost entirely prescriptive; little was known about how people evaluate or perceive argument, what influence warrant type has on individuals' overall estimates of argument strength or quality. The authors refer to literature on evaluation of warrants as basis for their research. These include works by Reinard (1984), Petty and Cacioppo (Cacioppo, Petty & Morris (1983); Petty & Cacioppo (1979, 1981, 1984 and 1986), Pentony (1986), Benoit (1987). There was little empirical research about the influence of critical thinking instruction on warrant evaluation (The authors refer to the work by Follert and Colbert (1983) (p.8-9). Where empirical studies had been conducted, their results were inconsistent. Research which used competitive debaters as respondents may not be representative of general student population because of self-selection bias. The present study compared students who had completed a course in argumentation with those who had no training in argumentation.
Theoretical and research background	Background to Toulmin's model given on p.2. The researchers chose to focus on three types of substantive warrants: reasoning by example; reasoning by analogy; and cause-to-effect reasoning, because they 'represent well-established, prototypical forms of substantive reasoning' (p.2) and also because high reliabilities had previously been obtained for inter-expert agreement on the classification of the three warrant types. A literature review of studies investigating the effectiveness of arguments based on the Toulmin model, argument quality and quantity, research on 'strong' and 'weak' arguments, and whether perception of argument strength is related to warrant type was included. No empirical research to date (1990) examining influence of warrant type on individual's overall estimates of argument strength or quality.
Study design	Survey of students with 61.6% having taken a lower-division course in critical thinking or argumentation and the remaining 38.4% had not taken a critical thinking course, but they had taken a lower-division, fundamentals of communication course. Therefore design of study is 'case control', using a questionnaire.
Outcome measures	Questionnaire: 18 arguments made up the stimulus material: strong and weak arguments based upon example, analogy and cause-and-effect reasoning (two controversial topics and one neutral topic). The participants were asked to assess each argument's potency on three dimensions of strength: logical, emotional and ethical – three response items were determined for each of the dimensions.
Study details	
Participants	393 students from two major western universities: mean age 21.5; 57.3% female; 62.6% Anglo-American; 20.1% Asian-American. 242 (62%) had

	taken a course in critical thinking argumentation, while the other 151 (38%) had not taken a course on critical thinking.
Intervention	242 students had taken a course in critical thinking or argumentation. No details given of the course.
Control	151 students had taken a course in the fundamentals of communication. No details given of the course.
Outcome	
Results	<p>The effect of critical thinking instruction on perceptions of arguments that were judged weak by experts in argumentation: significant differences for course instruction ($F = 5.3$, $df = 1/1149$, $p < .02$)</p> <p>Perceptions of weak arguments: Also warranted by course interaction: subjects who had instruction in critical thinking were better able to detect weakness in the example and analogy arguments than the subjects who had no critical thinking instruction ($F = 4.2$, $df = 2/1149$, $p < .02$).</p> <p>Course instruction in critical thinking improved the subjects' ability to detect logic weaknesses ($F = 9.7$, $df = 1/1149$, $p < .002$).</p> <p>A significant difference was found for coursework on argumentation or critical thinking ($F = 4.8$, $df = 1/1148$, $p < 0.03$).</p> <p>Subjects who had taken a course in critical thinking were more adept at detecting logic weaknesses in the example and analogy arguments than subjects who had not taken the course in critical thinking ($F = 3.1$, $df = 2/1149$, $p < .05$).</p> <p>Perceptions of strong arguments: Students who had taken a critical thinking course perceived the logic of strong arguments as less compelling ($F = 4.8$, $df = 1/1148$, $p < .03$).</p> <p>Students who had instruction in critical thinking perceived less ethicality in the strong arguments than did students without critical thinking instruction ($F = 7.0$, $df = 1/1149$, $p < .01$).</p>
Reviewer summary	Case control study comparing students who had taken in course in critical thinking with those who had taken a course in communication, post-test.
Comments	<p>No way of knowing how the participants on each of the courses differed at the beginning, therefore possibility of selection bias. This limitation of the research is not discussed.</p> <p>No raw data, therefore only included in map not in in-depth review (Exclude 1)</p>

