Review of Educational Research
December 2008, Vol. 78, No. 4, pp. 1102–1134
DOI: 10.3102/0034654308326084

Instructional Interventions Affecting Critical Thinking Skills and Dispositions: A Stage 1 Meta-Analysis

Philip C. Abrami, Robert M. Bernard, Evgueni Borokhovski, Anne Wade, Michael A. Surkes, Rana Tamim, and Dai Zhang
Concordia University, Montreal, Quebec, Canada

Critical thinking (CT), or the ability to engage in purposeful, self-regulatory judgment, is widely recognized as an important, even essential, skill. This article describes an ongoing meta-analysis that summarizes the available empirical evidence on the impact of instruction on the development and enhancement of critical thinking skills and dispositions. We found 117 studies based on 20,698 participants, which yielded 161 effects with an average effect size (g+) of 0.341 and a standard deviation of 0.610. The distribution was highly heterogeneous (Q_T = 1,767.86, p < .001). There was, however, little variation due to research design, so we neither separated studies according to their methodological quality nor used any statistical adjustment for the corresponding effect sizes. Type of CT intervention and pedagogical grounding were substantially related to fluctuations in CT effects sizes, together accounting for 32% of the variance. These findings make it clear that improvement in students' CT skills and dispositions cannot be a matter of implicit expectation. As important as the development of CT skills is considered to be, educators must take steps to make CT objectives explicit in courses and also to include them in both preservice and in-service training and faculty development.

KEYWORDS: critical thinking, meta-analysis, achievement.

Critical thinking (CT), or the ability to engage in purposeful, self-regulatory judgment, is widely recognized as an essential skill for the knowledge age. Most educators would agree that learning to think critically is among the most desirable goals of formal schooling. This includes not only thinking about important problems within disciplinary areas, such as history, science, and mathematics, but also thinking about the social, political, and ethical challenges of everyday life in a multifaceted and increasingly complex world. Bailin and Siegel (2003) argued that "critical thinking is often regarded as a fundamental aim and an overriding ideal of education" (p. 188), and Sheffler (1973) contended that "critical thinking is of the first importance in the conception and organization of educational activities" (p. 1).

As well as being better students, a short-lived advantage, critical thinkers have a better future as functional and contributing adults. Let two examples suffice. Parents who can think through the challenges of child rearing and who act accordingly have a better chance of propagating a societal future of harmony and well-being. At a broader societal level, a democracy composed of citizens who can think for themselves on the basis of evidence and concomitant analysis, rather than emotion, prejudice, or dogma, is a plus—in fact, it sustains, builds, and perpetuates the democracy.

Educators are not alone in their concern about the urgency of teaching and learning CT. In the United States, "a national survey of employers, policymakers, and educators found consensus that the dispositional as well as the skills dimension of critical thinking should be considered an essential outcome of a college education" (Tsui, 2002, 740–741). In Canada, the cross-country consultation on the Canadian federal government's highly influential "Innovation, Knowledge, and Skills" policy recommended that schools, colleges and universities "should promote critical thinking . . . at all levels of education" (Government of Canada, 2002, n.p.). The Conference Board of Canada expressed the need for Canadians to improve their CT skills to strengthen Canada's innovation profile and competitive advantage in the knowledge-based global economy (Bloom & Watt, 2003).

What Is CT?

CT is by no means a new concept, but part of the problem facing practitioners and researchers alike is that it is a complex and controversial notion that is difficult to define and, consequently, to study. Furthermore, the tools of implementation (instructional interventions) are difficult to operationalize. Many definitions of CT have been proposed (Ennis, 1962, 1987; Facione, 1990a; Kurfiss, 1988; Lipman, 1991; Paul & Binker, 1990; Scriven & Paul, 1996; Siegel, 1988). One high-profile definition was developed by an American Philosophical Association Delphi panel of 46 experts, including leading CT scholars such as Ennis, Facione, and Paul:

> We understand critical thinking to be purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation, and inference, as well as explanation of the evidential, conceptual, methodological, criteriological, or contextual considerations upon which that judgment is based. . . . The ideal critical thinker is habitually inquisitive, well-informed, trustful of reason, open-minded, flexible, fair-minded in evaluation, honest in facing personal biases, prudent in making judgments, willing to reconsider . . . and persistent in seeking results which are as precise as the subject and the circumstances of inquiry permit. (Facione, 1990a, p. 3)

The Delphi Committee identified six skills (interpretation, analysis, evaluation, inference, explanation, and self-regulation), 16 subskills, and 19 dispositions (including inquisitiveness, open-mindedness, understanding others, and so on) that they associated with CT. These skills and dispositions provide a complex normative framework for understanding and assessing the qualities of human cognition.

Within CT theory and research, different conceptual traditions have emerged. According to psychological views, CT requires gaining mastery of a series of discrete skills or mental operations and dispositions that can be generalized across a variety of contexts. These skills include concepts such as interpreting, predicting,

analyzing, and evaluating. The obvious and primary appeal of the skills discourse on CT involves its transfer between contexts. D. N. Perkins and Salomon (1988) reflect the skills approach by claiming, "Students often fail to apply knowledge and skills learned in one context to other situations. With well-designed instruction we can increase the likelihood that they will" (p. 22).

Bailin and Siegel (2003) reflect the philosophical tradition by arguing that psychological conceptions of CT are problematic for three reasons: (a) It is impossible to determine whether particular mental operations correlate with particular cases of good thinking, (b) there is no particular set of procedures that is either necessary or sufficient for CT, and (c) terms denoting thinking (for example, *evaluating, interpreting, analyzing*) refer not to mental operations or processes but rather to different tasks requiring thinking (p. 181). They suggest CT is a normative concept that requires mastery of context-specific knowledge to evaluate specific beliefs, claims, and actions. From this perspective, *critical* normally means making sound judgments and claims that meet epistemologically acceptable standards. Siegel (1988) characterizes CT as being "appropriately moved by reasons" (p. 23) and emphasizes particular criteria that must be satisfied for reasons to count as appropriate. Paul (1990) adopts a similar view of CT by suggesting it requires the ability and disposition to evaluate beliefs effectively and to identify and assess their underlying assumptions.

How Is CT Measured?

The question of how CT is measured is as important as understanding how CT can be taught and learned. How will we know if one intervention is more beneficial than another if we are uncertain about the validity and reliability of outcome measures? Researchers in education and human cognition have employed numerous assessment tools that cover a broad range of formats, origins, psychometric characteristics, areas of application, and scope of constructs to be measured. This situation presents serious challenges in identifying, categorizing, and evaluating learners' outcomes in empirical research on CT in education. Even when researchers explicitly declare that they are assessing CT, there still remains the major challenge of ensuring that measured outcomes represent the construct of CT according to the operational definition adopted by reviewers.

There is also concern about the psychometric properties of the existing standardized measures of CT, such as the Watson-Glaser Critical Thinking Appraisal (WGCTA; G. Watson & Glaser, 1980), Cornell Critical Thinking Test (Ennis & Millman, 1985a, 1985b), California Critical Thinking Skills Test (Facione, 1990), and California Critical Thinking Disposition Inventory (Facione & Facione, 1992). A great deal of research has been undertaken to establish the validity and reliability of these measures—more for the WGCTA because it is the oldest—and the overall results are inconsistent (to say the least). In a recent analysis, Bernard et al. (2008) explored the intercorrelational structure of the subscales of the WGCTA, performing a factor analysis of the means of 70 studies of two versions of the test, and found that the best interpretation of these versions is represented by the total score rather than by the individual scores on the five subscales.

Is CT a Generic Skill?

No one would argue that CT is applicable across a range of disciplinary areas, but there is little consensus about whether it is a set of generic skills that apply across subject domains (engineering, arts, science) or whether it depends on the subject domain and context in which it is taught (Ennis, 1989). If it is generic, then CT should be taught in specialized courses that focus on CT skills (Royalty, 1991; Sá, Stanovich, & West, 1999). If CT is dependent on subject matter, then it should be learned by tackling concrete problems in specific disciplines (Halliday, 2000; G. Smith, 2002).

The *generalist* view, supported by Siegel (1988), contends that identifying informal fallacies of reasoning, such as *post hoc* and *hasty generalization* fallacies, is transferable between different contexts. In his view, CT is partially defined by detecting such fallacies without regard to specific subject matter, because errors of reasoning are based on argument design rather than content. Alternatively, the *specifist* position, represented in the views of McPeck (1981), argues against general CT capacities on logical grounds, because thinking is always linked to a specific subject domain:

> It makes no sense to talk about critical thinking as a distinct subject and therefore cannot profitably be taught as such. To the extent that critical thinking is not about a specific subject X, it is both conceptually and practically empty. (p. 5)

What Do We Know About Teaching Students to Think Critically?

It is not surprising that a search of the literature reveals thousands of documents devoted to issues related to the teaching of CT. There have been several reviews focusing on the effects of instruction on CT. For example, Adams (1999) summarized studies that reviewed changes in the CT abilities of professional nursing students. Allen, Berkowitz, Hunt, and Louden (1997, 1999) studied the impact of various methods for improving public communication skills on CT. Bangert-Drowns and Bankert (1990) reported some effects of explicit instruction for CT on measures of CT. Follert and Colbert (1983) analyzed research on debate training and CT improvements. Pithers and Soden (2000) reviewed methods and conceptions of teaching likely to inhibit or enhance CT. McMillan (1987) considered the effects of instructional methods, courses, programs, and general college experiences on changes in college students' CT. The Assessment and Learning Research Synthesis Group (2003) at University of London reviewed the impact of information and communication technologies on creative and CT skills.

What Are Some Instructional Approaches to Teaching CT?

In our analysis of CT instructional approaches, we will use Ennis's (1989) CT typology of four courses—general, infusion, immersion, and mixed—for classifying and describing various instructional interventions. In the general course, CT skills and dispositions are learning objectives, without specific subject matter content. In contrast, content is important in both the infusion and immersion approaches. CT is an explicit objective in the infusion course but not in the immersion course. In the mixed approach, CT is taught as an independent track within a specific subject matter. These four approaches—general, infusion, immersion, and mixed—will be assessed in this review for their instructional efficacy.

According to Ennis (1989), the *general* approach attempts to teach CT abilities and dispositions separately from the presentation of the content of existing subject matter offerings, with the purpose of teaching CT. Examples of the general approach usually do involve some content but do not require that there be content.

Using a quasiexperimental design, Riesenmy, Mitchell, Hudgins, and Ebel (1991) taught self-directed CT directly (using Ennis's [1989] general method) to 70 fourth- and fifth-grade students in St. Louis public schools. The authors expected that students who were taught the roles of four modes of thinking (task definer, strategist, monitor, and challenger) would perform better on a problem-solving posttest that demanded both lateral and vertical transfer of general thinking skills. This prediction was fulfilled by the results. Three groups of treated students outscored the control group: The group that wrote immediate posttests had greatly superior scores on average, a second group tested 4 weeks later outscored the controls, and a third group, tested 8 weeks later, also outperformed the control students.

