Reflections on Critical Thinking: Theory, Practice, and Assessment

Donald L. Hatcher
Baker University

Abstract

This autobiographical piece is in response to Frank Fair's kind invitation to write a reflective piece on my involvement over the last 30 years in the critical thinking movement, with special attention given to 18 years of assessment data as I assessed students' critical thinking (CT) outcomes at Baker University. The first section of the paper deals with my intellectual history and how I came to a specific understanding of CT. The second deals with the Baker Experiment (1988 to 2009) in combining instruction in CT with written composition, with a special focus on Deductive Reconstruction. The third section goes over our attempts at assessment and what we discovered by using three separate tests. The final section of the paper presents my conclusions about the challenges of teaching CT based on the Baker Experiment.

Keywords: Deductive Reconstruction, critical thinking, critical thinking and formal logic, critical thinking and written composition, critical thinking assessment, integrated approaches to critical thinking, standardized critical thinking tests.

I. Introduction

Like others in this series of articles, I want to thank Frank Fair for the invitation to, so to speak, "tell my story." It is somewhat unique because, unlike many who have been involved with the "critical thinking movement" (Fisher and Scriven, 1997; Johnson, 2012, 1996; Nosich, 2010, Paul, 2011; Scriven, 1976), my path did not involve a crisis or disenchantment with formal logic. In fact, an experience in graduate school led me to believe that formal logic, understood only as the propositional logic covered in most freshman logic texts, was a useful tool for clarifying arguments and thinking critically. Hence, while formal logic is portrayed by some as a useless tool for critical thinking (CT), I believed it is a mistake for CT teachers not to include it in their classes (Hatcher, 1999a, 1999b, 1986). How this belief that formal logic could be a valuable tool for understanding, creating, and evaluating arguments influenced my thinking about CT will be a theme of much of what follows. The first part of this reflection paper will go over the intellectual history that led me to endorse a specific conception of CT. The second will describe how this conception led to what I call "The Baker University (BU) Experiment," where my colleagues and I designed and taught a two-semester critical thinking and composition sequence that was required of all freshmen from 1990 to 2008. The third section will go over and reflect upon the considerable amount of assessment data collected over the 18 years we evaluated BU students' performance, including freshman to senior pre- to post-test scores. Finally, given the eighteen years of experience with that program and its assessment data, the final section will present some conclusions about teaching CT, students' ability to learn CT skills, and what an ideal CT program might look like.

II. The Path to a Specific Conception of Critical Thinking

In 1988, with the help of many of my BU University colleagues, especially Anne Spencer, we defined our conception of critical thinking (CT) as "thinking that attempts to arrive at a judgment only after honestly evaluating alternatives with respect to available evidence and arguments" (Hatcher and Spencer, 2005, p. 1). One might call it the "HEAP" conception of CT, that is, the "Honest Evaluation of Alternative Positions." I should point out that while this conception is my favored one, I have recently argued that we should give up the debate over what the "correct definition of CT" is and focus on skills necessary for effective critical thinking in any area (Hatcher, 2013). This definition served as a guiding principle in the development of the first edition of Anne Spencer's and my text, *Reasoning and Writing: From Critical Thinking to Composition* (1993), the Critical Thinking and Composition Program that was required of all BU University freshmen, and much of the research based on that program (Hatcher, 2011; 2009; 2006;1999a;1999b; 1998; 1997; 1995; 1992; 1990). So, what was the intellectual path that led to this conception of critical thinking?

In 1972, after previously flunking out of college and ending up serving in Vietnam, I was finally a senior majoring in English at Pittsburg State College in Pittsburg, Kansas. However, upon contemplating my future as a high school English teacher, I decided I had made a terrible mistake. It was easy to imagine spending evenings and weekends grading stacks of boring, badly-written English themes, which was not the way I wanted to live my life. (However, I am glad some good people are willing to grade all of those papers.) I knew that what

drew me to the study of literature was the ideas found there, rather than any aesthetic qualities. So, I decided to give up literature and major in philosophy. My advisor, knowing my lack of interest in the aesthetic qualities of literature, fully concurred with my decision. Fortunately, as it turned out, Pittsburg State College did not offer a philosophy major. So, in the fall of 1972, I transferred to the University of Kansas (KU) in Lawrence to study philosophy. Having the G.I. Bill gave me the luxury of relative financial security for the next few years. So, the fact that I might end up with an extended undergraduate education was not an issue. In addition, having grown up in the 60s, I thought having a rational life plan was evidence of a serious character flaw.

In those days, the KU philosophy department emphasized the study of the history of philosophy. The classics from Plato to Rawls (and Rawls was new stuff back then) were taught with a reverence usually afforded only sacred texts. Every course, except maybe Introduction to Logic, involved the close reading of primary texts. Skills that I would later learn were essential to CT were learned from the study of those works and from the professors who taught them. For example, on the first day of class in my Plato seminar, the late Professor Michael Young announced that we would not only be reading most of Hamilton and Cairns' *Collected Dialogues of Plato* (1978), but we would outline the arguments in each dialogue we read. The quality of these outlines would account for a large part of our grades. Similarly, in my Aristotle seminar, the late Professor Richard Cole announced that we would outline each paragraph of all assigned works in Richard McKeon's reader, *Introduction to Aristotle* (1973)—which turned out to be most of the book. Completing these assignments took hours of close reading and summation, but was well worth the effort. Not only did this exercise help me to understand the writings of Plato and Aristotle, it taught me the obvious truth so essential for CT: *in order to evaluate the reasonableness of a position or argument intelligently, one must first understand it*. To that end, one must be able to read closely, paraphrase when needed, and outline the arguments.

When I was an undergraduate philosophy major at KU, the value of these "sacred texts" or the wisdom of their authors was seldom questioned. Professor Young's love of Plato or Professor Cole's love of Aristotle was obvious, and questioning the reasonableness of these great thinkers was rare. It was so engrained that when I read Epictetus's claim that "death is nothing dreadful, or else Socrates would have thought it so" (p. 218), it did not occur to me that this might be an example of the appeal to false authority, and that what we really needed was an argument about the pros and cons of dying.

Only after reading a good deal of the history of philosophy did I realize that, because philosophers, even the most famous, did not say the same things about the same things, not everything said by the great thinkers could be true. To state the obvious, the history of philosophy is the history of critique. It consists of philosopher after philosopher critiquing those who either came before or were their contemporaries: Aristotle's critique of Plato, Locke's critique of Descartes, Marx's critique of Hegel, Rawls' critique of the utilitarianism, and so on. In fact, it seemed that the ability to understand and critique a position was a necessary condition for coming up with a new theory. This is because it is only through seeing the problems with an old theory is one moved to create a new one that might avoid those problems.

As it turned out, seeing the dialectical character of the history of philosophy was one of the avenues to my endorsing *fallibilism* as an epistemology. That is, one should always be open to the possibility that any belief, including beliefs about methods, could be wrong and should be revised. Later, this knowledge of the history of philosophy made the writings of Karl Popper (1965) and Harvey Siegel (1987; 1988) have a natural appeal. If it turns out that no one in the history of philosophy had the right answers, then it seemed reasonable to adopt a position that one ought to form beliefs in a tentative fashion, knowing that, even if one met his or her epistemic obligations, one might well be proven wrong, given new evidence and arguments. Honest inquirers, rather than some "true believer" who was wholly devoted to someone's prior philosophy, should continually evaluate the reasonableness of their beliefs, including methodologies.

With respect to the belief that formal logic could be a valuable tool for CT, a pivotal event occurred in the middle 1970s, years before I ever heard the words "critical thinking." In 1973, I took Introductory Logic using Copi's text on logic. A few semesters later, I enrolled in Professor Young's Plato Seminar. One day, while outlining one of the Platonic dialogues, I realized that many of Socrates' arguments followed the same deductive patterns I had studied in the logic class: *Modus Ponens, Modus Tollens, Disjunctive Syllogism,* or some combination of these. When outlining the dialogue, I discovered that it was easier to follow the arguments if I sketched them in formal notation in the margins of my book. For example, take Socrates' argument in the *Meno* that virtue is not knowledge. It is simply a series of *Modus Tollens* arguments: If virtue is knowledge (VK), then it can be taught (VT). If it can be taught (VT), there must be teachers (T). These teachers (T) are either the Sophists (S), any Athenian gentleman (AG), or poets (P). He then shows that none of the three candidates for teachers of virtue are acceptable (Not S, Not AG, and Not P). Hence, there are no teachers (Not T). If there are no teachers (Not T), then virtue cannot be taught (Not VT),

and therefore must not amount to knowledge (not VK) (Hatcher, 1996).

This accidental application of the simplest tools of formal logic to the arguments of Plato (and the arguments of many other treatises I later read) suggested that when these great thinkers and writers sat down to write an essay, they sketched their arguments in standard deductive form and then began to write. I hypothesized that this might explain why some writers were able to create such clear and powerful arguments in their writings, while others wrote in a way that seemed muddled and without clarity and focus. If the great thinkers of old proceeded in this way, it would be great to find a way to teach students to employ such a method. If they did, the quality of the essays of the average college student would be greatly enhanced. All that was needed was to first lay out the arguments for their thesis in standard deductive form and then evaluate them critically before writing the paper. This is the same thesis that is presented in Judge Aldisert's book *Logic for Lawyers*. Aldisert claims that the study of propositional logic and its application to argument should be required of all law school students. This was ten years before I heard the words "The Critical Thinking Movement."