Joiner, R. and Jones, S. (2003) The effects of communication medium on argumentation and the development of critical thinking. <i>International Journal of Educational Research</i>, 39(8): 861-871.	
Country of origin	UK
Setting	University of Bath
Objective	'...to compare the effects of different communication media on the quality of arguments and the development of argumentative reasoning' (p.861).
Rationale	Research suggests that high-quality argumentative reasoning (critical thinking) is not widespread in the adult population (p.862) (e.g. Kuhn, 1991).
Theoretical and research background	Background to research that has looked at interventions that use peer interaction to facilitate argumentative reasoning indicating that collaborative thinking and arguing might enhance quality of reasoning. Background to the use of asynchronous computer-mediated communication (CMC) and its possible strengths and limitations.
Study design	'Friendship' groups of about 4 randomly assigned to condition: therefore cluster RCT.
Outcome measures	Pre-test – mini essay to ascertain participants' opinion of corporal punishment, and a 21-point smacking scale. Post-test – identical to the pre-test except that there was also a 17-point self-reported measure of how students' opinions of corporal punishment had changed.
Study details	
Participants	73 undergraduate students (9 males and 64 females, ranging in age between 18-40. All participants were given two papers to read before the seminar discussions.
Intervention	The 'on-line group' participated in a seminar discussion over a period of two weeks using <i>Blackboard</i> a commercial virtual learning environment, n = 39. The seminar discussion was about the detrimental effects of smacking on children's development.
Control	The 'face-to-face group' participated in a seminar discussion for one hour, n = 34. The seminar discussion was about the detrimental effects of smacking on children's development.
Outcome	
Results	<p>Analysis of argument:</p> <ul style="list-style-type: none"> • Arguments in the students' mini essays in the pre-test and post-test were classified under 3 categories: Type I, Type II and Type III, using a framework by Kuhn (1991). • Within each category, arguments were further classified according to whether they provided alternatives to CP, whether they offer reasons for or against CPs (Types 1A, B, C & D). • Analysis of arguments was based on a scheme developed by Felton and Kuhn (2001). It comprised 3 broad categories: transactive questions, transactive statements and non-transactive statements. A transactive utterance is one that engages the partner in discourse. <p>3 patterns of utterances were used to classify dialogues: Corner Sequence (where the speaker corners the partner in an untenable position), Rebuttal, and the Block where the speaker rejects or counter argues the premise of a leading question posed by the partner.</p> <p>Results:</p> <p>No statistically significant differences were found between the two conditions in terms of improvement in the quality of students' arguments or in terms of change of opinion about corporal punishment. Argumentative reasoning was not facilitated by either face-to-face or computer-mediated discussion. Participants in the face-to-face condition were more likely to make transactive questions (utterances that request a response from a</p>

	partner), Table 1, and 'meta-?' transactive statements. The participants in the face-to-face condition made proportionally more 'add', 'agree', 'disagree' and 'refuse' transactive utterances, Table 3, and proportionally fewer 'unconnected non-transactive utterances. (Non-transactive utterances fail to connect to the partner's preceding utterance.) Finally, participants in the face-to-face condition showed twice as many strategic sequences as the participants in the CMC condition. Therefore, the quality of the argumentation used in face-to-face was higher than that used in the online discussions' (p.861).
Reviewer summary	Pre-test, intervention session consisting of either an asynchronous online discussion or a face-to-face discussion, and post-test.
Comments	Exclude from in-depth review due to limited nature of intervention (Exclude 2)