The *infusion* of CT requires deep, thoughtful, and well-understood subject matter instruction in which students are encouraged to think critically in the subject. Importantly, in addition, general principles of CT skills and dispositions are made explicit.

Zohar, Weinberger, and Tamir (1994) used the infusion method in the Biology Critical Thinking Project to support seventh-grade biology students in Israel in developing their CT skills (which included recognizing logical fallacies, distinguishing between experimental findings and conclusions based on findings, identifying tacit and explicit assumptions, avoiding tautologies, isolating variables, testing hypotheses, and identifying relevant information). A true experimental design was used to test the efficacy of the program on two dependent variables developed for this study, a general CT test (administered before and after the training) and a biology CT test (posttest only). Average scores were reported for nearly 500 students, and the results were highly favorable for the program, as experimental students registered much higher average gain scores on the biology CT test and also outperformed the control group on the general CT test.

In the *immersion* approach, subject matter instruction is thought-provoking, and students do get immersed in the subject. However, in contrast to the infusion approach, general CT principles are not made explicit.

An example of CT immersion during instruction is provided by Kamin, O'Sullivan, and Deterding (2002), who used digital video case simulations followed by group discussions as an instructional method. One group of students who viewed the cases on video discussed the case online, a second group saw the videos and discussed them face-to-face, and a third group received a text account of the case (rather than a video) and participated in face-to-face discussions. Content analysis of the discussions was used to assess CT demonstrated by each group; results showed that although video presentation seemed to facilitate CT, the online discussion group scored highest. The authors suggested that the written discussion format provided better opportunities for the students to concentrate on articulating their ideas.

The *mixed* approach consists of a combination of the general approach with either the infusion or immersion approach. Under it, students are involved in subject-specific CT instruction, but there is also a separate thread or course aimed at teaching general principles of CT.

In a successful attempt at implementing Ennis's (1989) mixed CT instructional strategy, McCarthy-Tucker (1998) reported that high school freshman and sophomore students in English and algebra who received instruction in formal logic as a supplement to their curricular instruction showed much greater improvement (from pretest to posttest) on two standardized measures of thinking, the Test of Logical Thinking (TLT) and the Content Specific Test of Logic (CSTL), than untreated control participants.

Goals of This Systematic Review

This article describes an ongoing meta-analysis that summarizes the available empirical evidence we have examined to date on CT development and utilization in educational contexts. The core research question is: What instructional interventions have an effect on the development and effective use of CT skills and dispositions, and to what extent, and under what circumstances? Our objectives are as follows: First, we will summarize the evidence on the impact of instruction on CT. Second, we will examine how certain methodological aspects of individual studies (such as research design, type of CT measure, and the method of the effect size extraction) moderate the magnitude of this impact. We want to know, for example, whether the effect of instruction has generally been found to be smaller in true experiments and larger in preexperimental studies. We are also interested in how effect sizes vary with CT measures, expecting that the impact of instruction will be smallest on standardized tests and largest on teacher-made tests.

Third, the role of several substantive study features will be analyzed. In particular, we want to know how different types of instructional interventions affect CT skills, what impact pedagogical background (e.g., instructor training) has, and how calculated effect sizes vary with age and educational level and whether collaborative work was part of the treatment.

By summarizing our work completed to date, we hope to inform the research and practitioner communities of the findings and also to solicit feedback from scholars and educators about the next stages in the review, including alternative and fine-grained analyses. We hope eventually to develop an empirically validated model of effective CT instruction.

Method

Literature Search Strategies and Data Sources

An extensive literature search was designed to identify and retrieve primary empirical studies relevant to the project's major research question. The databases searched were ERIC, ProQuest Dissertations and Theses, PsycInfo, ABI/Inform Global on ProQuest, AACE Digital Library, OCLC PAIS International, EBSCO EconLit, EBSCO Academic Search Premier, Social Science Index, CSA Social Services Abstracts, and CSA Sociological Abstracts. The bibliographies of review articles and previous meta-analyses were scanned. In addition, the Educational Testing Service Test Collection database (see http://www.ets.org/portal/site/ets/) and ERIC were consulted to locate a comprehensive list of standardized tests that measure CT. The Web of Science database was also consulted to locate studies exploring the factor structure of CT standardized assessment tools. In this regard, a cited reference search was performed on the following authors: Clifford,

Follman, Frisby, Harris, Johnson, Loo, Lowe, McMurray, Ross, Simon, and Whimbey.

The descriptor *critical thinking* was used when possible; otherwise, it was searched as a keyword. For noneducation databases, this concept was combined with the search statement *educat** or *teach** or *learn** or *student**. All results were limited to empirical studies when possible (e.g., the document type code 143 was used in ERIC), or the abovementioned search statements were combined with the keywords: *control group* or *compar** or *study* or *studies*.

Search results span the 1960s through 2005, although the entire year of 2005 has not been searched, as the final searches occurred in the summer of 2005. A total of 3,720 studies were found.

Although the bulk of studies have likely already been identified, in the next stage of this project, all searches will be updated and further searches will be conducted in the following sources: British Education Index; Australian Education Index; CBCA Education; Education: A Sage Full-Text Collection; PubMed; Francis; various databases on the Evidence Network; and various Web search engines, including Google and Google Scholar.

Inclusion and Exclusion Criteria and Review Procedure

Decisions regarding whether to retrieve an article were based on a review of study abstracts. Decisions about whether to include studies in the review were based on reading the full text of the article. Both decisions were made by two researchers working independently who then met to discuss their judgments and to document their agreement rate. The following inclusion criteria were used: (a) accessibility—the study must be publicly available or archived; (b) relevancy—the study addresses the issue of CT development, improvement, and/or active use; (c) presence of intervention—the study presents some kind of instructional intervention; (d) comparison—the study compares outcomes that resulted from different types or levels of treatment (e.g., control group and experimental group, pretest and posttest, etc.); (e) quantitative data sufficiency—measures of relevant dependent variables are reported in a way that enables effect size extraction or estimation; (f) duration—the treatment in total lasted at least 3 hr; and (g) age—participants were no younger than 6 years old. If any of these criteria were not met, the study was rejected.

Along with studies employing experimental and quasiexperimental designs, we also included studies that used preexperimental designs (e.g., one-group pretest–posttest designs). Our intent was not to eliminate methodologically less sound studies a priori but instead to code for research methodology. Including studies of variable quality meant we could determine empirically whether and in what ways the findings differed by research design and, as necessary, use weighted multiple regression to remove variance associated with study quality. In fact, the bulk of research on CT does not consist of true experiments with "randomized control trials." This is largely because CT research has been conducted in classrooms, where randomization is difficult to achieve. Our approach allowed greater flexibility and greater statistical power in the analyses, because potentially promising findings from more typical applications (e.g., studies with higher external validity) are not excluded in advance of conducting the review (Abrami & Bernard, 2008).

Abstracts and full-text research reports were each coded for inclusion or exclusion by two raters. An individual coder's ratings, at each stage, were specified to range from 1 (*the study is definitely unsuitable for the purposes of the project*) to 5 (*the study is definitely suitable for the purposes of the project*); the midpoint of 3 (*doubtful but possibly suitable*) was designated as a vote in favor of including the study. In other words, ratings from 3 to 5 suggested either the retrieval of the full-text document (at the abstract review stage) or inclusion of the study in further analyses (at the full-text review stage), whereas ratings of 1 or 2 meant the elimination of the study from further consideration. Interrater agreement rate at these two stages of the review was calculated and reported in two different ways: (a) as a coefficient of correlation between ratings given by independent coders across all reviewed papers and (b) as a percentage of studies, with respect to which both coders agreed whether to reject the study or to continue analyzing it.

The extent of uniformity between coders was also documented with regard to effect size extraction and to the calculation of study features. Each effect size was coded by two raters, and two agreement rates were produced: (a) a number between 50 and 100 was assigned to each study to reflect the degree of agreement between the raters with regard to how many effect sizes should be extracted from each study, and this number was averaged across studies; and (b) a similar procedure was applied with regard to agreement as to which calculation procedures should be used to determine each effect size. As for study features coding, each study was assigned a rating according to the percentage of the features on which the raters initially agreed; all disagreements were discussed until a final accord was negotiated. All agreement rates were averaged across studies, and the average rates are presented in the Results section below.

Measuring CT

The issue of measuring CT is a complex one. Our research team (Bernard et al., 2008) factor analyzed CT subscale weighted means from 60 data sets for the WGCTA. A strong general factor emerged for two versions of the WGCTA, and so for this review, we focused on global indices of CT only instead of on individual subscale means that are often reported in the literature where the WGCTA is used. Although we have not performed a similar analysis on other popular standardized CT measures, our findings and decision as to how to proceed with the WGCTA suggested that we should follow this strategy uniformly.

To account for variation in CT assessment tools, we decided to categorize them as follows.

Standardized tests. These are well-established measures of CT or particular thinking skills and dispositions, such as WGCTA, Cornell Critical Thinking Test, California Critical Thinking Skills Test, and California Critical Thinking Disposition Inventory.

Tests developed and evaluations conducted by a teacher. This category includes, for example, the content analysis of students' responses to interview questions, open-ended questions, and essay type of tasks teachers used to address CT skills development in their students.

Tests developed by researchers (one or more of the study authors). These are nonstandardized measures developed by a researcher for use in a particular study, for example, Bonk, Angeli, Malikowski, and Supplee (2001) and VanTassel-Baska, Zuo, Avery, and Little (2002).

Tests developed by researchers (one of the study authors) who also taught the courses in question. These are developed by a researcher who was also the teacher or instructor of record. For example, in one of our included studies, the researchers (Zohar & Tamir, 1993) developed the Critical Thinking Application Test to evaluate performance in reasoning skills.

Secondary-source measures. These instruments are usually adopted from other sources with or without modifications. Researchers may use previously developed (standardized or unstandardized) instruments or modify them to meet the requirements of their research setting. For example, Feuerstein (1999) adapted the Language and Media Test developed by an Australian research group (Quin & McMahon, 1993) to suit her Israeli study according to local media texts.

Effect Size Extraction

Effect size is a standardized metric expressing the difference in two group means (usually a control and a treatment). Cohen's d (1) is the biased estimator of effect size.

$$d_i = \frac{\bar{X}_{Experimental} - \bar{X}_{Control}}{S_{Pooled}} \tag{1}$$

There are also two modifications of this basic equation: one for studies reporting pretest data for both experimental and control groups and another for a single-group pretest–posttest design. In other cases (e.g., t tests, F tests, p levels), effect size is estimated using conversion formulas provided by Glass, McGaw, and Smith (1981) and Hedges, Shymansky, and Woodworth (1989).