The material in one other course, Philosophy of Logic, taught by Professor Don Marquis, was also significant in the development of my understanding of CT. In that course we spent a good deal of time studying Brian Skyrm's book, *Choice and Chance: An Introduction to Inductive Logic* (1966). Much of the book was devoted to Hume's *Problem of Induction* and the various attempts to solve it. Hume pointed out that any generalization based on past experience was going to be uncertain because the inference assumed that future experiences--experiments, samplings, observations, etc.--would be like the past. In other words, inductive reasoning, which is the foundation of scientific inquiry, assumed that nature was and will be uniform. Of course, this assumption, *The Principle of the Uniformity of Nature,* could never be proven. In studying the problem and the various failed attempts to solve it, the more I understood that any premise in an argument based on inductive inferences would always be uncertain. This, as well as the dialectical character of the history of philosophy, further supported my views about *fallibilism* as the proper attitude with respect to inquiry, truth claims, and arguments. Given the problem of inductive inferences, even if one's beliefs were the conclusion of a valid deductive argument, any premise that was the product of inductive reasoning would always be uncertain.

This was a welcome realization because, while I liked the clean rigor of formal deductive logic, it avoided a philosophical point of view that I always considered problematic. That is, too many philosophers, from Plato to Descartes, demanded that knowledge claims be certain. As a result they were forced to avoid empirically-based claims. I remember remarking to a fellow graduate student that if Descartes had been a race car driver, he would have never entered a race because there were no perfect cars on the market. It would always be possible that, no matter how well built and thoroughly inspected the car might be, some system could malfunction.

It was in the Philosophy of Logic class that I first came across the term "critical thinking." Towards the end of the term, while working on a paper on the problem of induction, I read Sir Karl Popper's paper, "Science: Conjectures and Refutations" (Popper, 1963). Near the end of that paper, in his discussion of what separated science from pseudo-science, Popper makes the distinction between what he calls "the critical attitude" and "the dogmatic attitude." As he puts it, "The distinction between dogmatic and critical thinking, or the dogmatic and critical attitude, brings us right back to our central problem. For the dogmatic attitude is clearly related to the tendency to *verify* our laws and schemata by seeking to apply them and confirm them, even to the point of neglecting refutations, whereas the critical attitude is one of readiness to change them—to test them; to refute them; to falsify them, if possible" (p. 50). While Popper was speaking of science, it seemed to me that the dogmatic attitude applied to any person who held a belief or theory and only looked for evidence to support the belief, either ignoring counter-evidence or never seeking it. In Francis Bacon's terms, the dogmatic thinker is guilty of "proof by enumeration." The critical thinker, on the other hand, is one who is always looking for arguments and evidence that might falsify his or her belief or disconfirm the theory in question. In other words, the critical thinker is one who is "honestly evaluating alternatives in terms of available evidence and arguments" (Hatcher and Spencer, p. 1). The justification is simply the dialectical history of science, philosophy, or any discipline. That is, historically growth in knowledge has generally been through criticism rather than through simply finding confirmation instances of one's favorite dogma.

Popper's insight about the value of testing or trying to falsify one's beliefs resonated nicely with what had already become one of my favorite works in the history of philosophy, John Stuart Mill's *On Liberty*. In the second chapter, Mill spends considerable time extolling the virtues of not only of knowing the arguments for one's favorite position, but also knowing and taking seriously the arguments against that position. As Mill so eloquently says,

> [O]n every subject on which difference of opinion is possible, the truth depends on a bal-

> ance to be struck between two sets of conflicting reasons. Even in natural philosophy, there is always some other explanation possible of the same facts; some geocentric theory instead of heliocentric, some phlogiston instead of oxygen; and it has to be shown why that other theory cannot be the true one; and until this is shown, and until we know how it is shown, we do not understand the grounds of our opinion. But when we turn to subjects infinitely more complicated, to morals, religion, politics, social relations, and the business of life, threefourths of the arguments for every disputed opinion consist in dispelling the appearances which favor some opinion different from it. The greatest orator, save one, of antiquity, has left it on record that he always studied his adversary's case with as great, if not still greater, intensity than even his own. *What Cicero practiced as the means of forensic success requires to be imitated by all who study any subject in order to arrive at the truth. He who knows only his own side of the case knows little of that. His reasons may be good, and no one may have been able to refute them. But if he is equally unable to refute the reasons on the opposite side, if he does not so much as know what they are, he has no ground for preferring either opinion.* The rational position for him would be suspension of judgment, and unless he contents himself with that, he is either led by authority or adopts, like the generality of the world, the side to which he feels most inclination (Mill, 1978, p. 34). [*Italics mine—D. H.*]

One could take much from this fecund passage, but two things made a lasting impression on me. First, when honestly evaluating an issue, the evaluation must take quite seriously the arguments in opposition to one's chosen position. Secondly, Mill implies that one must be open to changing one's mind after such an evaluation has been made. Much later, at the CT Conference in at Sonoma State University, when I heard Richard Paul speak of the distinction between "weak-sense" and "strong sense" critical thinking, I could not help but recall this passage, as well as Popper's distinction between "dogmatic thinking" and "critical thinking."

These simple ideas of summarizing arguments, putting them in valid deductive form, seeing the importance of understanding and taking seriously the arguments in opposition to one's own, turned out to be the foundation of what later became Baker University's Critical Thinking and Composition Program, an experimental venture in joining the formal logic and critical thinking with instruction in written composition that lasted eighteen years. Judging by the assessment results, it is an experiment that was successful, at least when compared to many other attempts to teach critical thinking. That is not to say there were not problems. More on that later.

III. The Baker University Program: An Experiment in Deductive Reconstruction

I was hired to teach philosophy at Baker University (BU) in the fall of 1978 to teach philosophy and a two-semester core general education course called "The Shaping of Western Thought." The mastermind behind the sequence was Dr. John English, a brilliant, widely read, history professor. The sequence was a natural for me as it included readings from many of the classic texts in philosophy: Plato, Aquinas, Rousseau, Mill, Marx, Spencer, Huxley, Dewey, and Sartre. Soon, I was asked to be on an *ad hoc* faculty committee charged with designing a capstone seminar required of all seniors. The course was supposed to integrate many of the skills and values of the liberal education BU students were receiving. It was called "Science, Technology, and Human Values." Its design required all seniors to choose a public policy issue brought about by current developments in science or technology, and then to research, prepare, present, and defend a fifteen- to twenty-page position paper that argued for a specific public policy over alternatives. Topics included issues like cloning research, water use policy, energy policy, computer privacy, file sharing policy, reproductive technologies, numerous medical issues, and defense policy, to name a few from a very long and ever-changing list.

When it became time to look for texts, we discovered that the issues in most texts that dealt with science, technology and society were dated. Because those planning the course wanted our seniors to grabble with current policy issues, we decided that the best way to achieve this goal was to allow the students' papers to serve as the primary text in the course. The chair of the *ad hoc* committee was the late Dr. Milford White, a well-respected and quite demanding professor of chemistry. He insisted that, besides presenting practical arguments for a proposed policy, significant parts of the papers would include an evaluation of the ethics of the proposed policy and *include a clear presentation of and response to possible objections or alternatives to the proposed policy*. It is perhaps worth noting that, while I had not yet formulated the HEAP conception of CT, the importance of "honestly evaluating alternatives" or the arguments that opposed one's preferred position turned out to be valued not only by philosophers such as J. S. Mill, but was apparently seen as equally important in the minds of practicing scientists.

The capstone began in 1979. Shortly after the senior capstone began, I realized the students needed a text-

book that acquainted them with the fundamentals of the scientific method, its strengths and limitations, with the standard ethical theories used in applied ethics and how to apply them, and with the nature and limits of technological development, as well as assessment. Such a text would be a great help to the students. So, in 1987, the first edition of my text, *Science, Ethics, and Technological Assessment* was published. Another issue arose early on, when those who taught the course complained that many seniors were seriously challenged (to put it mildly) when asked to produce such a paper. The primary difficulty was students did not understand how to construct or evaluate arguments. Given that their current general education requirements did not include anything having to do with logic and argument evaluation, this student difficulty should not have been a surprise.

In an attempt to solve this shortcoming, the initial plan was to integrate critical thinking skills into courses across the curriculum. BU's Academic Dean, Rick Torgerson, noticed a brochure for a conference at Sonoma State University on "critical thinking." The description convinced Torgerson (and me) that critical thinking skills might just be what our seniors needed to do a better job on their capstone projects. In 1983, with the help of a small grant from the Council of Independent Colleges, I attended my first CT Conference at Sonoma State University. My charge was to find good presenter who was willing to do a week-long workshop for the BU faculty on integrating CT skills into their courses.

As Ralph Johnson pointed out (2012), I too found the Sonoma Conference to be a very stimulating intellectual experience. Here were scores of teachers and scholars who shared an appreciation of the value of reasoned argument and who shared my long-held belief that the ability to understand and honestly critique a position, whether one's own or another's, was the most important part of college education.

That first year I attended scores of presentations, and subsequently we invited Jerry Nosich to come to BU the following summer and help us out. Over the fifteen years (1983-1998) that I attended the Sonoma Conference, I met many scholars and teachers who I still consider among my closest friends and most valued colleagues: Ralph Johnson, Mark Weinstein, Jerry Cederblom, Bob Ennis, Bill Dorman, Connie Missimer, Ed Damer, Harvey Siegel, Jerry Nosich, Ian Wright, Carol LaBar, Sharon Bailin, and Mark Battersby, to name only a few. A few years later, with the help of other grants, many of these folks were invited to do workshops at BU as we developed our freshman sequence in CT and Composition. Many from the Sonoma Conferences joined the Association for Informal Logic and Critical Thinking (AILACT), and they still today give or comment on papers at the AILACT sessions at one of the three yearly American Philosophical Association (APA) meetings, as well as other conferences such as Ontario Society for the Study of Argumentation (OSSA). I cannot say enough about the value of that network of scholars, teachers, and good friends. Richard Paul, his colleagues, and his student workers at Sonoma State University deserve tremendous credit for their hard work to set up these wonderful conferences. One of the greatest things about the Sonoma Conference was the opportunity to test drafts of their papers on CT theory or pedagogy that would later be published. Many of the editorial board members for *Informal Logic* or *INQUIRY: Critical Thinking Across the Disciplines* would be in the audience at the paper presentations and provide valuable feedback.