Kozma, R.B. (1991) The impact of computer-based tools and embedded prompts on writing processes and products of novice and advanced college writers. <i>Cognition and Instruction</i> , 8(1): 1-27.	
Country of origin	USA
Setting	A midwestern comprehensive university and a midwestern research university
Objective	To examine the impact of computer-based tools and embedded and topical and rhetorical prompts on college writers; specifically, to 'examine the interaction between the cognitive skills of learners and two software packages designed to support these components (higher level structural considerations and rhetorical planning) of the writing process' (p.4)
Rationale	There was little research to date on the effects of software packages to facilitate conceptual planning in writing.
Theoretical and research background	Theoretical background to the differences between novice and expert writers (particularly in relation to the work of Bereiter and Scardamalia and Flower and Hayes). Background to computer-based tools.
Study design	Students were randomly assigned to a particular computer-based writing software package – either Acta (n=13) or Learning Tool (n=15) in addition to MacWrite, whilst some students only used MacWrite (n=13) (control group).
Outcome measures	Pre- and post-tests: measures of cognitive strategies used during the writing process using think-aloud protocol methodology and scored for the proportion of time that subjects engaged in conceptual planning (process planning, rhetorical planning and structure planning) (interrater reliability for coding ranged between 0.7 and 0.88); argumentative tasks scored for overall quality, organisation, effectiveness of argument, and the extent to which the paper exhibited an awareness of an audience, using a 6-point primary-traits scale to rate the overall quality of a composition, the effectiveness of an argument, and the extent to which the composition reflected an awareness of the intended audience (interrater reliability ranged from 0.84 to 0.88).
Study details	
Participants	Participants were 'novice' and 'advanced' writers. The novices were 21 students enrolled in an introductory English composition course. The advanced writers were 20 students who had taken at least two writing courses (one of which was a course in argumentative writing).
Intervention	Intervention i: practice assignment (argumentative writing) using MacWrite plus Acta with or without rhetorical prompts Intervention ii: practice assignment (argumentative writing) using MacWrite plus Learning Tool with or without rhetorical prompts
Control	Practice assignment using MacWrite with or without rhetorical prompts. Students 'were randomly assigned to a particular computer-writing software package' – Acta or the idea organizer, Learning Tool. 'The prompting treatment consisted of a series of questions related to the topic, the audience, and the writer's goals for the composition' (p.6). The experimental group (18) received the prompts; the control group (23) did not. The intervention/prompt is given in full in the appendix (pp.26-7).
Outcome	
Results	Significant effects for both software and prompts. 'Writers did the least conceptual planning with the word processor (6.32%); more with the ideal organiser, Learning Tool (10.02%) and the most with the outliner, Acta (12.58%). Writers did more conceptual planning with embedded prompts (11.53%) than without them (8.08%)' (p.9). There was a significant interaction: 'the effects of software and prompts were compounded' (p.9). There was no interaction between writing level and treatments: 'There were

	no more benefits for novice writers than there were for advanced writers' (p.10). However, 'compositions were no better when organisational software or rhetorical prompts were used' (p.11).
Reviewer summary	This is an individual RCT with two intervention groups and one control group, with pre- and post-tests.
Comments	Exclude from in-depth review due to lack of data (standard deviations) (Exclude 1); very short intervention – one practice composition (Exclude 2); and Exclude 3 on basis that although outcome measures were argumentative writing, interventions were peripheral to this.

Larson, M.M., Britt, M.A. and Larson, A.A. (2004) Disfluencies in comprehending argumentative texts (Experiment 2). <i>Reading Psychology</i>, 25: 205-224.	
Country of origin	USA
Setting	University
Objective	To test the effectiveness of a tutorial to aid students in argument comprehension, specifically to aid evaluation of argument during reading (p.208). 'The effect of reading goal was also examined as a factor to improve readers' argument comprehension' (p.216).
Rationale	'Many high-school graduates have difficulty forming and understanding written arguments' (p.206). Attention to discourse markers which can help to appropriately connect the various argument elements into a coherent structure (p.208). The results of Experiment 1 (Larson <i>et al</i> , 2004, Exp.1) suggested that students may benefit from instruction in evaluating argument when reading.
Theoretical and research background	Theory: the use of argument schemas for comprehension of arguments, eg, Toulmin, 1958; claim supported by reasons which can also be supported. Research: in argument comprehension, argument production ability. The study also draws on prior research that has shown 'that readers' representation of the structure of narrative and expository text is aided by signals that clarify the structure' (p.207).
Study design	Individual RCT – no details. 2x2 design: 2 training (argument tutorial or no-feedback practice) x 2 reading goal (comprehension vs. rebuttal) between subjects design. 'Participants were randomly assigned to one of four conditions when they arrived for the experiment' (p.218).
Outcome measures	Participants were given four texts (from Exp. 1 – see p.217 for details) and asked to identify the main claims and reasons. They had to write down the author's main claim and list any reasons mentioned to support the claim (p.219). 'The Comprehension participants were asked to list an unmentioned reason to support the author's main claim while Rebuttal participants were asked to list an unmentioned rebuttal to the author's main argument. After completing the packet, participants rated the strength of each of the author's reasons' (10-point scale from very weak to very strong with justification).
Study details	
Participants	n=70 native English-speaking students from an introductory-level Psychology class at Northern Illinois University.
Intervention	The short tutorial 'defined key argument terms, challenged common misconceptions about arguments, and explained a series of steps to comprehend written arguments. Participants were provided an example and an opportunity to practice the presented steps' (p.216). See pp.217-8 for a full description of the intervention. n = not stated.
Control	Participants received no instruction but received an opportunity to practice the task for the same amount of time. n = not stated.
Outcome	
Results	'there was a significant Training x Reading goal interaction' (p.219)
Reviewer summary	This is an individual RCT with a factorial design and post-tests only.
Comments	No raw data (means, sds and ns), therefore not possible for reviewer to re-calculate effect sizes. Also not stated how many participants in each of the four arms of the experiment. Therefore exclude from in-depth review (Exclude 1 – insufficient data).