To correct for bias in small samples, d was converted to the unbiased estimator g (Hedges & Olkin, 1985), as follows:

$$g \cong \left(1 - \frac{3}{4N-9}\right)d \tag{2}$$

The Q_T statistic was used to test for homogeneity of effect sizes (Hedges & Olkin, 1985). Q_T is a homogeneity statistic that is most commonly used in assessing a collection of effect sizes or correlation coefficients. When all findings share the same population value, Q_T has an approximate χ^2 distribution with $k - 1$ degrees of freedom, where k is the number of effect sizes or correlations. If the obtained Q_T value is larger than the critical value, the findings are determined to be significantly heterogeneous, meaning that there is more variability in the effect

sizes or correlations than chance fluctuation would allow around a single population parameter.

Q_T is a notoriously sensitive measure of heterogeneity that becomes more sensitive as sample size increases. Higgins and Thompson (2002), and more recently, Huedo-Medina, Sanchez-Meca, Marin-Martinez, and Botella (2006), recommend the use of I^2 as a complement to the interpretation of Q_T. I^2 represents heterogeneity in proportional terms as the percentage of variability in point estimates that is caused by heterogeneity rather than sampling error. Higgins and Thompson tentatively suggest the following interpretations of I^2: "Mild heterogeneity might account for less than 30 per cent of the variability in point estimates, and notable heterogeneity substantially more than 50 per cent" (p. 1553).

Study Features

To explain variability in effect sizes, coded study features were individually assessed and also entered collectively in weighted multiple regressions. The following methodological features were tested: (a) type of research design (preexperimental, quasiexperimental, or true experimental), (b) type of CT measure (standardized, teacher-made, researcher-made, teacher- and researcher-made, or secondary-source measures), and (c) effect size extraction method (calculated from descriptive statistics, estimated with no assumptions made, or estimated with assumptions). The following substantive features were coded: (a) age level of participants (elementary or 6–11 years, early secondary or 11–15 years, high school or 16–18 years, undergraduate education, graduate education, or adult learners outside of formal school settings), (b) intervention type (general, infusion, immersion, or mixed), (c) pedagogical grounding of the intervention (instructor receives special training for teaching CT; extensive observations of course activities were reported to describe their relevance to CT skills development; the course curriculum was described in detail, including how its components were linked to the objective of CT skills development; or CT was simply declared to be among course the objectives, with no provision of any supporting information), and (d) peer collaboration (collaborative learning was part of the course, or no indication of collaborative work was reported).

Results

To date, searches of ERIC, PsycInfo, ABI, and Dissertation Abstracts International revealed 3,720 studies that were potentially suitable for further scrutiny. Judging from the review of abstracts, we marked 1,380 studies for retrieval (for dissertations, a representative sample was selected), of which 158 were retained after full-text review. This report includes data from 117 studies that were selected for inclusion. This is a much larger number of studies than any other review has previously included. Interrater agreement was as follows:

- Abstract review, 94.9% ($r = .794$, $p < .01$)
- Full-text review, 94.6% ($r = .833$, $p < .01$)
- Agreement on numbers and categorization of effect sizes, 94.1%
- Agreement rate on effect size calculation, 98.5%
- Agreement rate on study features coding, 84.1%

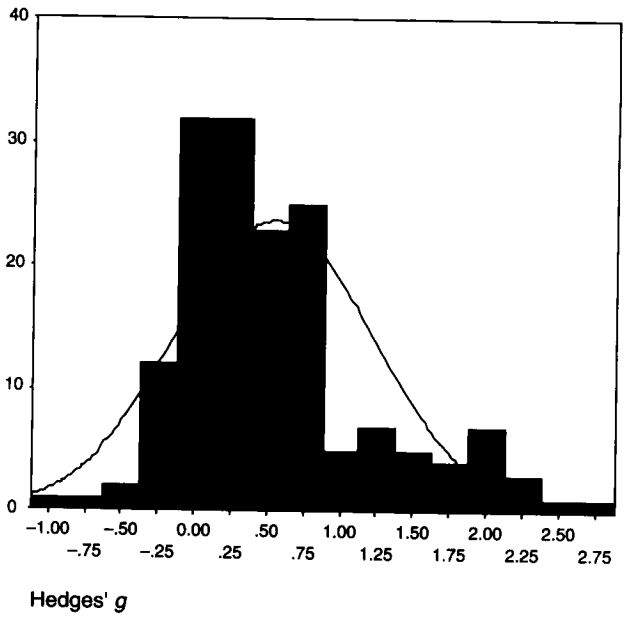

FIGURE 1. *Distribution of weighted effect sizes (Hedges' g).*

Overall Analysis

The 117 studies yielded 161 independent effect sizes, including 27 effect sizes from true experiments, 74 effects sizes from quasi experiments, and 60 effect sizes from preexperiments. Figure 1 shows the distribution of g, and Tables 1 and 2 show the general statistical analyses associated with it. The distribution is positively skewed (skewness = 1.606) and platykurtic (kurtosis = 2.876), with an average effect size (g+) of 0.341 (k = 161, N = 20,698) and a standard deviation of 0.610. The standard error of the mean is 0.01. This represents an average advantage of about one third of a standard deviation for the treatment over the control. The distribution is highly heterogeneous (Q_T = 1,767.86, p < .001). Because the I^2 estimate here is greater than 0.90, we conclude that the distribution suffers from severe heterogeneity. This substantially weakens any claim that the average effect size is representative of population parameters. Significant heterogeneity does, however, open up the possibility of exploring effect size variability in terms of other study features.

Overall, the moderate average effect is sympathetic with the view that instruction improves CT skills and dispositions. But the large variability means the finding is neither uniform nor consistent, requiring further exploration to determine whether methodological and substantive features explain the differences among study findings and can account for the wide range of findings (i.e., from −1.0 to +2.75).

TABLE 1
Overall mean effect size (g+) and statistics

		Effect size		*95% confidence interval*		*Test of null*
Overall outcome	k	g+	SE	*Lower*	*Upper*	z
Summary statistics	161	.34	.01	.31	.37	23.58*

*$p < .001$.

TABLE 2
Heterogeneity analysis (Q and I^2) for overall mean effect size

	Heterogeneity			
Overall outcome	Q *value*	df *(Q)*	p	I^2
Summary statistics	1767.86	160	.00	90.95

Outlier Analysis and Publication Bias

Outlier analysis. Outlier analysis seeks to determine if the removal of a certain number of effect sizes from a distribution of effect sizes greatly increases the fit of the remaining effect sizes to a simple model of homogeneity without drastically affecting substantive interpretation of the recalculated mean effect size. Three approaches are described by Hedges and Olkin (1985): one that involves a visual examination of a forest plot of the data, another that involves examining relative residuals, and a third that is based on the magnitude of the Q statistic for each effect size, in conjunction with what is known as a "one-study-removed" analysis. These methods will not always agree as to the outliers that should be considered for deletion, but according to Hedges and Olkin, they often do.

This review is marked by high heterogeneity of effect size (Q_T = 1,767.86) around an average effect size of $g+ = 0.34$. An examination of the results of the one study removed analysis revealed that the removal of any one effect size did not substantially affect $g+$. Only two studies on the negative tail raised $g+$ by as much as 0.03; one study on the positive tail lowered $g+$ by 0.02, and another lowered it by 0.04. All of these recalculated results remain within the 95th confidence interval of the original $g+$.

Examination of the Q statistic and relative residuals for each effect size revealed eight potentially outlying effect sizes, all related to large-sample studies (i.e., standard errors less than 0.10). All of these effect sizes had a Q statistic greater than 100. Three had a Q statistic above 200. Table 3 shows the changes to both $g+$ and to Q_T when three and eight effect sizes are removed.

While the reduction of Q_T is substantial, especially when eight effect sizes are removed (33%), neither removal comes close to producing a better-specified

TABLE 3
The effect of removing three and eight potential outliers from the distribution of effect sizes

Number removed	*Percentage removed*	*Effect size (g+)*	*Change in g+*	*Percentage change*	Q_T	*Percentage change*
0	0.0	.340	.0	0.0	1767.86	0.0
3	1.86	.414	+.074	+21.76	1578.92	−10.69
8	4.97	.418	+.078	+22.94	1183.34	−33.06

model (i.e., homogeneous). In addition, the removals produce an increase in *g*+, in both cases, of more than 20% (i.e., 22% and 23%, respectively). Both of these recalculated average effect sizes are above the 95th confidence interval of the original *g*+ of 0.34. Therefore, we decided not to remove effect sizes from the original distribution.

Publication bias. The question answered through the analysis of publication bias is, "Are there a significant number of studies with null results that have not been uncovered through a search of the literature to nullify the effects found in the meta-analysis?" Two statistical approaches are generally considered acceptable for assessing this bias. Classic fail-safe *N* determines the number of null effect studies needed to raise the *p* value associated with the average effect above a specified level of α (in this case, 0.05). This test, performed in Comprehensive Meta-Analysis (Borenstein, Hedges, Higgins, & Rothstein, 2005), revealed that 30,539 additional studies would be required to nullify the effect. The second test, Orwin's fail-safe *N*, estimates the number of missing null studies that would be required to bring the mean effect size of the meta-analysis to some specified trivial level. For the purpose, 0.10 was set as the trivial value. The number of missing null studies to bring the current mean effect to 0.10 is 388. Combined, these estimates suggest that there is little if any publication bias in this collection of effect sizes.

Methodological Features

Research designs. Table 4 shows the subgroup and total group effect sizes, and Table 5 contains the results of the heterogeneity analyses. The weighted means for subgroups ranged from +0.31 to +0.36. Although the variability within each of the subgroups is significantly heterogeneous, the variability explained among subgroups groups is not significant, $Q_B(2) = 2.28, p = .32$. Based on this, we decided to combine the data from all research designs in the subsequent analyses.

Abrami and Bernard (2008) discuss various approaches to dealing with study quality in systematic reviews. These approaches include treating study quality as an inclusion criterion, presenting and analyzing the results separately for each study quality category, weighting studies for quality, and testing and adjusting (as needed) for differences in effect size caused by study quality. The advantages of the latter approach include increasing review sample size, maximizing representativeness and generalizability, and statistically controlling for differences attributable to design weaknesses.

TABLE 4
Effect size (g+) for each type of research design.

		Effect size		95% confidence interval		Test of null
Research design	k	g+	SE	*Lower*	*Upper*	z
1. Preexperimental	60	.31	.02	.26	.36	12.73*
2. Quasiexperimental	74	.36	.02	.32	.40	18.01*
3. True experimental	27	.34	.04	.26	.42	8.49*

*$p < .01$.