At the end of the 1983 conference, I invited Jerry Nosich to do a week-long workshop at BU. In looking back, why I chose Jerry is clear. As I have pointed out, from my days in graduate school when I saw that many of the arguments in philosophical texts could be easily summarized and put in standard deductive form (*Modus Ponens, Modus Tollens, Disjunctive syllogism,* or some combination of these*),* I believed that the simplest method for argument evaluation was what is now called Deductive Reconstruction. Summarize the arguments identifying the conclusion and supporting reasons, then, if needed, add a major premise to make the argument valid. Critique then amounted to focusing on the reasonableness of the premises. In his first book, *Reasons and Arguments* (1982), Jerry's approach was very similar. However, as he reported in his reflection piece (2010), he has now changed his mind about what he considers the proper methodology. I have not changed mine.

Jerrry came to BU in the summer of 1984 and gave a four-day workshop, addressing half the faculty in the mornings and the other half in the afternoons. Then we socialized in the evening. He showed us how we could evaluate the reasons for a position by treating the conclusion and reasons as an enthymeme, and then adding the major premise to turn the argument into a valid deductive argument, i.e. Deductive Reconstruction. His method is summarized in his text (Nosich, 1982, p. 142). as follows :

> Step 1. Paraphrase the argument so that you are sure you understand it.
>
> Step 2. Break the argument down into premises and conclusion.
>
> Step 3. Arrange the premises and conclusion in their logical order.
>
> Step 4. Fill in the missing premises needed to make the argument valid.

> Step 5. Criticize the argument for validity and the premises for truth.

By 1984, I had become the Director of the capstone program, Science, Technology, and Human Values. So, months after Jerry's workshop, I visited the Dean's Office and asked to look at the syllabi my colleagues had turned in for their classes. As far as I could tell, nothing had changed. At that point I began thinking seriously about developing a CT course that focused not only on evaluating arguments, but also on using the Deductive Reconstruction model to help students create clear and forceful arguments for their papers. However, I had no hope of developing a program that would be required of everyone and convincing our faculty to endorse the new requirement.

In 1987, BU's President, Dr. Dan Lambert, who was well aware of the problem of preparing all seniors for their capstone project, suggested, with due concern for the quality of BU's graduates, that I might apply for a Fund for the Improvement of Post Secondary Education (FIPSE) grant. If successful, I could use the funds to develop a course that taught students how to apply CT skills to writing papers. He pointed out that we had a relatively successful senior capstone program that had been in existence for eight years, showing that we were serious about supporting the program, but that it had a problem (the need for greater student preparation) that could be solved if we just had the resources to develop a required course that focused on CT and writing. This, he said, is the sort of problem that often gets funded, and he was right.

Before writing the grant application, I, and some of my colleagues who wanted to be involved in the project, went to our faculty and explained the situation. We had a problem that a successful grant application might solve, but the faculty needed to know that, if the grant application was funded, we would be adding at least one required course to our general education program. Adding non-departmental core requirements to a general education program is always risky. If multiple sections of a required course are needed, they must be taught by someone, and it is a zero-sum game. That is, each section of a course in the core that is taught by a professor from some department is a course in some major that cannot be offered. Unlike the traditional distribution requirements that comprise a typical liberal arts general education core, such non-departmental core requirements can often so reduce the major offerings in small departments that the major is lost. In spite of this, the BU faculty said to go ahead and apply for the FIPSE grant. Luckily, because of significant grant funding for the program from 1991 to 2005, department chairs did not have to face the "zero-sum problem" of staffing non-departmental core courses. We were able to use grant money to hire adjuncts to teach courses in the departments whose faculty members were teaching in the critical thinking and composition sequence. Money makes everything easier. (Consequently, I do not think it was accidental that when the grant funding ran out in 2005, various BU administrators started talking about the need for a new general education program.)

The FIPSE grant application went beyond CT and written composition and added the study of primary texts often taught in the humanities. Frankly, I never expected the application to be successful. However, David Arnold, our terrific FIPSE grant facilitator later told me that there was never any question about funding when the FIPSE Director first read the proposal. Given our situation with a quality senior capstone and the ever-increasing perceived value of CT in higher education, our proposed solution seemed to make very good sense. More importantly for FIPSE, it was the sort of solution that, if successful, other institutions might use.

So, in 1988, funded by FIPSE grants from the U.S. Department of Education ($68,500 and $106,110), a group of humanities faculty and I began planning the freshman critical thinking and written composition sequence. As it turned out, the basic idea of combining CT and composition had considerable appeal to outside funding agencies. That is, from 1991 to 2005, various parts of the program were supported and enhanced by four generous grants from the Hall Family Foundation ($175,000, $175,000, $396,000, and $120,000). Sometimes, one gets really lucky!

Given the grant support, I think it is fair to say that the BU program was developed and sustained under near ideal conditions. Financial support from various grants, totaling over $1,000,000 for fourteen years, allowed those working on the project the luxury to do what few groups charged with designing such programs are able to do. First, the faculty members were given one course released time each semester for two years. Second, the grants allowed many of those working on the BU freshman sequence to attend a number of Richard Paul's wonderful CT Conferences at Sonoma. Third, as we were working on the courses, we had funds to bring in some of the better-known and highly respected scholars in the area of CT. They came to campus and gave workshops on their conceptions of CT and how to teach it. They also evaluated our plans and materials for the freshman sequence, including drafts of the text we were writing for the course. These scholars included Harvey Siegel, Ralph Johnson, Ed Damer, Connie Missimer, and Jerry Cederblom. Prior to Harvey Siegel's visit in 1988, as a point of departure for the project, the professors working on the courses worked through his fine book *Educating Reason* (Siegel, 1988). I thought that we could make better progress on designing the courses

if we all understood and endorsed much of Harvey's defense of CT as an educational ideal. Fourth, and perhaps most importantly, from 1992 to 2003, the grant money from the Hall Family Foundation allowed my colleague Anne Spencer and me to put on two-week long summer workshops for those BU faculty who were going to teach a section of the required course for the first time, as well as anyone who was interested in integrating CT skills that were covered by all freshmen into their courses. The usual format of our summer workshops was to bring in a well-know expert on CT and have him or her do a day-long workshop on his or her specific conception of CT and how to teach it. Ralph Johnson, Connie Missimer, and Jerry Cederblom were regulars. Ralph was an immediate hit with the participants. It was clear to all that this man knew a whole lot about CT and the challenges of teaching it well. Ralph often started a workshop with a quote from Don Marquis, "If you make students think they are thinking, they will love you. If you make students actually think, they will hate you." (Given some of our student evaluations, that was indeed a fair warning.)

Jerry Cederblom was a natural complement to the rest of the workshop because we both used Deductive Reconstruction as the primary method of argument evaluation (Cederblom & Paulsen, 2006). He provided the participants with some good strategies for evaluating the major premise of a *Modus Ponens* or *Modus Tollens* argument. After the presentations by experts, the format of the workshops was to have the participants work through our text, *Reasoning and Writing: From Critical Thinking to Composition* (Hatcher and Spencer, 2006), and, on the last day of the workshop, do presentations on how they would integrate the material and skills taught in the workshop into their courses. The two-year funding from FIPSE, also allowed us the time to plan the first-semester course and run trial sections of it taught by faculty from different disciplines. Finally, the grants funded one-course released time each semester for its director. As a result, we ran weekly staff meetings where those teaching sections of the CT and composition courses went over the material in the text that would be taught that week. Released time allowed me to develop an instructors' manual with chapter summaries, teaching strategies, and answers to all of the exercises. With such institutional support, the financial resources made available by the grants, and a genuine desire by those involved in the project to teach our students to write clear, well-argued, and academically rigorous critical papers, I believed that if this program could not be successful, how could any succeed?

Throughout the planning of the program, and especially in writing the text, one concern was always whether what we were doing satisfied the requirements for critical thinking courses as described in California's Executive Order 338. I agreed with Michael Scriven that anything less might well be guilty of what he later called "The Prostitution of the Critical Thinking Requirement" (1991, 1992). Currently, Executive Order (EO1065) states relatively the same requirements as EO 338:

> *In critical thinking courses, students will understand logic and its relation to language; elementary inductive and deductive processes, including an understanding of the formal and informal fallacies of language and thought; and the ability to distinguish matters of fact from issues of judgment or opinion . . . students will develop the abilities to analyze, criticize, and advocate ideas; to reason inductively and deductively; and to reach well-supported factual or judgmental conclusions (p. 7).*

To this day, I do not think any course that does not at least include the knowledge and skills listed in this Executive Order can legitimately claim to be a course in critical thinking. However, that is not to say that other courses that do not actually teach these skills cannot be valuable by including assignments that ask students to exercise these skills (Hatcher, 2013).

The courses began, not unlike other critical thinking courses, by defining and explaining the nature and importance of critical thinking, understood as, "the honest evaluation of alternatives with respect to available evidence and arguments" (Hatcher and Spencer, 2006, p.1; Hatcher, 2000). In looking back on this definition, it is obvious how our focus on preparing students to write clear, well-researched, and well-argued research projects, especially in their senior capstone course, influenced the formulation of this definition. The honest evaluation of alternatives is exactly what we want students to do when they are researching an issue prior to writing a paper. Otherwise, the research turns out to be only a search to find support for one's prejudices. To this end, we emphasized only three basic skills: (1) the ability to understand a position, (2) the ability to evaluate the evidence and arguments given in its support, and (3) the ability to articulate one's judgment. Because this course was to be required of all freshmen, we were trying to keep things as simple as possible. For example, one might contrast this to Bob Ennis' well-known list of 14 dispositions and 12 skills, with 93 subcategories, not counting informal fallacies (Ennis, 1987).

