
9.

The Institutional Assessment of Critical Thinking

A Fifteen-Year Perspective

Donald L. Hatcher

Sometimes good things happen accidentally. People inherit money from distant relatives whom they have never met. Some very lucky people meet the loves of their lives quite by chance. In education, too, good things may happen accidentally. Let me describe an instance of my own.

In the early 1970s, I was simultaneously enrolled in an introductory logic course and a seminar in Plato. One day, after having studied some of the standard deductive patterns of reasoning (*modus ponens*, *modus tollens*, *disjunctive syllogism*) in the logic class, I was working through one of the Platonic dialogues when I realized that many of Socrates's arguments followed the same patterns I had learned in the logic class. I discovered that it was easier to follow the arguments if I sketched them in formal notation in the margins of the book. This was a useful exercise because a significant part of our grade was determined by the quality of our outlines of the dialogues we read (my professor was committed to a fundamental principle of critical thinking: that students cannot adequately criticize what they have not first understood, and outlining what one reads is a good way to achieve this first end).

The accidental application of the simplest tools of formal logic to the arguments of Plato (and the arguments of many other treatises read in graduate school) suggested an idea. Perhaps, when

these great thinkers and writers sat down to write an essay, they sketched their arguments in standard deductive form, and then proceeded to write. I hypothesized that this might explain why some writers were able to create such clear and powerful arguments, whereas others wrote in a way that seemed muddled and unfocused. If the great thinkers of old proceeded in this way, why not find a way to teach college students today to employ this method? It seemed to me that the essays of average college students would be greatly enhanced if they first sketched the arguments for their theses in standard deductive form and evaluated them critically before writing. This was ten years before I heard the phrase “critical thinking.”

These simple ideas were the genesis of what became, years later, Baker University’s Liberal Arts Program, an experiment in joining the disciplines of logic and critical thinking with instruction in written composition. This is an experiment supported by \$1,000,000 in grant funding (sometimes one gets really lucky).¹ Judging by the assessment results, it is an experiment that is relatively successful when it is compared to many other attempts to teach critical thinking.²

My chapter in this book explains the development and operation of the Baker program and reports on our ongoing assessment efforts. Because our approach to teaching critical thinking was unique — with some skeptics saying we were not teaching composition, others claiming that we were not doing justice to logic and critical thinking — careful assessment has been an extremely important part of the program. In addition to providing evidence for the success of our approach, our model shows how easily assessment can be implemented, and provides fifteen years of data that others can use for comparison.

The ability to compare is one of the great benefits of standardized critical thinking tests. These assessments allow us to compare the educational outcomes of different attempts to teach critical thinking. Such comparisons are the best way for teachers of critical thinking to find ways of teaching critical thinking that work (and do not work), and such comparisons, in turn, can inform the development of our testing instruments. In keeping with the latter,

one of the interesting, aspects of the Baker history of assessing critical thinking is our ability to compare the results of the two tests that we used. From 1990 to 1996, we used the *Ennis-Weir Critical Thinking Essay Test* (E-W), and from 1996 to the present, we have used the *California Critical Thinking Skills Test* (CCTST). Although the results have been positive in both instances, they have not been the same.

The History of Baker's Liberal Arts Program

Baker University's General Education Program of fifty-plus college hours contains three specially designed courses required of all students: a two-semester freshmen sequence (LA 101 and LA 102) and a senior capstone (LA 401). The freshmen sequence, "Critical Thinking and Effective Writing" and "Ideas and Exposition," provides all Baker freshmen with instruction in formal logic and critical thinking skills, and shows how this knowledge can be used successfully in writing expository prose. The senior capstone seminar, "Science, Technology, and Human Values," asks each senior to choose a public policy issue brought about by current scientific or technological developments, and then to research, prepare, present, and defend a fifteen- to twenty-five-page position paper that argues for a specific public policy with respect to the issue. Topics include cloning, water-use policy, energy policy, reproductive practices, numerous medical issues, and defence policy, to name a few of over one hundred possible issues. A significant part of the paper includes a critical analysis and response to alternative policies or objections to the proposed policy. Students must consider the ethical consequences of each alternative under consideration.