TABLE 5
Heterogeneity analysis (Q and I^2) for type of research design

	Heterogeneity			
Research design	Q *value*	df (Q)	p	I^2
1. Preexperimental	406.77	53	.00	94.464
2. Quasiexperimental	1099.67	86	.00	86.796
3. True experimental	259.14	26	.00	85.659
Total within (Q_W)	1765.58	158	.00	
Total between (Q_B)	2.28	2	.32	
Total (Q_T)	1767.86	160	.00	90.95

Other methodological features. Two additional methodological study features were coded and analyzed: (a) type of CT measure and (b) method of extraction. Tables 6 and 7 show the breakdown of type of measure. Since Q_B was significant (indicating significant heterogeneity among levels), Bonferroni post hoc tests were conducted among pairs according to Hedges and Olkin (1985, p. 162). Each comparison is tested with $z(\hat{\gamma}) = \hat{\gamma}/\hat{\sigma}^2$ compared with the $100(1 - \alpha/2l)$ percentage point on the unit normal distribution, where l is the number of actual comparisons. This holds the family of comparisons at $\alpha = .05$. Levels containing fewer than 10 effect sizes were considered too unstable for analysis, so teacher-made tests ($k = 7$) were not included. Results revealed that standardized measures, teacher- and researcher-made, and modified secondary-source measures were not different from one another but that they were all different from researcher-made measures.

Likewise, methods of effect size extraction were compared (see Tables 8 and 9). Extraction from descriptive statistics (i.e., means and standard deviations) and estimated with no assumptions (e.g., exact t values, exact probabilities) are more accurate than estimation with assumptions (e.g., $p < .05$ plus effect direction), leading most often to an underestimation of the true effect size. Although significant, this between-group effect (Q_B) was weaker than that for measures. Bonferroni comparisons revealed no differences among the levels.

TABLE 6
Effect size (g+) for type of measure

Type of measure	k	Effect size		95% confidence interval		Test of null
		g+	SE	Lower	Upper	z
1. Standardized	91	0.24	.02	0.20	0.28	12.29*
2. Teacher made	7	1.43	.12	1.19	1.66	11.78*
3. Researcher made	31	0.65	.03	0.59	0.71	21.23*
4. Teacher/researcher	17	0.16	.04	0.08	0.23	4.27*
5. Secondary source	15	0.29	.05	0.18	0.39	5.44*

*$p < .01$.

TABLE 7
Heterogeneity analysis (Q and I^2) for type of measure

Type of measure	Heterogeneity			
	Q value	df (Q)	p	I^2
1. Standardized	566.10	90	.00	84.10
2. Teacher made	31.92	6	.00	81.20
3. Researcher made	654.55	30	.00	95.42
4. Teacher/researcher	199.02	16	.00	91.96
5. Secondary source	83.99	14	.00	83.33
Total within (Q_W)	1535.59	156	.00	
Total between (Q_B)	232.27	4	.00	
Total (Q_T)	1767.86	160	.00	90.95

TABLE 8
Effect size (g+) for method of effect size extraction

Method of effect size extraction	k	Effect size		95% confidence interval		Test of null
		g+	SE	Lower	Upper	z
1. Descriptive statistics	103	.35	.02	.32	.39	19.22*
2. Estimated (no assumptions)	30	.49	.04	.42	.56	13.62*
3. Estimated (with assumptions)	28	.20	.03	.14	.26	6.47*

*$p < .01$.

Substantive Study Features

Four substantive study features were coded: (a) age of participants, (b) type of intervention, (c) pedagogical grounding of the intervention, and (d) presence or absence of collaboration.

TABLE 9
Heterogeneity analysis (Q and I^2) for effect size extraction method

	Heterogeneity			
Type of measure	Q *value*	df (Q)	p	I^2
1. Descriptive statistics	1297.17	102	.00	92.14
2. Estimated (no assumptions)	193.92	29	.00	85.05
3. Estimated (with assumptions)	236.13	27	.00	88.57
Total within (Q_W)	1727.22	158	.00	
Total between (Q_B)	40.64	2	.00	
Total (Q_T)	1767.86	160	.00	90.95

TABLE 10
Effect size (g+) for categories of age of participants

		Effect size		95% confidence interval		Test of null
Age category	k	g+	SE	*Lower*	*Upper*	z
1. Elementary education (6–10 years)	24	.52	.05	.42	.62	10.38*
2. Secondary education (11–15 years)	34	.69	.03	.63	.75	22.57*
3. High school (16–18 years)	8	.10	.06	–.03	.22	1.56
4. Undergraduate postsecondary	80	.25	.02	.21	.29	12.30*
5. Graduate postsecondary	4	.62	.11	.41	.82	5.78*
6. Adults	9	.32	.07	.19	.44	4.83*

*$p < .01$.

Age. The results for age of participants are shown in Tables 10 and 11. The intergroup difference (Q_B) was significant, so post hoc comparisons were conducted. Again, categories with fewer than 10 effect sizes were not interpreted. The post hoc results revealed that elementary-age participants (ages 6–10) were not significantly different than secondary-age participants (ages 11–15) but that they were both significantly higher than undergraduate postsecondary students.

Type of intervention. The Q_B for type of intervention was significant, and post hoc analyses were performed. Mixed instructional approaches that combine both content and CT instruction significantly outperformed all other types of instruction. Immersion methods significantly underperformed all other approaches (see Tables 12 and 13).

Pedagogical grounding. This factor produced the strongest between-group effect ($Q_B = 446.16$), suggesting that the differences among the categories are the greatest. These results are shown in Tables 14 and 15. Examination of the means

TABLE 11
Heterogeneity analysis (Q and I^2) for categories of age of participants

	Heterogeneity			
Age category	Q *value*	df (Q)	p	I^2
1. Elementary education (6–10 years)	101.50	23	.00	77.34
2. Secondary education (11–15 years)	765.50	33	.00	95.69
3. High school (16–18 years)	47.09	7	.00	85.14
4. Undergraduate postsecondary	489.99	79	.00	83.88
5. Graduate postsecondary	24.97	3	.00	87.99
6. Adults	36.02	8	.00	77.79
Missing data	23.71	1	.00	95.78
Total within (Q_W)	1488.79	154	.00	
Total between (Q_B)	279.07	6	.00	
Total (Q_T)	1767.86	160	.00	90.95

TABLE 12
Effect size (g+) for type of intervention

		Effect size		*95% confidence interval*		*Test of null*
Type of intervention	k	g+	SE	*Lower*	*Upper*	z
1. General critical thinking skills	39	.38	.03	.31	0.45	11.29*
2. Infusion	52	.54	.03	.49	0.59	21.14*
3. Immersion	48	.09	.02	.05	0.13	4.10*
4. Mixed	22	.94	.06	.82	1.05	16.16*

*$p < .01$.

TABLE 13
Heterogeneity analysis (Q and I^2) for type of intervention

	Heterogeneity			
Type of intervention	Q *value*	df (Q)	p	I^2
1. General critical thinking skills	172.09	38	.00	77.92
2. Infusion	939.46	51	.00	94.57
3. Immersion	164.68	47	.00	71.46
4. Mixed	195.78	21	.00	89.27
Total within (Q_W)	1472.02	157	.00	
Total between (Q_B)	295.84	3	.00	
Total (Q_T)	1767.86	160	.00	90.95

TABLE 14
Effect size (g+) for pedagogical grounding of intervention

Pedagogical grounding of intervention	k	Effect size		95% confidence interval		Test of null
		g+	SE	Lower	Upper	z
1. Instructor training	16	1.00	.04	.92	1.07	25.09*
2. Extensive observations	46	0.58	.04	.51	0.65	16.62*
3. Detailed curriculum description	25	0.31	.04	.22	0.40	6.91*
4. Critical thinking among course objectives	74	0.13	.02	.09	0.17	6.97*

*$p < .01$.

and post hoc analysis revealed that instructor training significantly outperformed the other three categories. In addition, the following differences were found (Category 2 > Categories 3 and 4; Category 3 > Category 4).

A follow-up regression analysis was performed to determine the joint contribution of intervention type and pedagogical grounding to overall variation. These two study features were dummy coded and run in hierarchical weighted multiple regression. The analysis revealed that in combination, these two instructional variables accounted for 32% of the total variation in effect size. However, Q_W remained heterogeneous.

Collaborative learning conditions. A final analysis of substantive study features was conducted on the basis of the presence or absence of collaborative learning conditions associated with each study and therefore each effect size. Tables 16 and 17 present the outcomes of this analysis. Although there is a 0.10 difference between the presence and absence of collaboration, and the Q_B of 9.25 is significant, this was considered a relatively minor difference compared with other instructional variables.

It should be reiterated that in all of these analyses of substantive study features, there was not one homogeneous mean effect size. This means that although the mean differences reported in this analysis are suggestive, they are by no means definitive.

Discussion

The data (161 effect sizes from 117 studies, including 27 true experiments) suggest a generally positive effect of instruction on students' CT skills. However, the findings are not uniformly positive, and we found some evidence of negative effects. Nevertheless, given the complexity of the CT construct, we were surprised to find such positive effect sizes and are more curious now about understanding the data than when we began our investigation.

We explored several substantive features in an attempt to explain the variable findings. We were especially interested in instructional variables and found that

TABLE 15
Heterogeneity analysis (Q and I^2) for pedagogical grounding of intervention

	Heterogeneity			
Pedagogical grounding of intervention	Q *value*	df (Q)	p	I^2
1. Instructor received training	532.43	15	.00	97.18
2. Extensive observations	295.53	45	.00	84.77
3. Detailed curriculum description	111.71	24	.00	78.52
4. Critical thinking among course objectives	382.02	73	.00	80.89
Total within (Q_W)	1321.69	157	.00	
Total between (Q_B)	446.16	3	.00	
Total (Q_T)	1767.86	160	.00	90.95

TABLE 16
Effect size (g+) for the presence and absence of student collaboration in critical thinking interventions

		Effect size		*95% confidence interval*		*Test of null*
Collaboration	k	g+	SE	*Lower*	*Upper*	z
No	102	.31	.02	.28	.35	18.39
Yes	59	.41	.03	.36	.46	15.02

TABLE 17
Heterogeneity analysis (Q and I^2) for the presence and absence of student collaboration in critical thinking interventions

	Heterogeneity			
Collaboration	Q *value*	df (Q)	p	I^2
No	1031.95	101	.00	90.31
Yes	726.65	58	.00	92.02
Total within (Q_W)	1758.60	158	.00	
Total between (Q_B)	9.25	2	.01	
Total (Q_T)	1767.86	160	.00	90.95

both the type of CT intervention and the pedagogical grounding of the CT intervention contributed significantly and substantially to explaining variability in CT outcomes. Together, these two instructional variables explained 32% of the variance in effect sizes, meaning that improved CT skills and dispositions are associated with how CT instruction is provided.