Anne Spencer and I thought that showing students the importance of CT was important because otherwise the course might be seen as just one more troublesome requirement to be fulfilled with minimum effort and promptly forgotten (Bailin, 1999; Hatcher, 2001). With this in view, as an example of the importance of CT we showed how many social problems, such as those caused by prejudice against women and minorities, result from

people basing beliefs on insufficient evidence. Most importantly, we argued that a college education, at least those not focused only on job training, typically asks students to read material that expresses competing points of view. In many, if not most, disciplines, no single paradigmatic way of looking at issues has been found. So, if students are not to become cynical about discovering which position is more rational, they need tools for honestly evaluating the reasonableness of the alternative positions. Of course, for the more pragmatic student, we could not pass up showing them that the skills of being able to understand, evaluate, and articulate a position are essential tools for most professions. Students also read Plato's "Allegory of the Cave" in an attempt to help them recognize that many of their beliefs and values were not the products of a careful consideration of alternatives, but rather were functions of whatever the beliefs and values were projected on the wall of their specific "cave cultures" when they were young.

After showing the importance of critical thinking, instruction in basic critical thinking skills followed. Again, the three basic skills were (1) the ability to understand arguments, (2) the ability to evaluate arguments, and, (3) with the knowledge of valid deductive argument patterns, the ability to develop and articulate strong arguments in their papers. An ideal summary of a passage would include identifying the position (conclusion) and then stating the reasons (premises) given in its support, as well as any support given for the reasons. For example, a summary of a very simple arguments might take the following form "Smith believes X because A, B, C, and D." More complex arguments would of course have more complex summaries. We spent only three to four weeks studying the basics of deductive and inductive logic, and the more common informal fallacies. We also asked students to apply the tools to supplemental readings from Bacon, Paley, and Russell to illustrate how the skills could be used.

During the final weeks of the first semester we provided instruction in how the logical tools of Deductive Reconstruction can be applied to writing expository papers. An expository paper was understood to mean any paper where the student put forth a position and supported it with evidence and arguments. Students used some of the standard deductive argument patterns (*Modus Ponens, Modus Tollens,* and *Disjunctive Syllogism*) to construct arguments in support of various positions they might defend in a paper. For example, one way to argue for a position is to employ what we called a *Modus Tollens* strategy. Students began by negating the position in question and then showing how denying the position leads to unacceptable consequences, and so the position in question should be supported. For example, suppose one wanted to argue for teaching critical thinking to all students. One argument might go something like, "If we don't teach critical thinking, citizens will be easily duped by politicians. We do not want easily duped citizens in a democracy. Hence, we should teach critical thinking to all students."

In the spirit of "the honest evaluation of alternatives," after researching their topics, we asked students to construct the best arguments they could on both sides of an issue or question before deciding upon a thesis, and then evaluate the arguments. This involved evaluating the strength of the premises intended to support the position.

After evaluating the arguments both for and against a position, students constructed a thesis. The thesis was basically an outline of the major points of the entire paper. It included main objections, the position to be defended, and the supporting reasons. For example, a thesis for a paper for requiring CT of all students might look something like this: "In spite of the fact that not all students can learn the logic necessary to think critically, a CT course should be required of all students because CT skills are essential for most jobs, for intelligent voting, and for doing well in college." Students then created outlines for their position papers where each argument would be later turned into at least a paragraph in the paper.

After turning in their outlines, students set up a conference time and met with their teachers to discuss the outline. The focus of the conference was on the thesis, on the strength of the arguments given in its support, on the quality of the research, i.e., had the student looked at articles from respectable sources on each or different sides of the issue, and on whether the arguments on the other side and the responses were fairly treated. If the outline was acceptable, the student then began working on a draft. Once the draft was turned in and evaluated by the instructor, a draft conference was scheduled to go over the paper, looking at all of the points evaluated in the outline, plus grammar, mechanics, and the quality of the writing. All papers followed the same five-part pattern with an introduction, stating a clarification of the thesis, providing supporting reasons and arguments, addressing possible objections and replies to the objections, and finally a summation and conclusion. All were graded by the same grading rubric. (See appendix A.)

While we adopted the Deductive Reconstruction model, there are at least two concerns with it that need to be addressed. First, one troubling consequence of the "Deductive Reconstruction" model is there can apparently be valid arguments with reasonable premises both for and against a position. That is, the model seems to allow that both A and not-A can be reasonable positions. However, rather than be troubling, this consequence can

provide an explanation for why in many areas there are "deep disagreements." When the reasons in support of some contested position are listed, the preference of one specific position over another is relative to the value or weight the persons in the dispute place on the reasons they claim support their position (Hatcher, 2003). For example, in the debate in the U.S. over gun ownership, for some people, their firearm collection is one of the most important things in their lives. So, while pointing out the thousands of handgun deaths each year in the U. S. is recognized as a relevant reason for banning handguns, it does not carry as much weight relative to other reasons with those who dearly love their handguns. When trying to understand the current deep disagreements on important social issues, it is helpful to recognize that people might share the same values but rank them differently. The real issue is not so much having different values, but their weighting.

If this is so, I believe much could be gained if we forced students (and others) to justify with what they believe to be good arguments the ranking of the relevant values they bring to the table. One might then be able to evaluate those arguments and perhaps overcome what is often a stalemate between people who have such deep disagreements. This assumes, of course, that people's positions can be altered by the critical evaluation of their evidence and arguments—a necessary condition for "the honest evaluation of alternatives."

A second issue facing the Deductive Reconstruction model is to provide an explanation for how a series of reasons that individually do not provide sufficient support for a position can become sufficient when they are combined with other reasons that are also individually insufficient. In other words, why do we suppose that by combining a series of weak arguments we can ultimately create a good argument? For example, consider the argument, "We should support capital punishment because it reduces crime." On the Deductive Reconstruction model, that argument requires adding a major premise, "If something reduces crime, then it should be supported." This premise is clearly false. It would entail allowing many ethically and socially unacceptable practices, e.g., cutting off the hands of petty thieves, putting surveillance cameras in every room of every home and building, etc. However, the proponent of capital punishment might recognize this and add a series of additional reasons, none of which is individually sufficient to support his or her position, but when combined strengthen his or her position significantly: A) Capital punishment makes for safer prisons, for both inmates and guards, than life imprisonment. B) Capital punishment is in line with the basic principle of retributive justice: let the punishment be equal to the crime. C) Capital punishment provides satisfaction to the families of victims.

Perhaps a fruitful way to think about this apparently enigmatic phenomenon is to consider each reason as a judge in a court of law might think of a number of pieces of circumstantial evidence. While no one bit of circumstantial evidence is enough to convict the criminal, when a string of such bits of evidence is introduced, the person's guilt becomes more and more likely. Each new bit can make it less and less likely that the person on trial is innocent. Likewise, a string of relevant but individually insufficient reasons can be combined to provide sufficient support for a position.

Another way to understand how a string of individually inconclusive arguments might become conclusive is to think of evaluating reasons and arguments as analogous to how the quality of used cars is established. When *Consumer Reports* ranks used cars, it lists sixteen different things for each model. No one thing is conclusive, and not all carry the same weight. However, when all of the various items are added together (engine, transmission, fuel system, etc.), it allows those doing the ranking to conclude that one make and model is superior to others. The same can be said of competing arguments where lists of insufficient reasons are given for differing positions, and, when combined, the individually insufficient reasons become sufficient.

There are surely other issues surrounding the use of Deductive Reconstruction, otherwise more CT teachers would have adopted this simple approach to understanding and evaluating arguments. However, in my opinion, these are the most glaring.

The second semester of the course asked students to apply these same critical thinking skills and strategies to five sets of readings and write five additional critical papers, all including the same basic parts (albeit not necessarily in the same order): thesis, support, counter-arguments or objections, replies, and conclusion. Students were required to follow the same process. Where in the first semester, all sections used Anne Spencer's and my text, in the second semester teachers were free to choose any set of readings, as long as the papers followed the same format and went through the same process and were graded by the program's agreed upon grading rubric. This is because, when we all used the same text chosen by popular vote from among three alternatives, the vote was often 5-4-3. As a consequence the majority of us would be saddled with a text we would not choose to teach—not a pedagogically healthy situation. It is pretty evident that people cannot teach well material they do not much care for or material that is out of their area of expertise. Later, I saw that this also applies to the teaching of CT. But more on that later.

IV. Assessment Efforts: Eighteen Years of Puzzling Results

A. Assessment with the Ennis-Weir Critical thinking Essay Test (EW)

The question of whether or not to assess the new program was never an issue. The people from FIPSE made it clear that, if we were to fulfill our obligation and produce a program that others might emulate, evidence of its worth based on sound assessment methods was essential. In consultation with Stephen Norris, co-author with Bob Ennis of *Evaluating Critical Thinking* (1989), we chose to assess the critical thinking element of the program with the Ennis-Weir Critical Thinking Essay Test (E-W). Because the sequence integrated instruction in writing with critical thinking, Norris recommended this essay test. The E-W asks students to respond to an eight-paragraph Letter to the Editor with a nine-paragraph letter of their own that evaluates the reasoning put forth in each paragraph of the Letter to the Editor, judging whether it was good or bad and supporting their judgment with reasons. In a final paragraph, the response includes a judgment of the Letter to the Editor's overall argument.

According to the test manual, the E-W covers relevance, level of support, inconsistency, ignoring alternative explanations, equivocation, circularity, conditional inference, over-generalization, and emotive language (Ennis & Weir, 1985, p. 1). Both the format and the skills being tested seemed to match what we were trying to teach in the new program. One exception is the E-W does not put much emphasis on deduction. The highest possible score is 29.

The pre-test was given to all freshmen the first week of the fall semester. To encourage students to take the pre-test seriously, instructors told them that we were part of an important research project for the U. S. Department of Education and to do their very best. Also, because the post-test score counted 10% of the second semester's final exam and because some actually do worse on the final, we told students they could use either the pre- or post-test score, depending on which was higher as part of their final grade. The data in Tables 1, 2, and 3 indicate the outcomes for the E-W for the six years that we used it, including comparison group scores and freshman to senior scores for BU students.