Nussbaum, E.M., Hartley, K., Sinatra, G.M., Reynolds, R.E. and Bendixen, L.D. (2004) Personality interactions and scaffolding in on-line discussions. <i>Journal of Educational Computing Research</i>, 30(1-2): 113-136.	
Country of origin	USA
Setting	University undergraduate introductory educational psychology class
Objective	To evaluate the effectiveness of a tool ('note starters') designed to scaffold and improve the level of reflection and argumentation in on-line discussions on students' counter argumentation, and the interaction of this tool with students' personality characteristics (p.114). The hypotheses were: (1) 'Low levels of assertiveness or openness to ideas, or high levels of anxiety, would reduce levels of disagreement' (p.118) (2) '(the) instructional interventions (note starters and cases) would increase levels of disagreement by prompting students to think more deeply' (p.118)
Rationale	
Theoretical and research background	Theoretical background in personality research (McCrae and Cost, 1979, 1992); empirical research in on-line discussion (e.g. Marttunen, 1998; Marttunen and Laurinen, 2001); empirical research linking argumentation variables to learning outcomes, improvements in writing skills and better individual problem solving ability. (p.115).
Study design	Individual RCT, with participants randomly assigned to groups and groups randomly assigned to conditions. (p.119).
Outcome measures	On-line discussion in groups of four (each student posted four messages, one initial response to the question and three responses to the other group members' initial posting) and on-line survey (personality inventory) completed at the beginning of the study to assess individual students' personalities.
Study details	
Participants	48 undergraduates enrolled in an introductory educational psychology class with self-reported moderate levels of computer experience; 83% female.
Intervention	'Note starters': a menu of phrases, from which students choose a phrase to begin the first sentence of a discussion note. They are a form of scaffolding intended to encourage students to think more deeply through the construction of explanations and arguments (p.116). See p.119 and Fig 1 for details about 'note starters'.
Control	No 'note starters', with or without elaborated cases. 'Elaborated cases' are 'detailed scenarios about a classroom situation contained in the initial problem statement. Cases may make the problem situation more meaningful to students and, by increasing engagement, increase the number of problem-centred moves' (p.117). 2x2 crossed design – 4 conditions – with or without note starters and with questions or elaborated cases.
Outcome	
Results	There was a higher percentage of disagreement in the note starter condition than in the control condition for both subcategories of disagreement as well as overall. Using computed average disagreement scores for each topic from 0 (consistent agreement) to 1 (consistent disagreement) there was a significant main effect for note starters ($F = 4.6, (1,37) p < 0.05$). In addition note starters were shown to be most beneficial for students who had low levels of assertiveness, openness to ideas and anxiety. 'Note starters appeared to encourage students who were not naturally curious and inquisitive (i.e. open to ideas) or assertive to consider opposing viewpoints. However, students who were anxious benefited less from note starters' (p.125). The results differ from previous studies (e.g. Marttunen 1998), which

	appear to suggest that students tend to agree with each other during online discussions at university level.
Reviewer summary	A 2x2 study design, with randomized allocation on an individual basis. Using an online discussion format, the study sets out to determine whether an intervention consisting of note starters actually improves students' ability to counter-argue. It concludes that the intervention is successful, but that there are different degrees of success according to the personality characteristics of the participants.
Comments	A small-scale study. The intervention is minimal: phrases used to stimulate counter-argument. The results are positive, but the size of the sample and the minimalism of the intervention suggest that claims are suitably modest, and that further research is called for both in note starter technique, between types of personality/learner, and within or without online discussion. Insufficient raw data given (means, standard deviations, numbers), therefore exclude from in-depth review (Exclude 1). In addition intervention is too short (Exclude 2).