Course content and the curriculum matter. Although all the subgroup means for the type of CT intervention were significantly greater than zero, they were not uniformly so. The mixed method, where CT is taught as an independent track within a specific content course (e.g., Boodt, 1984; Browne, Haas, Vogt, & West, 1977; Brownell, 1953; Crow & Haws, 1985; Hartman-Haas, 1984; Harty, Woods, Johnson, & Pifer, 1986; Klassen, 1983; Loesch-Griffin, 1986; Martin, Craft, & Sheng, 2001; Marzano, 1989; McCarthy-Tucker, 1998; Robinson 1987; Stoiber, 1991), had the largest effect, whereas the immersion method, where CT is regarded as a by-product of instruction, had the smallest effect. Moderate effects were found for both the general approach, where CT skills are the explicit course objective, and the infusion approach, where CT skill are embedded into the course content and explicitly stated as a course objective. The smallest effects, for the immersion method, occurred where CT skills were not an explicit course objective. Such indirect instruction is the least effective approach. Whether it is taught separately of content or embedded within content seems like a less important distinction empirically. This is an important finding for the design of courses. Making CT requirements a clear and important part of course design is associated with larger instructional effects. Developing CT skills separately and then applying them to course content explicitly works best; immersing students in thought-provoking subject matter instruction without explicit use of CT principles was least effective.

Pedagogy matters. When instructors received special advanced training in preparation for teaching CT skills (e.g., Arlo, 1969; Daley, Shaw, Balistrieri, Glasenapp, & Piacentine, 1999; Feuerstein, 1999; Hook, Jacobs, & Crisp, 1969; MacPhail-Wilcox, Dreyden, & Eason, 1990; Martin et al., 2001; McCarthy-Tucker, 1998; McConney, McConney, & Horton, 1994; Riesenmy et al., 1991; VanTassel-Baska et al., 2002; Zohar et al., 1994; Zohar & Tamir; 1993), or when extensive observations on course administration and instructors' CT teaching practices were reported, the impacts of the interventions were greatest. By contrast, impacts of CT were smallest when the intention to improve students' CT was only listed among the course objectives and there were no efforts at professional development or elaboration of course design and implementation. These results suggest that better outcomes can be achieved through active, purposeful training and teacher support both at the preservice and in-service levels. The results also implicate the design of courses in which CT skills are addressed, either generically or in concert with course content. To maximize impact requires both the willingness to incorporate CT instruction and explicit strategies and skills to do it effectively.

We also found that collaboration among students while developing CT skills appears to provide some advantage but that this effect is minor compared with other substantive instructional study features. Recently, theoretical work has appeared in the literature of distance and Web-based education relating CT and online collaboration (e.g., Fahy, 2005; Garrison, Anderson, & Archer, 2001; C. Perkins & Murphy, 2006). Because most of the empirical research that has appeared to date has been qualitative in nature, it is difficult to judge now what this adaptation of classroom practice will yield.

These findings make it clear that improvements in students' CT skills and dispositions cannot be a matter of implicit expectation. As important as the development of CT skills is, educators must take steps to make CT objectives explicit in

courses and to integrate them into both preservice and in-service training and faculty development. If the outcome is worth it, the effort is worth it. One unanswered question, of course, is whether results that are achieved in particular classrooms have a lasting or transitory effect and, in this sense, whether the outcome is worth the effort. It seems unlikely that a "hit-or-miss" approach to developing thinking skills, or any skill for that matter, will yield satisfactory results that extend very far into the future.

The next steps in our review of the evidence include even more complete and more detailed examinations of the evidence. There are more methodological and substantive features we wish to explore, including treatment duration, subject matter, student characteristics, instructor characteristics, and so on. We hope to be able to account for further variation in effect size based on these study features so that we can sort out the various conditions of intervention type, learner type, subject matter, and so on that better explain why high levels of success are achieved in certain classroom applications and not in others.

Arguably the most illusive and, at the same time, most important aspect of the work to be completed concerns the quality of CT interventions. Specifically, we wonder whether and to what extent implementation fidelity is related to effect sizes both overall and for specific CT interventions.

Our future work will extend from this Stage 1 review to a finer-grained analysis of methodological and substantive issues, and we will attempt to more completely answer the question: What are the specific elements of instructional practice that develop better CT and learning outcomes for students? Much like a single primary investigation cannot address every critical hypothesis, a single review may not address every question about the collection of evidence. Studies follow studies and so too should reviews follow reviews.

These initial results appear to contradict some of the previous reviews (e.g., McMillan, 1987) that have found that instructional interventions have little effect on the development of CT skills and dispositions. However, we cannot make final claims about the state of the population until the wide heterogeneity in study findings has been further reduced. We decided to report these intermediate results both to distribute the tentative findings to educators and scholars and also to get feedback while the work is in progress. We hope others find these preliminary data as interesting and compelling as we do. At the same time, we welcome comments and criticisms that may lead us to improve the quality of our undertaking.

If the learning of CT skills and the development of the disposition to think critically can be influenced through instruction, as it registers on posttest measures of CT, that is one thing of importance that can result from the continuation of this work. We hope to draw new attention to CT instruction in schools and possibly suggest some new directions for pedagogy, and given our wider interests, both with and without the use of technology (Bernard et al., 2004).

Otherwise, efficiency concerns need to be explored. In other words, we need to consider whether to change curricula to address development of CT skills, without neglecting subject matter content. Undoubtedly, we will encounter in this question the debate about fundamental educational competencies and where in this schema to place CT.

Finally, a more important question that may never be fully answered is the following: How much does learning to think critically in classroom settings lead to

better citizens who are more discerning, people who are more analytical in their professions, or parents who can think carefully through the variety of choices that face them in raising a family in a complex and challenging world? That is the true test of CT instruction, specifically, and indeed, the true test of education in general.

Note

An earlier version of this article was presented at the 2006 annual meeting of the American Educational Research Association in San Francisco. This study was supported by grants to Abrami and Bernard from Fonds québécois de la recherche sur la société et la culture and the Social Sciences and Humanities Research Council of Canada. The authors express appreciation to Anna Peretiatkovicz for her assistance and contributions.

References

References marked with an asterisk indicate studies included in the meta-analysis.

Abrami, P. C., & Bernard, R. M. (2008). *Statistical control vs. classification of study quality in meta-analysis.* Manuscript in preperation.

Adams, B. L. (1999). Nursing education for critical thinking: An integrative review. *Journal of Nursing Education, 38*(3), 111–119.

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

Allen, M., Berkowitz, S., Hunt, S., & Louden, A. (1997, November). *Measuring the impact of forensics and communication education on critical thinking: A meta-analytic summary.* Paper presented at the annual meeting of the National Communication Association, Chicago. (ERIC Document Reproduction Service No. ED413625)

Allen, M., Berkowitz, S., Hunt, S., & Louden, A. (1999). A meta-analysis of the impact of forensics and communication education on critical thinking. *Communication Education, 48*(1), 18–30.

*Anderson, T., Howe, C., Soden, R., Halliday, J., & Low, J. (2001). Peer interaction and the learning of critical thinking skills in further education students. *Instructional Science, 29*(1), 1–32.

*Angel, B. F., Duffey, M., & Belyea, M. (2000). An evidence-based project for evaluating strategies to improve knowledge acquisition and critical-thinking performance in nursing students. *Journal of Nursing Education, 39*(5), 219–228.

*Annis, L. F., & Annis, D. B. (1979). The impact of philosophy on students' critical thinking ability. *Contemporary Educational Psychology, 4*(3), 219–226.

*Arburn, T. M., & Bethel, L. J (1999, March). *Assisting at-risk community college students: Acquisition of critical thinking learning strategies.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, Boston. (ERIC Document Reproduction Service No. ED448016)

*Arlo, R. (1969). The relative effectiveness of inductive and expository teaching of principles of general semantics upon the critical reading ability of 9th-grade students (Doctoral dissertation, New York University, 1969). *Dissertation Abstracts International, 30,* 2411A. Retrieved May 25, 2006, from ProQuest Digital Dissertation + Theses Fulltext database. (ERIC Document Reproduction Service No. ED046923)

Assessment and Learning Research Synthesis Group. (2003). *A systematic review of the impact on students and teachers of the use of ICT for assessment of creative and critical thinking skills.* Social Science Research Unit, Institute of Education, University of London.

*Athanassiou, N., McNett, J. M., & Harvey, C. (2003). Critical thinking in the management classroom: Bloom's taxonomy as a learning tool. *Journal of Management Education, 27*(5), 533–555.

*Ayoub, J. L. (1998). The effect of classroom use of an electronic group support system on student critical thinking, performance and satisfaction (Doctoral dissertation, University of Arizona, 1998). *Dissertation Abstracts International, 58*, 3030A.

Bailin, S., & Siegel, H. (2003). Critical thinking. In N. Blake, P. Smeyers, R. Smith, & P. Standish (Eds.), *The Blackwell guide to the philosophy of education* (pp. 181–193). Oxford, UK: Blackwell.

Bangert-Drowns, R. L., & Bankert, E. (1990, April). *Meta-analysis of effects of explicit instruction for critical thinking.* Paper presented at the annual meeting of the American Educational Research Association, Boston. (ERIC Document Reproduction Service No. ED 328614)

*Bartlett, D. J., & Cox, P. D. (2002). Measuring change in students' critical thinking ability: Implications for health care education. *Journal of Allied Health, 31*(2), 64–69.

*Bean, A. T. (1969). *Some comparisons between conventional college teaching methods and a composite of procedures involving large lecture groups, seminars, and reduced class time.* Unpublished doctoral dissertation, North Texas State University. (ERIC Document Reproduction Service No. ED0320320)

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

Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., et al. (2004). How does distance education compare to classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research, 74*(3), 379–439.

Bernard, R. M., Zhang, D., Abrami, P. C., Sicoly, F., Borokhovski, E., & Surkes, M. A. (2008). Exploring the structure of the Watson-Glaser Critical Thinking appraisal: One scale or many subscales? *Thinking Skills and Creativity, 3*, 15–22.

*Bledsoe, J. C. (1955). A comparative study of values and critical thinking skills of a group of educational workers. *Journal of Educational Psychology*, 408–417.

Bloom, M., & Watt, D. (2003). Solving Canada's innovation conundrum: How public education can help. Ottawa, ON: Conference Board of Canada.

*Blum, M., & Spangehl, S. (1979). *The role of programmed instruction for sequential skill development in higher education.* Unpublished manuscript, University of Louisville, KY. (ERIC Document Reproduction Service No. ED197670)

*Bonk, C. J., Angeli, C., Malikowski, S. R., & Supplee, L. (2001). Holy cow: Scaffolding case based conferencing on the web with preservice teachers. *Ed at a Distance, 15*(8). Retrieved May 25, 2006, from http://www.usdla.org/html/journal/AUG01_Issue/article01.html

*Boodt, G. M. (1984). Critical listeners become critical readers in remedial reading class. *Reading Teacher, 37*(4), 390–394.