While the effect size gains on the E-W were significant, we tried to further validate our findings by

Table 1

Comparison of Ennis-Weir Critical Thinking Essay Test Pre- and Post-Test Mean Scores for Baker University Freashmen 1990-1996

BU FRESHMEN	Pre E-W mean	Post E-W mean	Gain	Effect Size Gain/SD
1990-91 n = 169	6.3	12.4	6.1	1.19
1991-92 n = 119	9.4	12.2	2.8	0.51
1992-93 n = 178	6.8	12.6	5.8	1.05
1993-94 n = 178	8.1	14.1	6.0	1.13
1994-95 n = 164	7.5	13.0	5.5	1.00
1995-96 n = 169	6.9	12.9	6.0	1.13
Means (n =977) SD = 5.5	7.5	12.8	5.3	.96

testing students in two comparison groups. One was a standard logic class using Copi's *Introduction to Logic*, and the other was a critical thinking class. I am not sure what text they were using. The results of the comparison groups are in Table 2. Given the data in Table 2, if we could generalize from the comparison groups' performance, it was pretty clear that our two-semester course integrating CT skills with writing strategies was getting better results on the E-W than either a standard logic course or the one-semester CT course. However, our mean score of only 12.8 out of a possible 29 was only 44%--not a passing grade in most classes. So, while the effect-size gains were significant, the actual performance by our students was not so good.

Finally, after four years, we were able to test BU seniors in order to get an idea of how well not only the freshmen sequence was working, but whether our summer workshops with faculty from across the disciplines were getting the desired results. Table 3 has freshman to senior data for five years.

In looking back on the program, I can only speculate why the freshmen in BU's integrated, two-semester sequence did so much better on the E-W than the comparison groups who were taking the more traditional classes in logic and critical thinking (Hatcher, 2006). Educational research is notoriously uncertain. Compelling answers would take more controlled experiments where we carefully isolate as many variables as possible, e.g., teaching methods, textbooks, and teacher preparation. Nonetheless, there are some obvious differences with

Table 2

Comparison of Ennis-Weir Mean Test Scores Between a Standard Logic Class and a Two Semester CT Class

Class	Pre-test mean	Post-test mean	Difference	Diff./SD
Standard Logic (F94) n = 44	11.2	9.5	-1.7	-0.31
Critical Thinking (S92) n = 23	12.1	13.7	+1.6	+0.29

SD = 5.5

Table 3

BU Freshmen to Senior Comparison of Ennis-Weir Mean Scores

Year of Graduation	Freshman mean score	Senior mean score	Gain	Gain/SD
1995 (n = 119)	9.4	14.6	5.2	0.94
1996 (n = 88)	7.1	14.1	7.0	1.27
1997 (n = 80)	6.8	14.8	8.0	1.45
1998 (n = 58)	8.8	19.1	10.3	1.87
1999 (n = 42)	7.3	17.4	10.1	1.84
Means (n = 387) SD = 5.5	7.9	16.0	8.1	1.47

respect to our freshman sequence that may be causally related to the difference in performance between our students and the comparison groups.

One possible contributing factor may have been the simplicity and the repeated application of critical thinking skills in the two-semester sequence. Almost everything covered in the sequence focused on developing skills to evaluate the arguments found in what students read and the papers they prepared to write. The narrative we told students was simple: "Learn these skills and strategies and you will write better papers in all of your classes." Such simplicity and repetition may have made it easier for students to internalize the basic critical thinking skills, and apply them successfully to the E-W test. If simplicity and a unity of focus are important factors for student learning, it is possible that traditional logic courses confuse students by trying to cover too much material: deduction (with proofs), induction, informal fallacies, and sometimes quantification theory. By comparison, our two-semester sequence devoted only the first six weeks to the study of critical thinking and the logic. Most of what students covered early in the sequence was then applied repeatedly to what they read and in writing their papers. Because the focus was always on writing papers, the logical tools were seen as something that had obvious and immediate use in their educations. It is not just a set of skills needed to pass a test and then be forgotten (Hatcher, 2006, 2001).

A second contributing factor may be the time students spent using the skills since this was an obvious difference between our approach and the approach of the comparison groups. One would expect that a two-semester sequence, where relatively simple skills are repeatedly applied to papers, should yield better outcomes than the traditional one-semester stand-alone courses in critical thinking or logic. Being taught these standard logical skills and then applying them repeatedly to real papers for the rest of the academic year should get better results than one-semester courses. This evidence, more than anything else, argues for developing *a* serious *critical thinking across the curriculum* approach, where many instructors ask students in different disciplines to evaluate competing positions in their disciplines with respect to the evidence and arguments, and write papers defending their judgments. If the same song is sung often enough, then students might actually learn it. Unfortunately, when different teachers play the academic game by different rules, then students have a hard time deciding what is important and what is peripheral, let alone how to evaluate the rationality of a position (Bailin, 1999; Hatcher, 2001).

During the six years that we used the E-W test as an assessment tool, the data was encouraging. Hence, assuming the E-W's validity, then any claim is false that integrating CT with instruction in written composition and using Deductive Reconstruction as a primary method for constructing and evaluating arguments, is not a useful approach to CT instruction (Hatcher, 1999a). The numerous critiques of using formal logic to teach CT (Johnson, 2012, 1996; Nosich, 2010, Paul, 2011; Fisher and Scriven, 1997; Scriven, 1976) were simply not supported by our data. Compared to other studies, a freshman gain of nearly a full standard deviation was (and remains) substantial, even though the percentage correct was only 44%. For example, Pascarella and Terenzini estimate a freshman to sophomore effect-size gain of +0.34 and a freshman to senior gain of 0.55 (2005, p. 157). For students in the BU program, the three-year freshman to senior mean gain on the E-W was nearly three times that. Of course the problem may be that the effect size gains reported in the literature are based on other tests, such as the California Critical Thinking Skills Test or the Cornell Level Z CT Test rather than the E-W. Nonetheless, one advantage of using effect size as a measure is it would seem theoretically possible to compare outcomes, even if different tests were used. (More on this comparison later.)

I should point out that some, who would otherwise be inclined to use the E-W, might shy away from it because they believe, as an essay test, grading will be time-consuming and inter-grader reliability will be difficult to achieve. However, our experience showed it is fairly easy to achieve inter-grader reliability of .85 or better using pairs of bright, well-trained student workers. Also, grading time can be reduced if researchers choose a random sample of the essays and grade only those as opposed to grading every student's essay for assessment purposes. (We learned this lesson too late after doing double-blind grading of 1447 E-W essays.) One practical problem with grading the test is the graders should not know if they are grading a pre- or post-test. So, the grading cannot take place until after the post-tests are given. For us, that made a nice summer project for pairs of student workers.

B. Assessment with the California Critical Thinking Skills Test (CCTST)

In the fall of 1996, we changed tests and began to do pre and post testing with the California Critical Thinking Test (CCTST) (Facione and Facione, 1994). One reason for the change was a growing concern about the unusually strong post-test gains of our seniors. By 1999, the effect size gain from freshman to senior year was 1.84, and that seemed too high. I speculated that, after six years, the eight paragraphs that make up the E-W test may have become public, and seniors were in fact studying for the test. Secondly, while our inter-grader reliability was always high, as the teams grading the essays graduated and new teams took their place, there was really no way

to tell if the scores given by one pair of graders one year was comparable to the scores given be another team the next year. So, we switched.

The CCTST is a widely used test. It has been used in numerous other schools (see Table 5) and has been translated into several different languages. So, at least ideally, we believed it should allow an accurate comparison between our students' mean scores and those CT courses or programs trying other approaches. It is a 34-item multiple-choice exam that tests for many of the skills normally associated with critical thinking: interpretation, argument analysis and appraisal, deduction, logical puzzles, and induction. The data for the eight years of using the CCTST follows in Tables 4, 5, and 6. (Much gratitude is due Molly Ireland--now Dr. Molly Ireland--my dedicated and accomplished student assistant from 2004 to 2006, for grading hundreds of CCTST tests, entering the names and data on spreadsheets, and doing the statistics found throughout this paper and other papers dealing with the BU Experiment.)

It should be of interest that both the McMaster University and University of Melbourne courses used computer-assisted instruction to supplement in-class work (Hitchcock, 2004). Their positive gains indicate that computer exercises have a positive role to play in enhancing the critical thinking skills and test scores. It would be interesting to see their gains if they used the E-W essay exam.

When compared to the E-W, the relatively small gains on the CCTST were a bit disconcerting. With the average freshman year gain of only +2.6 points or an effect size of 0.57, BU students did only a little better than the mean gain of +2.42 points or an effect size of 0.54 of for the comparison groups (Hitchcock, 2004). This, however, was not close to the much larger gains on the E-W. We could not help but wonder which test more accurately reflected our students' CT abilities. (More on this later.)

Table 4
BU Freshmen Pre- and Post-Test Scores Using the California Critical Thinking Skills Test (CCTST)
Fall 1996 to Spring 2005

Academic Year	Pre-test mean	Post-test mean	Gain	Gain/SD
F96/S97 (n = 152)	15.14	18.49	3.35	0.75
F97/S98 (n = 192)	14.50	17.17	2.67	0.60
F98/S99 (n = 171)	15.81	17.90	2.09	0.46
F99/S00 (n = 153	15.91	18.28	2.50	0.53
F00/S01 (n = 184)	16.00	18.52	2.37	0.51
F01/S02 (n = 198)	15.30	17.47	2.17	0.48
F02/S03 (n = 221)	15.60	18.20	2.60	0.57
F03/S04 (n = 169)	15.40	18.10	2.70	0.60
F04/S05 (n = 170)	15.10	18.10	2.80	0.62
Means (n = 1617)	15.40	18.00	2.60	0.57

(Note: The standard deviation used with the CCTST is always 4.52. That was the SD used when the test was validated and that was also used by our comparison groups (Hitchcock, 2004.)