The senior capstone, LA 401, began thirty years ago in 1979, and it was not long before the faculty members who were teaching sections of the course realized that many of our seniors were seriously challenged when we asked them to write a critical or argumentative paper. The primary difficulty was their lack of understanding logic: what arguments were, how one constructs

them, and how one evaluates them. To address this shortcoming, we began planning the required freshmen critical thinking and composition sequence, LA 101 and LA 102, in 1988. This project was funded by two grants from the Fund for the Improvement of Postsecondary Education (FIPSE) provided by the United States Department of Education. It has since been supplemented with a series of four grants from the Hall Family Foundation. A good deal of the Hall Grant money has gone towards faculty development, dissemination, and assessment. Those in the Hall Family Foundation are committed to the idea that the Baker method of teaching writing and critical thinking needs to be circulated more widely in education.

Although the primary reason for developing the freshmen critical thinking and composition sequence was to better prepare Baker seniors for the LA 401 capstone experience, Baker faculty members believed, more generally, that critical thinking skills are the skills that students need if they are to evaluate alternative positions and write carefully argued papers for any of their courses.³ The critical thinking and composition sequence thus provides all of our entering students with skills essential for success in their college courses. The teaching of these skills includes instruction in paraphrasing and summarizing difficult readings; logical techniques for evaluating the reasonableness of beliefs and arguments; and logical strategies for developing strong arguments to support students' ideas used in papers across the curriculum.

The Critical Thinking and Composition Sequence

What are the Baker freshmen courses like? For those who worked on the Baker project, getting clear on what exactly we meant by critical thinking was extremely important. We understood that our conception of **critical thinking** would greatly influence both the structure and content of the courses. We examined some of the standard definitions of critical thinking and were not enamoured with any of them.⁴ We wanted a definition that would be as clear

and concise as possible, so that both we and the unconvinced would know what we were talking about when we discussed the new sequence. The definition needed to be easy to explicate to students, faculty, and administrators, showing why critical thinking is an essential educational goal.

We wanted a definition that referred specifically to the criteria that should be used for critical judgment. Otherwise, one could not expect agreement over what counts as a reasonable position. The definition should imply that critical thinking has broad educational utility, that it is applicable to many disciplines. It should be obvious from the definition that students in art, literature, political science, or history can benefit from learning logic and critical thinking skills. The definition, moreover, should allow people to distinguish critical thinking from other cognitive activities such as creative thinking, problem-solving, and logical inference. It should provide enough guidance to faculty to allow them to construct tests and assignments to assess whether students have acquired the appropriate skills and dispositions.

Given all of these constraints, the definition we chose defines critical thinking as “thinking that tries to arrive at a judgment only after honestly evaluating alternatives with respect to available evidence and arguments.”⁵ Properly understood, we believed that this definition could provide the needed foundation for a course integrating instruction in logic and expository prose. That is, when a student is assigned a position paper, the process will include the honest evaluation of alternative positions before the position to be defended is chosen. This means getting clear on the arguments for each alternative, and then evaluating their strengths and weaknesses. The paper’s thesis will be the position with the strongest support and weakest objections.

Our courses begin, like many other critical thinking courses, by explaining the nature and importance of critical thinking. The text *Reasoning and Writing: From Critical Thinking to Composition* (2006) gives a number of arguments, both practical and theoretical, for the value of critical thinking instruction. We show how many social problems, such as those resulting from prejudice against women and minorities, are the result of basing beliefs on

insufficient evidence and hasty generalizations. In addition, many personal problems, especially among the young, stem from poor judgment or a failure to evaluate honestly the available alternatives before making a decision. We begin the course by reading Plato's "Allegory of the Cave" in an attempt to get students to recognize how many of their ideas are a function of values projected on the walls of their specific "caves" when they were young. This approach to the beginning of the course clearly supports Hare's position that the claim "critical thinking texts and courses tend to teach political *conformity*" is indeed fallacious (see Hare, this volume). It is difficult to free students from the effects of living in a specific culture, and its values and ideas, in a few classes, but we do try to convince them that becoming a critical thinker is in their own interest.