Nussbaum, E.M. (2005) The effect of goal instructions and need for cognition on interactive argumentation. <i>Contemporary Educational Psychology</i>, 30(3): 286-313.	
Country of origin	USA
Setting	Introductory educational Psychology classes
Objective	<p>To 'examine the effect of goal instructions on students' reasoning and argumentation in an interactive context (discussing a topic online)' (abstract). To examine the effect of goal instructions (making explicit the goals of an activity) on students' argumentative writing in an interactive discussion context (Web-based discussions); to explore the effect of the need for cognition (NFC) and to examine the interactions between cognition and goal instruction. The six hypotheses are given on page 294. 'In summary, it is hypothesised that a goal 'to persuade others' will result in more adversarial discourse; the Goal 'to explore' will result in more explanatory discourse and – when combined with more specific goals – in better arguments. Specific goals will have main effects on specific argumentation elements (i.e. number of reasons, counterarguments, or rebuttals). Finally, 'need for cognition will be associated with greater depth and complexity' (p.294).</p> <p>Independent variables: NFC and different type of goal instruction. Dependent variables: type and complexity of arguments. Details given relating to what is meant by 'argument': 'a series of claims that bear logical relations to one another, specifically relations of support or opposition' (p.291). The framework for analysing arguments also given: global claim, supporting claims, opposing claims, contingent claims, divergent claims. Note that the framework differs from Toulmin's model where arguments consist of global claim, supporting grounds and a warrant (linking grounds to global claim).</p>
Rationale	Rationale for study based on assertion that 'how best to encourage students to generate complex arguments that examine multiple sides of an issue' is an 'unanswered question' (p.287).
Theoretical and research background	Background to argumentation as a 'device for promoting deep and elaborative learning' (p.287). Empirical research examining 'the effect of goal instructions on students' argumentative writing' (p.287) (e.g., Page-Voth and Graham, 1999) including comparing the effects of general goals with the effects of specific goals.
Study design	Cluster RCT: 'Students were placed randomly in groups of 3' (p.295). 'Randomization of subjects to both discussion group and condition helped eliminate any internal confounds and provided experimental control' (p.296). Therefore participants were allocated as individuals to the groups but the groups were allocated to the conditions. 3x3 randomized design using 224 undergraduate students (180 constituted the final sample). Students were randomly assigned to three conditions. Specifically, the design was 3x3 factorial, with a general goal (Explore, Persuade or None) crossed with a specific goal (Reasons, Counterarguments, or None).
Outcome measures	Two surveys administered by the Web-CT (electronic discussion board): need for cognition and prior attitudes towards television violence. Both surveys on 5-point Likert scale with responses ranging from 1 (strongly disagree) to 5 (strongly agree). NFC form: 18 items. Prior attitudes towards TV violence: 12 items (p.295). A conceptual framework for analysing arguments was developed. 'In this framework, an argument consisted of a series of claims that bear logical relations to one another, specifically relations of support or opposition. The framework was designed to apply to interactive arguments, where one