Table 5

Comparative Results From Other Groups Using the CCTST to Measure Critical Thinking Gains

Group	Pre-test mean	Post-test mean	Gain	Gain/SD
1990 Test Validation (n = 262)	15.94	17.38	1.44	0.32
2000 U. of Melbourne (n = 50)	19.50	23.46	3.96	0.88
2001McMaster University (n = 278)	17.03	19.22	2.19	0.49
2001 Monash University (n = 174)	19.07	20.35	1.28	0.28
2002 U. of Melbourne (n = 117)	18.85	22.10	3.35	0.73
Means (n = 831)	18.08	20.50	2.42	0.54

Table 6

Comparison of BU Freshman to Senior Scores on the CCTST 2000-2006

Year of graduation	Scores as Freshmen	Scores as Seniors	Gain	Gain/SD
2000 (n = 102)	15.2	19.4	4.2	0.93
2001 (n = 79)	14.3	18.3	4.0	0.88
2002 (n = 86)	15.8	19.2	3.4	0.75
2003 (n = 65)	15.8	19.7	3.9	0.87
2004 (n = 88)	15.9	20.2	4.3	0.95
2005 (n = 67)	15.3	20.5	5.2	1.15
2006 (n = 96)	15.6	20.5	4.9	1.09
Means (n = 455)	15.5	19.8	4.3	0.95

The average gain on the CCTST from the freshman to senior year was +4.3 points or 0.95 of a standard deviation, +1.7 points better than the +2.6 point average gain during the freshman year. While statistically, the literature claims this is better than expected (Arum and Roksa, 2011; Pascarella and Terenzini, p. 574), when I looked at these scores as a percentage of the possible points, it was pretty discouraging. A mean pre-test score of 15.5 out of 34 is only 46% correct. The mean post-test score of 19.8 out of 34 shows that in four years the student mean rose to only 58%. I cannot help but wonder whether this is a reasonable gain to expect given that students have had two semesters of CT instruction in the freshman year and then three additional years taught by faculty who claimed, at least in our summer workshops, to be committed to integrating the CT skills learned by our freshmen into their courses. My intuition is that it is reasonable to expect a greater gain. However, the 0.95 gain by the BU students was nearly double the average freshman to senior gains of 0.50 reported by Pascarella and Terenzini (2005, p. 574). This should have made me feel better about the program, even though, by traditional grading standards, the senior performance of 58% was still a failing grade.

C. Assessment Using the Cornell Level Z CT Test: Fall 2005 to Spring 2009

In the fall of 2005, we changed tests again and began using the *Cornell Level Z Critical Thinking Test* (Ennis, et al., 2004). We only have pre-post data on freshmen for four years. Of all of the multiple-choice exams, the Cornell was ranked highest by Fisher and Scriven (1997). The Cornell Level Z Critical Thinking Test (CLZ) is a 52-question multiple-choice test that can be completed in 50 minutes. According to the test booklet, the majority of the test items focus on deduction (n = 17) and induction (n = 23). The rest of the items test for observation, credibility, assumptions, and meaning (Ennis, et al., 2004). One reason for the switch to the CLZ was to compare our students' gains with those on other standardized tests. If the difference between the effect-size gains on the E-W and CCTST was so large, I was curious whether there would also be significant differences among the Cornell Level Z Critical Thinking Test (CLZ), the E-W, and the CCTST. (See Table 8) This would further support the theory that putative student outcomes were relative to the assessment tool used rather than actual CT skills. That would indeed be disturbing news for the assessment community. However, as Table 8 indicates, the difference between the effect size gain for the CCTST and the CLZ was not particularly significant: +0.57 and +0.52. The great difference was between the E-W (+0.97) and the other two tests (mean = 0.58).

One very troubling result of our assessment efforts with the CLZ was the range of student gains on the test relative to their first-semester instructor--the semester when the logic and CT skills covered in the test are supposed to be taught. While the mean gain for 684 students over the four years was 2.6 points or an effect size of +0.52, the mean gains relative to instructor ranged from +6.0 to -1.0. The +6.0 gain is an effect size of +1.20—a major gain relative to the 0.50 predicted by Pascarella and Terenzini (2005). However, what this shows is something most teachers probably already know: clearly some people are much better at teaching the material in a CT class than others, and therefore faculty development is essential when a course is required and has multiple sections. Secondly, the range of scores shows that one should be skeptical of reports of successful assessment data that are the product of one section of some course taught only one instructor. For example, prior to calculating the effect size gains relative to instructors, when Solon reported effect size gains of 0.87 on the CLZ for a class of 25 students that integrated CT and material in

Table 7
BU Freshman Scores for Cornell Test Level Z: Fall 2005 to Spring 2009

Academic Year	Pre-test mean	Post-test mean	Gain	Gain/SD
2005-2006 (n = 193)	27.2	30.0	2.8	0.56
2006-2007 (n = 175)	26.3	29.4	3.1	0.62
2007-2008 (n= 166)	26.8	28.7	1.85	0.36
2008-2009 (n = 150)	25.6	28.4	2.7	0.54
Means (n = 684) SD =5.0	26.5	29.1	2.6	0.52

General Psychology (2003), I was convinced that that was solid evidence that an integrative approach was far superior to all others (Hatcher, 2006). Now, such a conclusion may not have been warranted. It may just be the Tom Solon is one of those lucky professors who can actually teach CT, but if others tried his approach, they would fail miserably.

D. Comparing the Three Tests: Which Test More Accurately Measures Students' CT Abilities?

While the assessment data did not tell us that the BU experiment was a great success (gaining on average 2 to 3 points on a 52 question test after two semesters of instruction is hardly something to brag about), the disparity between effect size gains among the three tests deserves attention. Even though the students went through the same program (albeit with some variation in teachers), using the same CT text, and the same syllabus for eighteen years, the difference in average effect size gain during the freshman year between the E-W (0.97) and the CCTST and the CLZ (mean =0.54) is +0.43, nearly one half a standard deviation (Table 8). That is quite worrisome in this age of assessment and accountability.

Again, the same difference was true for the freshman to senior gains on the CCTST (0.95) and the E-W (1.47) (See Table 9 below). A difference of 0.52 of a standard deviation is a significant difference. If theses scores were made public, as was suggested in the Spellings Report (U.S. Department of Education, 2006), it would be the difference between claiming that one's institution had an average CT Program versus one of the best, when the only difference was what standardized tests were used for assessment. Before people start taking CT assessment scores too seriously, in the name of scientific replication, this is the sort of comparative research that needs to be done by many other CT teachers at numerous institutions.

The significant difference in effect size gains brings us to the essential question: Which test more accurately measures students' CT abilities? One response is to say that it depends on how one conceptualizes CT. If one thinks that a fairly clear understanding of deductive logic

Table 8

Comparison of Critical Thinking Effect Size Gains for Baker University Freshman Using Various Critical Thinking Tests From Tables 1, 4, and 7

Test	Pre-test mean	Post-test mean	Gain	Gain/SD
E-W (n = 977) SD = 5.5	7.5	12.8	5.3	0.96
CCTST (n = 1617) SD = 4.52	15.4	18.0	2.6	0.57
CLZ (n = 684) SD = 5.0	26.5	29.1	2.6	0.52

Table 9

Comparison of Effect Size Gains for BU Freshmen to Seniors Using the (EW) Ennis-Weir Test (SD = 5.5) and the (CCTST) California Critical Thinking Skills Test (SD = 4.52)

Test	Freshmen Pre-test mean	Seniors Post-test mean	Gain	Gain/SD
E-W SD = 5.5 (n = 387)	7.9	16.0	8.1	1.47
CCTST (n = 455)	15.5	19.8	4.3	0.95

and the ability to test scientific hypotheses are both essential skills for any student who claims to be a capable critical thinker, then the CCTST and CLZ would seem to be more accurate measures. Numerous questions on the tests focus on deductive validity (and much of what that concept entails) and how to test for the acceptability of or falsify a hypothesis. Hence, I cannot imagine students doing very well on the CCTST or the CLZ without having a CT course that spent s significant amount of time working on both deductive and inductive logic. So, if a class focuses only on informal logic and informal fallacies, students will struggle to do very well on the CCTST and the CLZ. However, if one's conception of CT does not emphasize scientific methodology and formal logic, as many informal approaches do not, then the E-W would assess student progress more accurately. That test focuses more on informal fallacies and the ability to evaluate evidence.

Another hypothesis to explain the difference in performance on the three standardized tests is the different formats. The act of taking E-W more closely resembled what BU students did throughout much of the two- semester sequence and what we wanted them to be able to do in real life more than the multiple-choice format of either the CCTST or CLZ. That is, we wanted students to read extended arguments, analyze their claims, evaluate their merit, and then articulate a thoughtful critique. Hence, given the BU students' experience, the act of taking the E-W was surely a more familiar and natural experience for them than working through a series of challenging multiple-choice questions on either the CCTST or CLZ. The CCTST is the sort of test, as Hitchcock (2004) and van Gelder (2004) have shown, that computer-assisted exercises can enhance students' performance significantly. Students can prepare by doing exercises that require the practice of discrete logical skills that can then be applied to the questions on the CCTST. But, it is not clear whether doing well on such exercises has any relationship to doing well in real situations that call for critical evaluation of reasons and arguments. So, because of its resemblance to real life situations that call for critical thinking, I believe that the E-W is in fact a better gauge of students' abilities to think critically *in real situations* than either the CCTST or the CLZ.

V. Concluding Thoughts

A. Results of Some Significance

So after 30 years of trying to teach students to think critically, and nearly 20 years of directing the BU Program, what conclusions can I reasonably draw?

First, it is clear that, even with a lot of help and a willing spirit, not everyone with a PhD can teach CT skills well. The material that seems quite natural to some with a specific philosophical orientation can be very difficult for others to master. But this should not be a surprise. The story of the post-modern critique of rationality may be analogous and helpful (Hatcher, 1994). In the 1980s, literature professors noticed that there could be several interpretations of a literary work, and there was no way to tell if one interpretation was the correct one. Then, against Aristotle's advice in the *Posterior Analytics* (Bk. I, Chaps. 9 and 32), they claimed professors should generalize and apply this discovery to all disciplines, including philosophy and the sciences. That turned out to be a mistake (Hatcher, 1991a).