After showing the importance of what we are asking students to learn, we follow with instruction in how to read, paraphrase, and summarize difficult prose and how to identify the arguments it contains. Because many students come to college with weak reading skills, learning to read carefully, with an eye to the evidence and arguments for any claim, is an essential skill. To address this, we spend a good deal of time teaching students how to paraphrase and ultimately summarize what they read. The goal is to read an argumentative passage, identify the position (conclusion), and identify the reasons (premises) given in support of the conclusion (e.g., "Smith believes X because A, B, C, and D"⁶).

Once students can identify and summarize arguments, the next step is instruction in argument evaluation. To this end we employ the technique of Deductive Reconstruction.⁷ That is, each of the arguments is put into standard deductive form — *modus ponens*, *modus tollens*, *disjunctive syllogism*, or some combination of these. The theory behind **Deductive Reconstruction** is the following: if the arguments are in a valid deductive form, then, for purposes of evaluation, the main question is whether the premises are reasonable or whether they need further support. Evaluating the level of support for the premises usually involves understanding inductive logic. We spend only three to four weeks

— an unusually brief time compared with other critical thinking courses — studying deduction, induction, and a few of the more common informal fallacies. There are other methods for evaluating arguments, but we decided to focus on these because of their simplicity, transferability among other disciplines, and usefulness in constructing arguments that will ultimately form the backbone of students' papers (remember my experience with logic and the Platonic dialogues). Most students have little trouble mastering the techniques we teach, though faculty who are not trained in philosophy sometimes struggle with the material when they first begin to teach it.

The final weeks of the semester provide instruction designed to show how the tools of Deductive Reconstruction are useful in writing expository papers.⁸ We teach students how to use some of the standard argument patterns (*modus ponens*, *modus tollens*, and *disjunctive syllogism*) to construct arguments in support of positions they might defend in a paper. For example, one way to argue for a position is to employ what we call a *modus tollens* strategy. Students begin by negating the position in question, show how this leads to unacceptable consequences, and conclude that the position in question should be supported. If we wanted to argue for teaching critical thinking to all students, such an argument might go something like this: "If we do not teach critical thinking, citizens will be easily duped by politicians. We do not want that in a democracy. Hence, we should teach critical thinking to all students."

In a spirit that embraces the honest evaluation of alternative positions, we ask students to construct the best arguments they can on both sides of an issue before deciding upon a thesis. Often, weak papers are the result of students picking a position, not because they have honestly evaluated the alternatives, but because it agrees with their deeply felt intuitions or "gut feelings." In such cases, students fail to recognize the extent to which they have been socialized by their culture to think in certain ways about specific issues — even though there may be good reasons for alternative conclusions. We use Plato's "Allegory of the Cave" to underscore this point.

After evaluating arguments for and against different sides of an issue, students construct theses and create their outlines for their position papers. They then meet with teachers to discuss an outline. The focus of the conference is the thesis and the strength of the arguments given in its support. If the outline is judged acceptable, the student begins writing a draft. This, too, is evaluated by the instructor. All papers follow the same four-part pattern: an introduction, clarification, and thesis; supporting reasons and arguments; possible objections and replies; and a summation and conclusion.