	<p>student may initially make a claim that frames the discussion and other students (as well as the first student) provide additional claims in response to the first claim' (p291).</p> <p>Each discussion note was scored for type of argumentation claims made (supporting, contingent, divergent) and level of claim (up to 3).</p>
Study details	
Participants	<p>224 undergraduates enrolled in five different sections of an introductory educational psychology course. Only 208 randomized due to some of the participants not completing the initial survey. Some participants did not complete all elements of the study and were therefore excluded leaving a final sample of 180 (see p 298 for details of how this final figure was arrived at). Participants (in final sample) had the following characteristics: 79% female; 89% Caucasian; ages ranged from 19 to 58 (median age was 27). Most of the participants were majoring in elementary education (56%), secondary education (27%) or some other type of education program (17%). The students were primarily juniors (62%) – ie third years – but 17% were sophomores (2nd years) and 21% seniors.</p>
Intervention	<p>All participants used the discussion board to discuss the following question: 'Does watching TV cause children to become more violent?' The question was posted on the notice board along with additional instructions. One person in the three-person group was randomly assigned to post the first note, after which the other two could post their notes. The additional instructions varied by goal condition which was one of the following: explore, persuade, reasons, counterarguments.</p>
Control	<p>This was provided by the 3x3 design in which the third group acted as a control with no interventions. The design was 3x3 factorial with general goal (Explore, Persuade, None) crossed with a specific goal (Reasons, Counterarguments, None).</p>
Outcome	
Results	<p>'The goals to (a) persuade and (b) generate reasons had the strongest effects on students' argumentation. Both resulted in more argumentation claims; the reason condition also increased the number of contingent claims by encouraging students to consider multiple factors. As hypothesised, persuasion resulted in more opposition and debate, as did the counterargument goal condition' (p.306).</p> <p>Effect of goal condition: Persuade and Reasons conditions – significant effect – $d = 0.42$ for persuade ($p < 0.001$) and $d = 0.60$ for reasons ($p < 0.001$).</p> <p>'Need for cognition was associated with more total argumentation and also with deeper or more complex arguments (providing reasons for reasons)' (p.306).</p> <p>The reason condition resulted 'in more contingency and divergence, and the persuasion condition in more opposition' (pp.300-1). The counter-argument condition resulted in more oppositional claims and the explore condition in more divergent claims. 'These results are generally in accord with the prior hypotheses, except that the general goal 'to persuade' was more effective than the specific goal to generate/refute counterarguments. The hypothesis that general goals (i.e. to explore) would have an effect by interacting with specific goals was also not supported' (p.301).</p> <p>In summary: 'the persuade condition resulted in debates, the reason condition in responses that integrated different sides of the issue together and considered multiple factors, and the explore condition in the most divergence' (p.306).</p> <p>Counter-argument appeared to be less successful in the interactive, online context. (p.307).</p>
Reviewer summary	

Comments	<p>The need for cognition (reasoning) has not been linked to the generation of arguments until this study, claims the author, although it has long been associated with the reading and comprehension of arguments. Nussbaum concludes that the interactive context for discussion may have led to better argumentation in this case [but see comments below]. The conclusion appears to be that better reasoning is linked to a higher profile for the function of argument (ie persuasion, in this case), e.g., 'persuasion resulted in deeper and more complex arguments' (p.306).</p> <p>The researcher also concludes that the online asynchronous nature of the discussion may have allowed more reflection than in face-to-face discussion, which may have contributed to the results of the study.</p> <p>The limitations of the study are acknowledged, but it is also claimed the 'real power of this study is in documenting what sort of goal instructions are the most effective in this context' (p.309). The reasoning instructions may have fostered an exploratory 'argumentation frame' (ibid.) and it is these instructions, claims the researcher, that have produced such exploratory, interactive discourse.</p> <p>The researcher states the tentativeness of the assumption that web-based/online discussion is more interactive and more constructivist than studies of individual student writing, and suggests that 'it is an open question...whether goal instructions will have an effect in an interactive, on-line environment' (p.289).</p> <p>Authors' conclusion: 'finding the right goal instructions to trigger the appropriate argumentation frame is important in creating interactive discourse that is educationally productive' (p.309) and concentration on the functions and reasons for argument is more effective in making argumentation more complex than specific concentration on counterargument or other heuristic scaffolding devices.</p> <p>Course instructors not informed of the nature of the interventions but the author taught one of the sections (possibility of bias, teacher effect).</p> <p>Scoring of notes was undertaken blind to condition. Groups where not all 3 students posted notes or which contained less than 80% of the requested notes were eliminated as were outlying cases – therefore possibility of bias introduced here. Multilevel modelling used to control for clustering effects, therefore two levels of analysis – student level and group level.</p> <p>Significance level set at 0.01 because of number of tests performed.</p>
----------	--