Similarly, about the same time, I believe philosophy professors, especially those schooled in the Enlightenment tradition who were already inclined to value reasoned arguments above all else in education, saw the value of CT and concluded that critical thinking should be essential for all college graduates. With the help of people like Richard Paul, Ralph Johnson, Bob Ennis, Harvey Siegel, and others involved with the Sonoma conferences, the "Critical Thinking Movement" took shape. Soon, it was assumed that all professors would see the light and integrate CT skills into their courses. What a noble goal! However, my research indicates that some students can learn some of these skills, but many cannot. Some teachers, especially some philosophy teachers, are good at teaching the skills, but others cannot—even under ideal conditions. In fact, after the program ended in 2009, some teachers who had taught sections of the course for years confessed that they never quite understood the formal logic.

This should not have been a surprise. Nearly twenty years ago, the research by Richard Paul and his colleagues concluded that many college professors who endorsed the value of CT neither understood what it was or possessed the logical skills taught in typical CT courses (Paul, Elder, and Bartell, 1995). Why should this be a surprise? If calculus were deemed essential for the educated person, it would be no surprise if many teachers were not able to teach it. If playing a musical instrument well were deemed essential for human well-being, it would not surprise us if many teachers could not learn or teach music. I think Michael Scriven was on target when he wrote, " I believe reasoning is the most worthwhile, though one of the hardest, subjects to teach" (1976, p. xiv).

Second, it follows from the first concern that, given the disparity of assessment results relative to individual teachers, I no longer believe that critical thinking across the curriculum (CTAC) is a reasonable goal (Hatcher, 2013). If a fair number of teachers in a program such as BU's could not teach CT skills, some even making some of their students worse, how reasonable is to believe that teachers across the curriculum can teach the skills?

Third, the assessment results from three separate tests indicate that many students are not learning CT skills. While there were always an effect-size gain of +0.50 or better on the tests, after a year-long sequence, close to 30% of the students scored lower on the post-test than the pre-test. This suggests that many students are simply not inclined and maybe not able to learn the CT skills. Perhaps CT is something only a few people in special subcultures become really good at. My wife, who is an RN with many circles of friends, often remarks when we are at a social event with a group of academics, especially if they are philosophers, that we behave quite differently from other groups of her friends. That is, that we seemingly like to argue, almost about anything. She says, in her experience, most people do not behave like this. So, if her observations are representative, does that mean we academics are indeed anomalies and internalizing CT skills and dispositions is something only a few people are inclined to do? As Garrison Keillor said, "One can teach a bear to ride a bicycle, but that is not what the bear really wants to be doing." Given the choice, they will never ride bicycles. Perhaps the same is true for humans and CT.

What is unfortunate is that if many people in a democracy lack the reasoning skills covered in CT classes, great social harm can follow. This is not the case if citizens lack math skills or musical talent. All that is needed is that *some* are very good at music or math.

What if it turns out that many students are simply not capable of becoming critical thinkers? What if Aristotle was wrong, and we are not *rational animals*, but are instead hard wired to form and retain irrational beliefs and practices? There is plenty of research on this by social and cognitive psychologists, and their conclusions are not favorable to teaching CT (Ariely, 2008; Brafman and Brafman, 2008; Burton, 2008; Gilovich, 1991; Marcus, 2008; Tavris and Aronson, 2006, to name a few of many). If these psychologists are correct in their assessment of human dispositions and abilities (or lack thereof), what are the consequences for teaching CT? Is it any wonder that the gains on the tests are minimal? What should our expectations be if we are realistic?

Fourth, I think many courses that are called "critical thinking" are in fact not. This idea is nothing new. As noted earlier, in 1991, Michael Scriven published a timely paper warning us of "The Prostitution of the Critical Thinking Requirement." At that time, California had its Executive Order 338 requiring courses in CT of all graduates. It also carefully listed the knowledge and skills that were expected: students should "understand logic and its relation to language; elementary inductive and deductive processes, including an understanding of formal and informal fallacies, . . . (and) the abilities to analyze, criticize and advocate ideas, to reason inductively and deductively; and to reach well-supported factual and judgmental conclusions" (California, p. 7). It was clear then to Scriven, and I believe remains true today, that this sort of instruction is not provided in many classes that are called "critical thinking classes."

Finally, I think the good news is the assessment data from the BU Experiment showed that there is no reason to claim that formal logic cannot be an effective tool for teaching CT. This is good news because, in my own circle of acquaintances outside of the critical thinking movement, many are not enamored with the informal approaches found in many critical thinking texts. The BU Experiment showed that a program can be built around the method of Deductive Reconstruction and have results as good or better than other approaches using informal logic as the tool. Of course, when one looks at the percentage correct on any of the tests, one can see there is plenty of room for improvement, regardless of one's favorite method.

B. What Should Happen?

What, in an ideal world, should happen? First, in such a world, administrators would hire enough expert CT teachers, ones who had demonstrated success in teaching CT (Scriven, 2013), to cover multiple sections of a CT course of proven quality required of all freshmen. After all, according to an article in *Forbes*, CT is the skill most sought after by employers in 2013 (Casserly, 2012). Then, with the help of summer workshops akin to those that Anne Spencer and I did at BU, interested faculty members from a variety of disciplines would learn what the freshmen were learning about CT, and then they would design assignments for selected courses that required students to apply those CT skills. Then, following their freshman year, all students would be required to take one of these "applied CT courses" each year until they graduated as part of their general education requirements. While such a requirement was never part of BU's General Education Program, the CT test score gain that is most impressive are the freshman to senior effect-size gains for the E-W (+1.47) and the CCTST (+0.95). This surely was a function of so many of our faculty members participating in the summer workshops that went over the material the freshmen were learning. That sort of ideal program, I believe, would avoid what Michael Scriven appropriately called "the Prostitution of the Critical Thinking Requirement." Unfortunately, I doubt that university administrators will ever be convinced to hire the cadre of "expert CT teachers" necessary to offer enough sections of a required course. If the need was for more assistant football coaches, that would be another matter. Unfortunately, "philosopher college presidents" are about as rare as "philosopher kings," and willing the end of graduates who are critical thinkers seldom entails willing the

means.

References

Aldisert, R. J. (1997). *Logic for lawyers: A guide to clear legal thinking*. 3rd Ed. South Bend, IN: National Institute for Trial Advocacy.

Ariely, D. (2008). *Predictably irrational: The hidden forces that shape our decisions*. New York: Harper Collins.

Aristotle. (1973). *Introduction to Aristotle* (2nd Ed.). (Richard McKeon, Ed.) Chicago: University of Chicago Press.

Arum, R. and Roksa, J. (2011). *Academically adrift: Limited learning on college campuses*. Chicago, IL: University of Chicago Press.

Bailin, S. (1999). The problem with Percy: Epistemology, understanding, and critical thinking. *Informal Logic, (19)*2 & 3, 161-170.

Brafman, O and Rom Brafman. (2008). *Sway: The irresistible pull of irrational behavior*. New York: Broadway Books.

Burton, R. (2008). *On being certain: Believing you are right even when you aren't*. New York: St. Martins.

Casserly, M. (2012, December 12). The 10 skills that will get you hired in 2013. *Forbes*. Retrieved from http://www.Forbes.com/sites/megancasserly/2012/12/10/the-10-skills-that-will-get-you-a-job-in-2013/

California Executive Order 1065. Retrieved from <http://www.calstate.edu/eo/EO-1065.pdf>.

Cederblom, J. (2012, April). Problems with teaching how to reconstruct a common kind of argument. Paper presented at the AILACT Session of the APA Meetings, Seattle, WA.

Cederblom, J. and Paulsen, D. W. (2006). *Critical reasoning*. 6th Ed. Belmont, CA: Wadsworth Publishing.

Damer, T. E. (2001). *Attacking faulty reasoning*. 4th Ed. Belmont, CA: Wadsworth Publishing.

Ennis, R.H. (1987). A conception of critical thinking—with some curriculum suggestions. *APA Newsletter on Teaching Philosophy*. Summer, 1-5.

Ennis, R. H. (2011a). Critical thinking: Reflection and perspectives. Part I. *INQUIRY: Critical Thinking Across the Disciplines*. *26 (1), 4-18*.

Ennis, R. H. (2011b). Critical thinking: Reflection and perspectives. Part II. *INQUIRY: Critical Thinking Across the Disciplines*. *26 (2), 5-19*.

Ennis, R. H., Millman, J., & Tomko, T. N. (2004). *Cornell critical thinking tests, level X and level Z manual,(* 4th ed.) Seaside, CA: Critical Thinking Company.

Ennis, R.H. and Weir, E. (1985). *The Ennis-Weir critical thinking essay test*. Pacific Grove, CA: Midwest Publications.

Epictetus (1916). *Discourses and manual*. (P. E. Matheson, Trans.). New York: Thomas Nelson and Sons.

Facione, P. & Facione, N. C. (1994). *The California critical thinking skills test: Form A and B test manual*. Milbrae, CA: The California Academic Press.

Facione, P. (1986). Testing college level critical thinking. *Liberal Education,* Fall.

Fisher, A. and Scriven, M. (1997). *Critical thinking: Its definition and assessment*. Point Reyes, CA: Edgepress.

Gilovich, T. (1991). *How we know what is not so: The fallibility of human reason in everyday life*. New York: Free Press.

Hatcher, D.L. (1990). Reasoning and writing: Critical thinking for everyone? *INQUIRY: Critical Thinking Across the Disciplines*, (6) 4, 18-21.

Hatcher, D. L. (1991a). Can critical thinking survive the postmodern challenge?" *INQUIRY: Critical Thinking Across the Disciplines*, (7) 1, 16-17.

Hatcher, D. L. (1991b). When poets teach critical thinking. *Thinking*, (9) 3, 46-47.