The second semester of the freshmen course asks students to apply these same critical thinking skills and strategies to five sets of readings and to write five additional critical papers. All papers include the same basic parts — thesis, support, counter-arguments or objections, replies, and conclusion (though not necessarily in this order) — and are composed in a manner that follows the same process students used in the first semester. Although all sections of the course use the same text as a basis for the first semester's paper, teachers are free to choose any set of readings in the second semester, on the understanding that all the papers follow the same process and are graded according to the same rubric. Given that instructors come from many different disciplines, finding one text that all teachers felt equally enthusiastic about proved to be an unrealistic goal. These critical thinking courses differ from traditional courses in a number of ways. Unlike most critical thinking courses, they teach students to use formal logic and critical thinking skills to argue for and critique positions in their papers. The time spent on writing, probably 70 percent, far exceeds the time spent on instruction in the logic necessary for critical thinking. The Baker courses differ from traditional composition courses in so far as they emphasize only one type of paper: the argumentative essay. In addition, grammar is taught only in the context of student writing assignments. For example, upon returning a set of essays a teacher might spend half a class period going over the points of grammar found wanting in the papers or (better yet) choose to meet with each student to explain the

problems. Students must return their papers with all mechanical errors corrected before their grades are recorded.

Assessing the Baker Freshmen Courses with the E-W

In the fall of 2005, we began the fifteenth year of the freshmen program. Our assessment data continues to demonstrate that our approach is as good as or better than many more traditional alternatives to the teaching of critical thinking or writing.⁹ With the endorsement of Stephen Norris, we began assessing the critical thinking element of the LA Program with *The Ennis-Weir Critical Thinking Essay Test* (E-W). Because the sequence integrates instruction in writing with logic and critical thinking, this test was deemed to be the most appropriate. It asks students to respond, in writing, to an eight-paragraph letter to the editor, stating whether the reasoning in each paragraph is good or bad and supporting their judgments with reasons (see Johnson, this volume, for a more detailed description of the E-W). The pre-test is given to all freshmen the first week of the fall semester. We tell them that we are part of a large research project and to do their very best. The post-test is given as part of the final exam the last week of the spring semester and counts for about 3 percent of the student's total grade.¹⁰ This encourages students to take the post-test seriously. The data below indicates the outcomes for the pre- and post-tests for the six years that we used E-W.

Our experience using the E-W as an assessment tool leaves little doubt that our approach to teaching critical thinking achieved significantly better outcomes than the two comparison groups. Anyone who claims that an approach to teaching critical thinking that integrates written composition cannot work is thus shown to be mistaken. The same can be said of anyone who thinks that the only way to teach critical thinking is by using the standard approaches found in most informal logic texts. A freshman gain of nearly a full standard deviation in critical thinking skills is an impressive gain, and much better than the gain in the comparison groups.¹¹

Year	Pre	St.D.	Post	St.D.	Mean Gain	Diff in St.D. (Effect Size)
90/91 (n=169)	6.3		12.4		+6.1	+1.11
91/92 (n=119)	9.4		12.2		+2.8	+0.51
92/93 (n=178)	6.8		12.6		+5.8	+1.05
93/94 (n=178)	8.1		14.1		+6.0	+1.09
94/95 (n=164)	7.5		13.0		+5.5	+1.00
95/96 (n=169)	6.9		12.9		+6.0	+1.09
Mean (n=977)	7.5	+/-5.3	12.8	+/-5.7	+5.3	+0.97

*St.D. used is 5.5, the average St.D. pre- and post-term

Comparison Groups Using the Ennis-Weir Test

	Pre	Post	Diff.	Mean Gain in St.D.
Standard Logic Class F94 (n=44)	11.2	9.5	-1.7	-0.31
Standard CT Class S92 (n=23)	12.1	13.7	+1.6	+0.29
Mean (n=67)	11.7	11.6	-0.10	-0.02

Comparison of BU Freshmen Scores to Senior Scores on Ennis-Weir

	Fr.	Sr.	Diff.	Mean Gain in St.D.
Grads 1995 (n=119)	9.4	14.6	+5.2	+0.94
Grads 1996 (n=88)	7.1	14.1	+7.0	+1.27
Grads 1997 (n=80)	6.8	14.8	+8.0	+1.45
Grads 1998 (n=58)	8.8	19.1	+10.3	+1.87
Grads 1999 (n=42)	7.3	17.4	+10.1	+1.84
Mean (n=387)	7.9	16.0	+8.1	+1.47

Table 1: Comparison of Ennis-Weir Critical Thinking Essay Test pre-and post-test scores for Baker freshmen, 1990-1996

One might argue that the comparison groups started out with higher pre-test scores, and so could not be expected to gain as much. There may be something to this argument but it hardly accounts for the standard logic classes getting worse. The critical thinking class did have a higher post-test score, but the effect-size gain of 0.29 is less than the literature claims is average (an effect-size gain of 0.5 standard deviation is considered average¹²).