Borenstein, M., Hedges, L., Higgins, J., & Rothstein, H. (2005). *Comprehensive meta-analysis* (Version 2). Englewood, NJ: Biostat.

**Broward County home economics curriculum development project, July 1, 1967 to July 1, 1970. Final report.* (1970). Tallahassee: Florida State Department of

Education; Division of Vocational, Technical, and Adult Education. (ERIC Document Reproduction Service No. ED055158)

*Brown, L. E. (1970). An experimental study of the effectiveness of organizing learning experiences to achieve the objective of critical thinking in certain fifth grade social studies classes (Doctoral dissertation, Minnesota University, 1970). *Dissertation Abstracts International, 30*, 2898A. Retrieved May 25, 2006, from the ProQuest Digital Dissertation + Theses Fulltext database. (ERIC Document Reproduction Service No. ED046830)

*Browne, M. N., Haas, P. F., Vogt, K. E., & West, J. S. (1977). Design and implementation of an evaluation procedure for an innovative undergraduate program. *College Student Journal, 11*(4, Part 2), 10.

*Brownell, J. A. (1953). The influence of training in reading in the social studies on the ability to think critically. *California Journal of Educational Research*, 28–31.

*Carey, S., Evans, R., Honda, M., Jay, E., & Unger, C. (1988). *What junior high school students do, can, and should know about the nature of science and scientific inquiry* (Technical Report 88-2). Cambridge, MA: Educational Technology Center. (ERIC Document Reproduction Service No. ED295833)

*Champion, B. W. (1975). *An inductive approach to the discrimination of factual and inferential statements as an aspect of critical reading.* Unpublished doctoral dissertation, University of Temple, PA. (ERIC Document Reproduction Service No. ED116129)

*Colbert, K. R. (1987). The effects of CEDA and NDT debate training on critical thinking ability. *Journal of the American Forensic Association, 23*(4), 194–201.

*Comerford, S. S., Busk, P. L., & Roberts, W. L. (2000, April). *Epistemological beliefs of community-college students.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED442483)

*Commeyras, M., Pearson, P. D., Ennis, R. H., Garcia, G., & Anderson, R. C. (1992). *Dialogical-thinking reading lessons: Promoting critical thinking among "learning-disabled"* students (Technical Report No. 553). Champaign: University of Illinois at Urbana-Champaign, College of Education. (ERIC Document Reproduction Service No. ED343115)

*Crow, L. W., & Haws, S. G. (1985, April). *The effects of teaching logical reasoning upon students' critical thinking and science achievement.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, French Lick Springs, IN. (ERIC Document Reproduction Service No. ED255371)

*Curtis, C. K. (1980). Developing critical thinking skills in nonacademic social studies classes. *Alberta Journal of Educational Research, 26*(2), 75–84.

*Daley, B. J., Shaw, C. R., Balistrieri, T., Glasenapp, K., & Piacentine, L. (1999). Concept maps: A strategy to teach and evaluate critical thinking. *Journal of Nursing Education, 38*(1), 42–47.

*Dolan, S. L. (1976). The effect of value-conflict discussion on the critical thinking and reading of adults (Doctoral dissertation, University of Pittsburg, 1976). *Dissertation Abstracts International, 37*, 2737A. Retrieved May 25, 2006, from ProQuest Digital Dissertation + Theses Fulltext database. (ERIC Document Reproduction Service No. ED132543)

*Elliott, B., Oty, K., McArthur, J., & Clark, B. (2001). The effect of an interdisciplinary algebra/science course on students' problem solving skills, critical thinking skills and attitudes towards mathematics. *International Journal of Mathematical Education in Science and Technology, 32*(6), 811–816.

*Ellis, T. J. (2001). Multimedia enhanced educational products as a tool to promote critical thinking in adult students. *Journal of Educational Multimedia and Hypermedia, 10*(2), 107–123.

*English, J. D. (1999). Improvement in drafting instruction: Using graphic organizers and paired problem solving to enhance critical thinking and content knowledge (Doctoral dissertation, University of Kentucky, 1999). *Dissertation Abstracts International, 60*, 718A. Retrieved October 27, 2004, from ProQuest Digital Dissertation + Theses Fulltext database.

Ennis, R. H. (1962). A concept of critical thinking: A proposed basis for research in the teaching and evaluation of critical thinking ability. *Harvard Educational Review, 32*(1), 81–111.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. Baron & R. Sternberg (Eds.), *Teaching thinking skills: Theory and practice* (pp. 9–26). New York: W. H. Freeman.

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

*Ennis, R. H., Finkelstein, M. R., Smith, E. L., & Wilson, N. H. (1969). *Conditional logic and children* (Cornell Critical Thinking Readiness Project, Phase IIc). Ithaca, NY: New York State College of Agriculture at Cornell University, Department of Education. (ERIC Document Reproduction Service No. ED040437)

Ennis, R. H., & Millman, J. (1985a). *Cornell Critical Thinking Test, Level X.* Pacific Grove, CA: Midwest.

Ennis, R. H., & Millman, J. (1985b). *Cornell Critical Thinking Test, Level Z.* Pacific Grove, CA: Midwest.

*Facione, P. A. (1990). *The California Critical Thinking Skills Test—College level: Experimental validation and content validity* (Technical Report No. 1). Millbrae: California Academic Press. (ERIC Document Reproduction Service No. ED327549)

*Facione, P. A. (1990). *The California Critical Thinking Skills Test—College Level: Factors predictive of CT skills* (Technical Report No. 2). Millbrae: California Academic Press. (ERIC Document Reproduction Service No. ED327550)

Facione, P. A. (1990a). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Research findings and recommendations.* Newark, DE: American Philosophical Association. (ERIC Document Reproduction Service No. ED315423)

Facione, P. A. (1990b). *The California Critical Thinking Skills Test—College level: Interpreting the CCTST, group norms and sub-scores* (Technical Report No. 4). Millbrae: California Academic Press.

Facione, P. A., & Facione, N. (1992). *The California Critical Thinking Dispositions Inventory (CCTDI) and the CCTDI test manual.* Millbrae: California Academic Press.

Fahy, P. J. (2005). Two methods for assessing critical thinking in computer-mediated communications (CMC) transcripts. *International Journal of Instructional Technology and Distance Education, 2*(3). Retrieved July 30, 2006, from http://www.itdl.org/Journal/Mar_05/article02.htm

*Farynaiarz, J. V., & Lockwood, L. G. (1992). Effectiveness of microcomputer simulations in stimulating environmental problem solving by community college students. *Journal of Research in Science Teaching, 29*(5), 453–470.

*Feuerstein, M. (1999). Media literacy in support of critical thinking. *Journal of Educational Media, 24*(1), 43–54.

Follert, V. F., & Colbert, K. R. (1983, November). *An analysis of the research concerning debate training and critical thinking improvements.* Paper presented at the 69th annual meeting of the Speech Communication Association, Washington, DC.

*Freeman, B. (1995). A study of the efficacy of a developmental instruction design for non-traditional adult learners (nontraditional students) (Doctoral dissertation, Columbia University Teachers College, 1995). *Dissertation Abstracts International, 55*, 2248A. Retrieved October 27, 2004, from ProQuest Digital Dissertation + Theses Fulltext database.

*Frost, S. H. (1991). Fostering the critical thinking of college women through academic advising and faculty contact. *Journal of College Student Development, 32*(4), 359–366.

*Gadzella, B. M., Ginther, D. W., & Bryant, W. (1996, August). *Teaching and learning critical thinking skills.* Paper presented at the 26th International Congress of Psychology, Montreal, QC. (ERIC Document Reproduction Service No. ED405313)

*Garcia, T., & Pintrich, P. R. (1992, August). *Critical thinking and its relationship to motivation, learning strategies, and classroom experience.* Paper presented at the annual meeting of the American Psychological Association, Washington, DC. (ERIC Document Reproduction Service No. ED351643)

Garrison, D. R., Anderson, T., & Archer, W. (2001). Critical thinking, cognitive presence and computer conferencing in distance education. *American Journal of Distance Education, 15*(1), 7–23.

*Gelven, D. R., & Stewart, B. R. (2001). Developing critical thinking skills of tech prep students using applied communications. *Journal of Technology Studies, 27*(2), 102–107.

*Germann, P. J. (1989). Directed-inquiry approach to learning science process skills: Treatment effects and aptitude-treatment interactions. *Journal of Research in Science Teaching, 26*(3), 237–250.

*Gibbs, L. E., Browne, M. N., & Keeley, S. M. (1988). *Stimulating critical thinking through faculty development: Design, evaluation, and problems.* Unpublished manuscript, Wisconsin University, Eau Claire. (ERIC Document Reproduction Service No. ED306902)

Glaser, E. M., (1942). *An experiment in the development of critical thinking.* New York: AMS. (Reprint of 1941 edition)

Glass, G. V., McGaw, B., & Smith, M. L. (1981). *Meta-analysis in social research.* Beverly Hills, CA: Sage.

Government of Canada. (2002). *Knowledge matters: Skills and learning for Canadians* (Cat No. SP-482-02-02). Hull, QC: Human Resources Development Canada.

*Green, C. S. I., & Klug, H. G. (1990). Teaching critical thinking and writing through debates: An experimental evaluation. *Teaching Sociology, 18*(4), 462–471.

*Griffin, C. M. (1979). *The effect of the use of selected newspaper articles in the teaching of certain critical reading/thinking skills.* Unpublished doctoral dissertation, University of New Jersey, Rutgers, NJ. (ERIC Document Reproduction Service No. ED169479)

Halliday, J. (2000). Critical thinking and the academic vocational divide. *Curriculum Journal, 11*(2), 159–175.

*Hancock, B. W., Coscarelli, W. C., & White, G. P. (1983). Critical thinking and content acquisition using a modified guided design process for large course sections. *Educational and Psychological Research, 3*(3), 139–149.

Hare, W. (1985). *In defense of open-mindedness.* Kingston, ON: McGill-Queen's University Press.

*Hartman-Haas, H. J. (1984, April). *An evaluation of the holistic approach to improving thinking.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED242759)

*Harty, H., Woods, C. J., Johnson, D. A., & Pifer, S. (1986). *Academic challenge program: Evaluation report.* Columbus, IN: Bartholomew Consolidated School Corp. (ERIC Document Reproduction Service No. ED255371)

*Hawkes, M. (2001). Variables of interest in exploring the reflective outcomes of network-based communication. *Journal of Research on Computing in Education, 33*(3), 299–315.

Hedges, L. V., & Olkin, I. (1985). *Statistical methods for meta-analysis.* Orlando, FL: Academic Press.

Hedges, L. V., Shymansky, J. A., & Woodworth, G. (1989). *A practical guide to modern methods of meta-analysis.* (ERIC Document Reproduction Service No. ED 309 952)

*Heerman, C. E. (1983, October). *Effects of two comprehension treatments in a college reading laboratory.* Paper presented at the annual meeting of the College Reading Association, Atlanta, GA. (ERIC Document Reproduction Service No. ED236577)

*Heraty, N., & Morley, M. J. (2000). The application of the Structure of Intellect Programme: A manufacturing facility experiment. *Journal of Managerial Psychology, 15*(7), 691–715.