Hatcher, D. L. (1992). Epistemology and Pedagogy. *INQUIRY*, (10) 2, 14-16.

Hatcher, D. L. (1994) Critical thinking, postmodernism, and rational evaluation. *Informal Logic*, (XVI) 3,197-208.

Hatcher, D. L. (1995). Combining critical thinking and composition: The whole is greater than the sum of the parts. *INQUIRY: Critical Thinking Across the Disciplines*, (XV) 2, 20-36. Reprint: *Achieving Distance Learning*, (1996) University of South Africa Press.

Hatcher, D. L. (1996). Plato's 'Meno': A model for critical thinkers. *INQUIRY: Critical Thinking Across the Disciplines*, Vol. XVI, No. 1, 1-8.

Hatcher, D. L. (1997). Three theories of rationality: Some pedagogical consequences for critical thinkers. *INQUIRY: Critical Thinking Across the Disciplines,* (XVII) 2, 4-19.

Hatcher, D. L. (1999a). Why formal logic is essential for critical thinking. *Informal Logic, (19)*1, 77-89.

Hatcher, D. L. (1999b). Why we should combine critical thinking and written instruction. *Informal Logic, (19)*2 & 3, 171-183.

Hatcher, D. L. (2000). Arguments for another definition of critical thinking. *INQUIRY: Critical Thinking Across the Disciplines,* 20(1), 3-8.

Hatcher, D. L. (2001). Why Percy can't think: The departmentalization of critical thinking. *Informal Logic*, 21(2), 173-181.

Hatcher, D. L. (2003). The role of personal values in argument evaluation. *Informal logic @ 25:*

Proceedings of the Windsor conference. J. A. Blair, D. Farr, H. V. Hansen, R. H. Johnson, & C. W. Tindale, (Eds.), Windsor, Ontario: OSSA.

Hatcher, D. L. (2006). Stand-alone versus integrated critical thinking courses. *Journal of General Education*. (55) 3-4, 247-272.

Hatcher, D. L. (2009). Assessing critical thinking skills: A fifteen-year perspective. *Critical thinking, education, and assessment*. Althouse Press, University of Western Ontario, London, Ontario.

Hatcher, D. L. (2011). Which test? Whose scores? Comparing standardized critical thinking tests. *New Directions in Institutional Research*. Ed. Jeremy Penn. San Francisco: Jossey-Bass.

Hatcher, D. L. (2011). *Science, ethics, and technological assessment*.(5th ed.). Boston, MA: American Press.

Hatcher, D. L. (2013, May). Is critical thinking across the curriculum a plausible goal? Paper presented at the Ontario Society for the Study of Argumentation conference, University of Windsor, Windsor, Ontario.

Hatcher, D. L. & Price, L. (1998). Why critical thinking and composition belong together (and vice versa). *INQUIRY: Critical Thinking Across the Disciplines*, Vol. 17, No. 4, 19-30.

Hatcher, D. L. & Spencer. L. A. (2006). *Reasoning and writing: From critical thinking to composition*. 3rd Ed. Boston: American Press. (Original work published by Copley Publishing, 1992 and Rowman and Littlefield, 1993.)

Hitchcock, D. (2004). The Effectiveness of Computer Assisted Instruction in Critical Thinking. *Informal Logic*, 24 (3), 183-217.

Johnson, R. H (2009). The implications of the dialectical tier for critical thinking. In J. Sobocan and L. A. Groarke (eds.) *Critical thinking education and assessment*. London Ontario: Althouse Press.

Johnson, R. H. (1996). *The rise of informal logic*. Newport News, VA.: Vale Press.

Johnson, R. H. (2000). *Manifest rationality: a pragmatic theory of argumentation*. Mahwah, N.J.: Lawrence Erlbaum Associates.

Johnson, R.H. (2012). When informal logic met critical thinking. *INQUIRY: Critical Thinking Across the Disciplines*, 27(3), 5-14.

Johnson, R, H. & Blair, J. A. (1994). *Logical self-defense*. New York: McGraw-Hill.

Lipman, M. S. (2003). *Thinking in education*. 2nd Ed. New York: Cambridge University

Marcus, G. (2008). *Kluge: The haphazard construction of the human mind*. Boston: Houghton Mifflin.

Mill, J.S. (1978). *On liberty* (Elizabeth Rapaport, ed.). Indianapolis: Hackett Publishing.

Missimer, C. (1986). *Good arguments: An introduction to critical thinking*. Englewood Cliffs, NJ: Prentice Hall.

Norman, G.R., Sloan, J.A., and Wyrwich, K. (2003). Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Medical Care,* 41, 582-92.

Norris, S. P. and Ennis R. H. (1989). *Evaluating critical thinking*. Pacific Grove, CA: Midwest Publications.

Nosich, Gerald (1982). *Reasons and arguments*. Belmont: Wadsworth Publishers.

Nosich, G. M. (2010). From argument and philosophy to critical thinking across the curriculum. *INQUIRY: Critical Thinking Across the Disciplines,* 25(3), 4-13.

Nosich, G. M. (2012). *Learning to think things through: A guide to critical thinking across the curriculum. 4th ed*. Boston: Pearson.

Pascarella, E.T. & Terenzini, P. T. (2005). *How college affects students, vol. 2: A third decade of research*. San Francisco: Jossey-Bass.

Paul, R. W. (2011). Reflections on the nature of critical thinking, its history, politics, and barriers, and its status across the college/university curriculum. Part I. *INQUIRY: Critical Thinking Across the Disciplines,* 26 (3), 5-24.

Paul, R., Elder, L., Bartell, T. (1995). "Study of 38 Public Universities and 28 Private Universities to Determine Faculty Emphasis on Critical Thinking in Instruction: Executive Summary." Accessed March 15, 2013. http://www.criticalthinking.org/pages/study-of-38-public-universities-and-28-private-universities-/598.

Plato. (1973). *The collected dialogues of Plato*. (Edith Hamilton and Huntington Cairns, Eds.). Princeton: Princeton University Press.

Popper, K. (1963). Science: Conjectures and refutations in *Conjectures and refutations: The growth of scientific knowledge*. New York: Harper and Row, pp. 33-65.

Possin, K. (2013). A fatal flaw in the collegiate learning assessment test. *Assessment Updates, 25 (1)*. 8-12.

Scriven, M. (1976) *Reasoning*. New York: McGraw-Hill.

Scriven, M. (1991). The prostitution of the critical thinking requirement, Part I. *CT News,* 10(2), 1-5.

Scriven, M. (1992). The prostitution of the critical thinking requirement, Part II. *CT News,* 10(3-4), 6-12.

Scriven, M. (2013) Commentary on: Donald Hatcher's "Is critical thinking across the curriculum a plausible goal?" Paper presented at the Ontario Society for the Study of Argumentation conference, University of Windsor, Windsor, Ontario.

Siegel, H. (1987). *Relativism refuted*. Boston: D. Riedel Publishing.

Siegel, H. (1988). *Educating reason*. New York:

Routledge, Inc.
Skyrms, B. (1966). *Choice and chance: An introduction to inductive logic*. Belmont, CA.: Dickerson Publishing.
Solon, T. (2003). Teaching critical thinking: The more, the better! *The Community College Enterprise, 9(2),* 25-38.
Tavris, C. and Elliot Aronson. (2007). *Mistakes were made (but not by me): Why we justify foolish beliefs, bad decisions, and hurtful acts*. Orlando, FL: Harcourt.
U.S. Department of Education (USDOE), *A Test of Leadership: Charting the Future of U.S. Higher Education*. Washington, D.C.: U.S. Government Printing Office, 2006.
Van Gelder, T. & Bissett. M. (2004). Cultivating expertise in informal reasoning. *Canadian Journal of Experimental Psychology,* 58 (2), 142-152.

Author Information

Donald Hatcher is Professor of Philosophy in the College of Arts and Sciences at Baker University in Baldwin City, KS, 66006. From 1979 to 2009, he was Director of Baker's Liberal Arts Program described as "The Baker Experiment" in this reflection paper. From 1988 to the present, he serves as the Treasurer for the Association for Informal Logic and Critical Thinking (AILACT). He can be contacted at dhatcher@bakeru.edu.

Appendix A: Rubric for Grading Critical Thinking Papers

Comments:

Introduction/Thesis
Establishes importance of issue
Leads naturally to thesis
Thesis is clear and complete___/10
Argumentation
Definition and clarification present when necessary
Argumentation is clear and persuasive
Ideas are non-trivial and show evidence of student's own thinking
Objections/Replies
Objections are substantive, adequately, and fairly presented
Strong objections are not overlooked
Objections are answered adequately___/40
Conclusion
Summary of all issues is presented
Conclusion agrees with thesis without merely repeating___/5
Research
Sufficient research, reliable sources
Supporting evidence is well-integrated and credited properly
Proper MLA format is followed exactly
All sources cited on the works cited page are cited in the paper
Proper format is present for in-text citations___/15
Language Usage and Mechanics (see the back of this page for details)___/15
Overall Coherence and Organization
Paragraphs are adequately developed and organized
Transitions between paragraphs is clearly present/15

(Numbers in parentheses refer to sections in *The Bedford Handbook*)

Section	Error	Section	Error
(19)	Sentence fragment	(10-11, 18b, 26)	Faulty idiom
(20)	Comma splice	(16-18)	Faulty diction ("wrong word")
(20)	Run-on sentence	(36)	Apostrophe absent or misused
(21)	Faulty subject-verb agreement	(32-39)	Other punctuation fault
(27-28)	Other verb form error	(17)	Sexist, slang, or colloquial language
(22)	Faulty pronoun agreement	(40-45)	Misspelling or typo, faulty abbreviation or capitalization
(23)	Faulty pronoun reference	(53-54)	Faulty format (title, margins, citations, bibliography, etc.)
(24-25)	Faulty pronoun case		
(9)	Faulty parallelism		
	Other sentence structure fault		