Sanders, J.A., Wiseman, R.L. and Gass, R.H. (1994) Does teaching argumentation facilitate critical thinking? <i>Communication Reports</i>, 7(1): 27-35.	
Country of origin	USA
Setting	Two major western universities
Objective	<p>To examine whether ‘training in argumentation facilitated critical thinking by enhancing the ability to discern strong and weak arguments and whether such training increased the need for cognition, argumentativeness, perceived arguing effectiveness and decreased verbal aggressiveness’ (p.27).</p> <p>Hypotheses:</p> <ol style="list-style-type: none"> (1) ‘Training in argumentation enhances one’s ability to discern weak and strong arguments’ (p.28) (2) ‘Subjects trained in argumentation will self report higher levels of argumentativeness than subjects not trained’ (p.29) (3) ‘Subjects trained in argumentation will self report higher levels of need for cognition than subjects not trained’ (p.29) (4) ‘Subjects trained in argumentation will self report more positive perceptions of arguing effectiveness than subjects not trained’ (p. 29) (5) ‘Subjects trained in argumentation will self report lower levels of verbal aggressiveness than subjects not trained’ (p.29)
Rationale	<p>Very little empirical research had previously been ‘devoted to examining effectiveness in teaching critical thinking’ (p.28). Little empirical research about the influence of critical thinking instruction on warrant evaluation (the authors refer to the work by Follert and Colbert,1983). Few studies had ‘examined the relationship between critical thinking and argument instruction’ (p.28). Therefore a ‘paucity’ of research created a need for ‘more focused investigation’ (p.28).</p>
Theoretical and research background	<p>The authors had previously reported their own empirical research demonstrating that argumentation instruction improves critical thinking (Gass <i>et al</i>, 1990). Research identifying ‘argumentativeness and verbal aggressiveness as important variables’ related to argument style (Infante, 1982, 1987; Infante and Wigley, 1986). Prior research which demonstrated that ‘persons who are highly argumentative but low in verbal aggression are perceived as more competent communicators’ (Infante <i>et al</i>, 1990) and more effective at influencing others (Infante and Gordon, 1987) (p.29). Prior research which suggests that ‘need for cognition’ – ‘an individual’s tendency to engage in cognitive elaboration, e.g., to think about and reflect upon ideas, issues, topics, and messages (Cacioppo and Petty, 1982)’ (p.29) – is ‘positively related to argumentativeness and negatively related to verbal aggressiveness’ (p.29).</p>
Study design	<p>Survey of students who received instruction in either argumentation or introductory interpersonal communication using responses to questionnaires, therefore design is ‘case control’.</p>
Outcome measures	<p>Questionnaires given out in the first and last weeks of term assessing argument perception, perceived arguing effectiveness, argumentativeness, need for cognition and verbal aggressiveness.</p> <p>Three different orders of the arguments and scales were used.</p> <p>The measure of argument effectiveness consisted of three items: convincing-unconvincing, persuasive-unpersuasive, and effective-ineffective.</p> <p>Need for cognition scale: 18 items measuring respondents’ predispositions toward handling complex vs. simple problems.</p> <p>Argumentativeness scale: 20 items designed to assess respondents’ approach and avoidance feelings toward arguing in conversational settings.</p> <p>Verbal aggressiveness scale: 20 items reflecting respondents’ tendencies to attack the self-concepts of other people.</p>

	Self-assessment of argumentation skills: 5 items.
Study details	
Participants	Total of 357 students from two major western universities. Mean age = 21.7 (sd = 5); 51.4% female; 50.3% Euro-Americans. All participants were enrolled in general education courses.
Intervention	299 experimental group members who were receiving instruction in argumentation. No details given.
Control	58 comparison group members receiving instruction in introductory interpersonal communication. No details given.
Outcome	
Results	<p>Results revealed that argument instruction enhanced the ability to discern weak example and causal arguments, increased perceived arguing effectiveness and decreased verbal aggressiveness.</p> <p>Table 1: ANCOVA – to determine the influence of argumentation instruction on: need for cognition; argumentativeness and verbal aggressiveness. Statistically significant. Difference for verbal aggressiveness scale, suggesting that argumentation instruction reduced respondents' tendency for verbal aggressiveness.</p> <p>ANCOVA – to determine the influence of argumentation instruction on respondents' perceptions of strong and weak arguments using example, analogy and causal reasoning. There were no statistically significant effects for argument instruction on the perceptions of strong arguments. Weak arguments were judged less cogent by respondents receiving argument instruction.</p> <p>ANCOVA – to determine the influence of argumentation instruction on respondents' perceptions of their own argumentation skills. Statistically significant. Difference between the experimental and comparison groups – respondents receiving argumentation instruction assessed their argumentation skills more favourably than those respondents not receiving argumentation instruction.</p>
Reviewer summary	