Higgins, J. P. T., & Thompson, S. G. (2002). Quantifying heterogeneity in a meta-analysis. *Statistics in Medicine, 21*(11), 1539–1558.

*Hilgenberg, C., & Tolone, W. (2000). Student perceptions of satisfaction and opportunities for critical thinking in distance education by interactive video [Electronic version]. *American Journal of Distance Education, 14*(3), 59–73.

*Hill, L. (2000). What does it take to change minds? Intellectual development of preservice teachers. *Journal of Teacher Education, 51*(1), 50–62.

Hook, J. N., Jacobs, P. H., & Crisp, R. D. (1969). Illinois state-wide curriculum study center in preparation of secondary school English teachers: Final report. Urbana, IL: ISCPET and USOE.

*Housen, A. (2001, April). *Methods for assessing transfer from an art-viewing program.* Paper presented at the annual meeting of the American Educational Research Association, Seattle, WA. (ERIC Document Reproduction Service No. ED457186)

*Hudgins, B. B., & Edelman, S. (1988). Children's self-directed critical thinking. *Journal of Educational Research, 81*(5), 262–273.

Huedo-Medina, T. B., Sanchez-Meca, J., Marin-Martinez, F., & Botella, J. (2006). Assessing heterogeneity in meta-analysis: Q statistic or I[squared] index? *Psychological Methods, 11*(2), 193–206.

*Huff, M. T. (2000). A comparison study of live instruction versus interactive television for teaching MSW students critical thinking skills. *Research on Social Work Practice, 10*(4), 400–416.

Hyslop-Margison, E. J. (2003). The failure of critical thinking: Considering a virtue epistemology pedagogy. *Philosophy of Education Society Yearbook*, 319–326.

Hyslop-Margison, E. J., & Armstrong, J. (2004). Critical thinking in career education: The democratic importance of foundational rationality. *Journal of Career and Technical Education, 21*(1), 39–49.

*Inlow, F. H., & Chovan, W. (1993). Another search for the effects of teaching thinking and problem solving skills on college students' performance. *Journal of Instructional Psychology, 20*(3), 215–223.

*Kamin, C., O'Sullivan, P., & Deterding, R. (2002, April). *Does project L.I.V.E. case modality impact critical thinking in PBL groups?* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED464921)

*Keeley, S. M., Ali, R., & Gebing, T. (1998). Beyond the sponge model: Encouraging students' questioning skills in abnormal psychology. *Teaching of Psychology, 25*(4), 270–274.

*Kemp, S. G., & Sadoski, M. (1991). The effects of instruction in forming generalizations on high school students' critical thinking in world history. *Reading Research and Instruction, 31*(1), 33–42.

*Kintgen Andrews, J. (1988). *Development of critical thinking: Career ladder P.N. and A.D. nursing students, pre-health science freshmen, generic baccalaureate sophomore nursing students.* Unpublished manuscript, University of Minnesota School of Nursing, Minneapolis. (ERIC Document Reproduction Service No. ED297153)

*Klassen, P. T. (1983). Changes in personal orientation and critical thinking among adults returning to school through weekend college: An alternative evaluation. *Innovative Higher Education, 8*(1), 55–67.

*Klein, P., Olson, D. R., & Stanovich, K. (1997). Structuring reflection: Teaching argument concepts and strategies enhances critical thinking. *Canadian Journal of School Psychology, 13*(1), 38–47.

Kurfiss, J. G. (1988). *Critical thinking: theory, research, practice, and possibilities* (Ashe-Eric Higher Education Report No. 2). Washington, DC: Association for the Study of Higher Education. (ERIC Document Reproduction Service No. ED304041)

*Langer, P., & Chiszar, D. (1993). Assessment of critical thinking courses. *Perceptual and Motor Skills, 77*(3, Part 1), 970.

*Lee, M. O. C., & Thompson, A. (1997). Guided instruction in logo programming and the development of cognitive monitoring strategies among college students. *Journal of Educational Computing Research, 16*(2), 125–144.

*Lewanski, P. M. (1999). *The effect of computer-assisted library research on critical thinking skills acquisition of ninth-grade students.* Unpublished doctoral dissertation, United States International University, Nairobi, Kenya.

Lipman, M. (1991). *Thinking in education.* Cambridge, UK: Cambridge University Press.

*Liu, M., & Pedersen, S. (1998, June). *The effect of being hypermedia designers on elementary school students' motivation and learning of design knowledge.* Paper presented at ED-MEDIA/ED-TELECOM 98 World Conference on Educational Multimedia and Hypermedia and World Conference on Educational Telecommunications, Freiburg, Germany. (ERIC Document Reproduction Service No. ED428689)

*Loesch-Griffin, D. A. (1986, August). *Gender differences in the instruction and intelligent behavior of fifth-graders.* Paper presented at the annual meeting of the American Psychological Association, Washington, DC. (ERIC Document Reproduction Service No. ED284668)

*Lundsteen, S. W. R. (1963). *Teaching abilities in critical listening in the fifth and sixth grades.* Unpublished doctoral dissertation, University of California, Berkeley. (ERIC Document Reproduction Service No. ED063296)

*MacPhail-Wilcox, B., Dreyden, J., & Eason, E. (1990). An investigation of PAIDEIA program effects on students' critical thinking skills. *Educational Considerations, 17*(2), 61–67.

*Magnussen, L., Ishida, D., & Itano, J. (2000). The impact of the use of inquiry-based learning as a teaching methodology on the development of critical thinking. *Journal of Nursing Education, 39*(8), 360–364.

Marcuse, H., & Magee, B. (1984). Marcuse and the Frankfurt School: Dialogue with Herbert Marcuse. In B. Magee (Ed.), *Men of ideas: Some creators of contemporary philosophy* (pp.43-55). Oxford: Oxford University Press.

*Marsh, S. M. (1998, April). *Widening the lens of diversity: Motivating reflective journal writing.* Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA. (ERIC Document Reproduction Service No. ED418425)

*Martin, D. S., Craft, A., & Sheng, Z. N. (2001). The impact of cognitive strategy instruction on deaf learners: An international comparative study. *American Annals of the Deaf, 146*(4), 366–378.

*Marzano, R. J. (1989). *Evaluations of the "tactics for thinking" program: Summary report.* Aurora, CO: Mid-Continent Regional Educational Lab. (ERIC Document Reproduction Service No. ED377549)

*Matthews, D. B. (1989). The effect of a thinking-skills program on the cognitive abilities of middle school students. *Clearing House, 62*(5), 202–204.

*McBride, R., & Bonnette, R. (1995). Teacher and at-risk students' cognitions during open-ended activities: Structuring the learning environment for critical thinking. *Teaching and Teacher Education, 11*(4), 373–388.

*McCarthy-Tucker, T. S. N. (1998). Teaching logic to adolescents to improve thinking skills. *Korean Journal of Thinking and Problem Solving, 8*(1), 45–66.

*McConney, A. W., McConney, A., & Horton, P. (1994, March). *The effects of an interdisciplinary curriculum unit on the environmental decision-making of secondary school students.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, Anaheim, CA. (ERIC Document Reproduction Service No. ED368566)

McMillan, J. H. (1987). Enhancing college students' critical thinking: A review of studies. *Research in Higher Education, 26*(1), 3–29.

McPeck, J. (1981). *Critical thinking and education.* Oxford, UK: Martin Robertson.

*Meiss, G. T., & Bates, G. W. (1984, May). *Cognitive and attitudinal effects of reasoning message strategies.* Paper presented at the annual meeting of the International Communication Association, San Francisco. (ERIC Document Reproduction Service No. ED246519)

*Moffett, D. W. (1998, October). *Evaluating supplemental program implementation through measurements of students' critical thinking skills and writing ability.* Paper presented at the annual meeting of the Mid-Western Educational Research Association, Chicago. (ERIC Document Reproduction Service No. ED428093)

*Neimark, E. D. (1984, August). *A cognitive style change approach to the modification of thinking in college students.* Paper presented at the Conference on Thinking, Cambridge, MA. (ERIC Document Reproduction Service No. ED261301)

*Neman, R. L. (1972). A focus on problems of national interest in the college general chemistry laboratory: The effects of the problem-oriented method compared with those of the traditional approach (Doctoral dissertation, North Texas State University, 1970). *Dissertation Abstracts International, 33*, 6068A. Retrieved May 25, 2006, from ProQuest Digital Dissertation + Theses Fulltext database. (ERIC Document Reproduction Service No. ED111603)

*Norris, C., & Jackson, L. (1992). The effect of computer science instruction on critical thinking skills and mental alertness. *Journal of Research on Computing in Education, 24*(3), 329–337.

Norris, S. P. (1985). Synthesis of research on critical thinking. *Educational Leadership, 42*(8), 40–45.

*Norton, S., Halpin, G., & Halpin, G. (1985, November). The effects of an independent laboratory investigation on the critical thinking ability and scientific attitude of students in a general microbiology class. Paper presented at the annual meeting of the Mid-South Research Association, Biloxi, MS. (ERIC Document Reproduction Service No. ED264291)

*Parkinson, M. G., & Ekachai, D. (2002). The Socratic method in the introductory PR course: An alternative pedagogy. *Public Relations Review*, *28*(2), 167–174.

Paul, R. W. (1990). *Critical thinking: What every person needs to survive in a rapidly changing world.* Santa Rosa, CA: Foundation for Critical Thinking.

Paul, R. W., & Binker, A. J. A. (1990). Strategies: Thirty-five dimensions of critical thinking. In A. J. A. Binker (Ed.), *Critical thinking: What every person needs to survive in a rapidly changing world* (pp. 305–349). Rohnert Park, CA: Sonoma State University, Center for Critical Thinking and Moral Critique.

Paul, R. W., Elder, L., & Bartell, T. (1997). Study of 38 public universities and 28 private universities to determine faculty emphasis on critical thinking on instruction. Retrieved April 6, 2003, from http://www.criticalthinking.org/schoolstudy

Perkins, C., & Murphy, E. (2006). Identifying and measuring individual engagement in critical thinking in online discussions: An exploratory case study. *Educational Technology and Society*, *9*(1), 298–307.

Perkins, D. N., & Salomon, G. (1988). Teaching for transfer. *Educational Leadership*, *46*(1), 22–32.

*Perlman, J. S. (1953). An historical vs. contemporary problem solving use of the college physical science laboratory period for general education. *Journal of Experimental Education*, *21*, 251–257.

*Pierce, W., Lemke, E., & Smith, R. (1988, April). *Effects of a secondary curriculum for moral development and critical thinking.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED296940)

Pithers, R. T., & Soden, R. (2000). Critical thinking in education: A review. *Educational Research*, *42*(3), 237–249.

*Pitts Scarangello, A. C. (1971). *A study in developing critical thinking through a text in elementary school social studies.* Unpublished doctoral dissertation, University of Pennsylvania, Philadelphia. (ERIC Document Reproduction Service No. ED079718)

*Price, W. F., Wilmes, D. R., & Turmel, M. S. (1988, April). *Assessment of general education knowledge and competencies in Psy 101, Introductory Psychology.* Paper presented at the annual meeting of the American Educational Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED342443)

Quin, R., & McMahon, B. (1993). Monitoring standards in media studies: Problems and strategies. *Australian Journal of Education*, *37*(2), 182–197.

*Reed, J. H., & Kromrey, J. D. (2001). Teaching critical thinking in a community college history course: Empirical evidence from infusing Paul's model. *College Student Journal*, *35*(2), 201–215.

*Riesenmy, M. R., Mitchell, S., Hudgins, B. B., & Ebel, D. (1991). Retention and transfer of children's self-directed critical thinking skills. *Journal of Educational Research*, *85*(1), 14–25.

*Robinson, I. S. (1987). *A program to incorporate high-order thinking skills into teaching and learning for Grades K–3.* Unpublished doctoral dissertation, Nova Southeastern University, Fort Lauderdale, FL. (ERIC Document Reproduction Service No. ED284689)

Royalty, J. (1991). The generalizability of critical thinking: Paranormal beliefs versus statistical reasoning. *Journal of Genetic Psychology*, *156*(4), 477–488.

Sá, W. C., Stanovich, K. E., & West, R. F. (1999). The domain specificity and generality of belief bias: Searching for generalizable critical thinking skill. *Journal of Educational Psychology*, *91*(3), 497–510.

*Sanders, J. A., & Wiseman, R. L. (1994). Does teaching argumentation facilitate critical thinking? *Communication Reports*, *7*(1), 27–35.

*Saucier, B. L., Stevens, K. R., & Williams, G. B. (2000). Critical thinking outcomes of computer-assisted instruction versus written nursing process. *Nursing and Health Care Perspectives*, *21*(5), 240–246.

Scriven, M., & Paul, R. (1996). *Defining critical thinking: A draft statement for the National Council for Excellence in critical thinking*. Retrieved April 23, 2003, from http://www.criticalthinking.org/University/univlibrary/library.nclk

Sears, A., & Parsons, J. (1999). Principles of an ethic of critical thinking. In R. Case & P. Clark (Eds.), *The Canadian anthology of social studies* (pp. 171–177). Burnaby, BC: Pacific Educational.

Sheffler, I. (1973). *Reason and teaching*. Indianapolis, IN: Hackett.

Siegel, H. (1988). *Educating reason: Rationality, critical thinking, and education*. New York: Routledge.

*Sister Constantine, M. S. S. J. (1968). *An experimental study of the development of critical thinking skills of high school English teachers enrolled in a methods course* (Interim Report 32). (ERIC Document Reproduction Service No. ED019283)

*Smith, C. A., Strand, S. E., & Bunting, C. J. (2002). The influence of challenge course participation on moral and ethical reasoning. *Journal of Experiential Education*, *25*(2), 278–280.

Smith, G. (2002). Are there domain-specific thinking skills? *Journal of Philosophy of Education*, *36*(2), 207–227.

*Solon, T. (2001). Improving critical thinking in an introductory psychology course. *Michigan Community College Journal: Research and Practice*, *7*(2), 73–80.

*Soukup, F. (1999). *Assessment of critical thinking skills in associate degree nursing students*. Unpublished master's thesis, University of Wisconsin–Stout. (ERIC Document Reproduction Service No. ED430081)

*Starr, R. J. (1972). Structured oral inquiry improves thinking. *American Biology Teacher*, *34*(7), 408–409.

*Stekel, F. D. (1969). *Development of a more flexible physical science laboratory program for non-science majors with superior high school science backgrounds: Final report*. Stevens Point: Wisconsin State Universities Consortium of Research Development. (ERIC Document Reproduction Service No. ED053987)

*Stoiber, K. C. (1991). The effect of technical and reflective preservice instruction on pedagogical reasoning and problem solving. *Journal of Teacher Education*, *42*(2), 131–139.

*Thomas, J. W. (1975, April). *The effects of instruction on the critical thinking abilities of middle-school-age children*. Paper presented at the annual meeting of the American Educational Research Association, Washington, DC. (ERIC Document Reproduction Service No. ED105420)

*Thomas, M. J. W. (2002). Learning within incoherent structures: The space of online discussion forums. *Journal of Computer Assisted Learning*, *18*(3), 351–366.

*Thompson, C., & Rebeschi, L. M. (1999). Critical thinking skills of baccalaureate nursing students at program entry and exit. *Nursing and Health Care Perspectives*, *20*(5), 248–252.

Tsui, L. (1998, November). *A review of research on critical thinking.* Paper presented at the annual meeting of the Association for the Study of Higher Education. Miami, FL. (ERIC Document Reproduction Service No. ED427572)

Tsui, L. (2002). Fostering critical thinking through effective pedagogy: Evidence from four institutional case studies. *Journal of Higher Education, 73*(6), 740–763.

*Unrau, N. J. (1991, April). *The effects of explicit instruction on critical reading and argumentative writing: The task of reading and writing.* Paper presented at the annual meeting of the American Educational Research Association, Chicago, IL. (ERIC Document Reproduction Service No. ED336737)

*VanTassel-Baska, J., Zuo, L., Avery, L. D., & Little, C. A. (2002). A curriculum study of gifted-student learning in the language arts. *Gifted Child Quarterly, 46*(1), 30–44.

*Viscovich, S. A. (2002). The effects of three organizational structures on the writing and critical thinking of fifth graders. In P. E. Linder, M. B. Sampson, et al. (Eds.), *Celebrating the faces of literacy* (pp. 44–67). Readyville, TN: College Reading Association.

*Vockell, E. L., & Rivers, R. H. (1984, April). *Computerized science simulations stimulus to generalized problem solving capabilities.* Paper presented at the annual convention of the American Education Research Association, New Orleans, LA. (ERIC Document Reproduction Service No. ED253397)

*Walters, K. K., & Strode, S. L. (1991, November). *Two studies of the effects of annotation training on college students.* Paper presented at the annual meeting of the College Reading Association, Crystal City, VA. (ERIC Document Reproduction Service No. ED344202)

*Wang, W. (2001). The relative effectiveness of structured questions and summarizing on near and far transfer tasks. In *Annual proceedings of selected research and development [and] practice papers presented at the National Convention of the Association for Educational Communications and Technology* (pp. 445–451). Atlanta: North Carolina Wesleyan College. (ERIC Document Reproduction Service No. ED470123)

*Watson, D. L., Hagihara, D. K., & Tenney, A. L. (1999). Skill-building exercises and generalizing psychological concepts to daily life. *Teaching of Psychology, 26*(3), 193–195.

Watson, G., & Glaser, E. M. (1980). *Watson-Glaser Critical Thinking Appraisal.* San Antonio, TX: PsychCorp.

*Wesp, R., & Montgomery, K. (1998). Developing critical thinking through the study of paranormal phenomena. *Teaching of Psychology, 25*(4), 275–278.

*Westbrook, S. L., & Rogers, L. N. (1991, April). *An analysis of the relationship between student-invented hypotheses and the development of reflective thinking strategies.* Paper presented at the annual meeting of the National Association for Research in Science Teaching, Lake Geneva, WI. (ERIC Document Reproduction Service No. ED340605)

*Williams, R. L. (2003). *Critical thinking as a predictor and outcome measure in a large undergraduate educational psychology course.* Unpublished manuscript, University of Tennessee, Knoxville. (ERIC Document Reproduction Service No. ED478075)

*Wood, L. E. (1980). An "intelligent" program to teach logical thinking skills. *Behavior Research Methods and Instrumentation, 12*(2), 256–258.

*Zohar, A., & Tamir, P. (1993). Incorporating critical thinking into a regular high school biology curriculum. *School Science and Mathematics*, *93*(3), 136–140.

*Zohar, A., Weinberger, Y., & Tamir, P. (1994). The effect of the biology critical thinking project on the development of critical thinking. *Journal of Research in Science Teaching*, *31*(2), 183–196.

Authors

PHILIP C. ABRAMI is Professor, Research Chair, and Director of the Centre for the Study of Learning and Performance (CSLP), Concordia University, LB-581, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8; e-mail: *abrami@education.concordia.ca*. His research interests include educational technology, social psychologyof education and research synthesis.

ROBERT M. BERNARD is Professor of Education at Concordia University and a member of the CSLP, LB-581, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8; *e-mail: bernard@education.concordia.ca*. His methodological expertise is in statistics, research design and meta-analysis. His research interests include educational technology and distance learning.

EVGUENI BOROKHOVSKI holds a Ph.D. in Psychology and is a Research Project Coordinator at the CSLP, Concordia University, LB -581, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8; e-mail: *eborokhovski@education.concordia.ca*. His area of research interests includes cognitive and educational psychology, language acquisition and methodology of systematic reviews, meta-analyses in particular.

ANNE WADE (M.L.I.S.) is Manager and Information Specialist at the CSLP, Concordia University, LB-581, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8; *e-mail: anne.wade@education.concordia.ca*. Her expertise is in information literacy, information storage and retrieval, and research strategies. She has been a lecturer in the Information Studies Program in the Department of Education, Concordia University for over a decade; is Convenor of Campbell's Information Retrieval Methods Group; a member of Campbell's Education Coordinating Group and an Associate of the Evidence Network, UK. Wade has worked and taught extensively in the field of information sciences for twenty years.

MICHAEL A. SURKES is a doctoral student in Educational Technology at Concordia University, LB-581, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8; e-mail: *mike.surkes@education.concordia.ca*. His academic qualifications include degrees in physiological psychology, experimental psychology and philosophy, and his research concerns metacognition, metamotivation, and the development of complex and coherent conceptual frameworks.

RANA TAMIM is a Ph.D. Candidate in Educational Technology at Concordia University, LB-581, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8; e-mail: *rana.tamim@education.concordia.ca*. She is a part-time faculty member in the Department of Education and a research assistant at the CSLP. Her research interests focus on the role played by computer technology in facilitating learning. Other interests include science education, generative learning strategies, and meaningful learning.

DAI ZHANG is a Ph.D. Candidate in Educational Technology at Concordia University, LB-581, 1455 de Maisonneuve Blvd. West, Montreal, Quebec, Canada H3G 1M8; e-mail: *dai.zhang@education.concordia.ca*. She is the information technology educational adviser in John Abbott College and a research assistant at the CSLP. Her research interests focus on collaborative instructional design, E-learning system design and distance learning.