Steinkuehler, C.A., Derry, S.J., Levin, J.R., Kim, J.-B. (2000) <i>Argumentative reasoning in online discussion</i>. Paper presented at the Annual meeting of the American Educational Research Association, April, 2000.	
Country of origin	USA
Setting	An educational psychology course in a School of Education
Objective	To compare the effects of three forms of online discussion on memory, belief change and argumentation skill (p.1). Hypotheses: (1) Online discussion was expected to produce greater improvement in students' ability to reason argumentatively (2) Individualised forms of online instruction derived from the cognitive memory literature were expected to produce relatively better recall of the text (3) 'Perceptions of participation and importance of the topic for consideration would be greater for online discussion relative to the two individual conditions and would be correlated with reported and actual opinion change, which would also be greater for online discussion' (p.7)
Rationale	Previous research (see below) has led to unanswered questions, such as, is discussion equally productive in forums other than face-to-face ones? Therefore the research in the present paper was undertaken to address such questions.
Theoretical and research background	Theoretical traditions supporting the notion that social interaction is the primary catalyst for cognitive development (Piaget, 1976 and Vygotsky, 1978). Empirical work suggesting that group interaction in educational settings seems particularly well suited for improving argumentation skills, e.g. Kuhn <i>et al</i> (1997), found that discussion is the underlying catalyst for improvement in argumentation.
Study design	Cluster RCT. Students individually randomized to one of the three conditions, and then placed in pairs. The pair was the unit of analysis, therefore this is a cluster trial. The scores of the pairs were averaged together 'resulting in a final data set consisting of 15 data points in each condition with the pair as the unit of analysis' (p.7).
Outcome measures	Argument change from pre-test to post-test, transfer of argument skills, text recall, reported and directly assessed opinion change, and perceptions of productivity and participation. The two pre-test measures were: brief written essay and an opinion scale to assess opinion and opinion change. The six post-test measures were: short-answer questionnaire designed to assess recall for the given pro/con text; scale designed to measure opinion change; participants wrote a second essay; scale designed to assess opinions; participants wrote another essay to assess transfer of improvements in argumentative reasoning to a topic beyond the activity; scale to measure productivity and participation. Segmentation: topic and transfer essays were coded by two coders. Agreement ranged between 0.88 and 90%. Analytic scheme: topic and transfer essays were coded by two coders into three broad categories: functional arguments, non-functional arguments, nonjustificatory arguments. Agreement ranged between 0.82 and 0.9.
Study details	
Participants	105 undergraduates enrolled in an educational psychology course. Most of the participants were in the School of Education; other subjects included communicative disorders, occupational therapy, and nutrition. Attrition = 15.
Intervention	Reading of a pro/con text followed by online discussion in pairs n = 30
Control	Control i: reading of a pro/con text followed by a form of individualised study technique derived from cognitive memory literature: self-explanation n = 30. Control ii: reading of a pro/con text followed by a form of individualised

	<p>study technique derived from cognitive memory literature: repeated summarisation and study n = 30. See pp.10-11 for a full account of the discussion, self-explanation and summarisation conditions.</p>
Outcome	
Results	<p>Across conditions, there was a statistical decrease from pre-test to post-test in the proportion of nonjustificatory statements made, with no statistical differences among the three conditions. There was no decrease in the proportion of non-functional statements made, either across or among conditions. Across conditions, there was a statistical increase in the number of metacognitive statements made, with no difference among conditions. In summary: across conditions, participants' argumentative reasoning improved from pre-test to post-test. However, there were no statistical differences among conditions ('Of the six variables used to investigate improvement in argumentative reasoning, only one variable differentiated among conditions – the number of statements made that contained or acknowledged the relevance of evidence. In particular, summarisation produced greater use of evidence than did self-explanation and discussion' p.21). For recall test the difference among the three conditions was statistically significant, $p < 0.001$.</p>
Reviewer summary	<p>This is a cluster RCT consisting of pre-tests, the reading of two arguments, an intervention or one of two control conditions, and post-tests. The intervention/control condition was the only variable that differed across conditions. The intervention lasted 45 minutes in each condition.</p>
Comments	<p>15 participants excluded from analysis – not clear at which point they were excluded but probably after randomization, but not clear how these 15 students had been 'paired'. Excluded from in-depth review due to short length of intervention (45 minutes) (Exclude 2).</p>