Why did the freshmen in Baker's integrated, two-semester sequence do so much better on the E-W than the comparison groups who were taking the more traditional classes in logic and critical thinking? Educational research is notoriously uncertain and definitive answers would take more controlled experiments that carefully isolated as many variables as possible, e.g., teaching methods, textbooks, and teacher preparation. We have not been able to carry out an extensive program of research along these lines, but there are some obvious aspects of our freshmen sequence

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

Key characteristics of our classes are simplicity and the repeated application of the critical thinking skills we emphasize in our two-semester sequence. Almost everything covered in the sequence aims to develop skills for evaluating the arguments found in what students read and what they write. Such simplicity and repetition may make it easier for students to internalize the basic critical thinking skills and apply them successfully to the E-W. Beyond that, 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. In the two-semester sequence, we devote only the first six weeks to the study of the principles of critical thinking and logic. Most of what students cover early in the sequence is then applied repeatedly to what they read and in writing their papers. The logical tools are seen as something that have obvious and immediate use in students' educations — not as just a set of skills needed to pass a test and then to be forgotten.

In part because of our emphasis on repetition, the time our students spend using the skills we teach distinguishes our approach from that experienced by the comparison groups. Looked at from this point of view, it is not surprising that a two-semester sequence, in which relatively simple skills are repeatedly practised for twenty-three weeks, yields better outcomes than broader, traditional one-semester courses in critical thinking or logic. Our experience provides evidence of the value of an “across-the-curriculum” approach to critical thinking, in which all instructors ask students to evaluate positions by the standards of evidence and argument appropriate to their discipline. If the same song is sung often enough, most students learn it. When different teachers play the game by different rules, then students have, in contrast, a difficult time deciding what is important and what is peripheral, and are less able to evaluate the rationality of a position.¹³

Another reason our students may have taken critical thinking more seriously than those in the comparison groups is our emphasis on the value of a logical critique to most of the things

they read and write. If we are successful in this, then students will use the techniques we teach, not only in assignments for our courses, but in assignments for other courses, and in reading and writing other material every day. In such a context, it is plausible to suppose that they may be more inclined to learn the skills we teach.

During the time in which we used the E-W for assessment purposes, our research indicated that one-semester courses in critical thinking make a fairly small difference in students' abilities to think critically. In contrast, student performance is significantly enhanced by a two-semester sequence that teaches the logical tools needed for "the honest evaluation of alternative positions" and then requires that students apply this knowledge to expository writing.

Hopefully, other educators interested in assessing student critical thinking skills can learn from our experiment and share their assessment data with the wider educational community. Some may be reluctant to use the E-W because it is an essay test and time-consuming to grade, and because one might imagine that it would be difficult to achieve inter-grader reliability. But our experience shows that it is possible to achieve inter-grader reliability of 0.85 or better using well-trained student workers, and grading time can be reduced if researchers choose a random sample of the essays and grade only those, instead of grading all students' essays for assessment purposes. We learned the latter lesson too late to take advantage of it — after double-blind grading of 1,447 E-W essays (sometimes one is *unlucky*).

Assessing the Baker Approach with the CCTST

In the fall of 1996, we began to do pre- and post-testing with the ***California Critical Thinking Skills Test*** Form A (the CCTST). One reason for the change was concern about the growing post-test gains of our seniors. By 1999, the effect-size gain by the graduating seniors was 1.84, and that seemed unreasonably high. We hypothesized that the material on the test must be public, and

the seniors were using it to study for the test. The data for the eight years during which we used the CCTST follows.

Freshmen	Pre	St.D.	Post	St.D.	Diff.	Mean Gain in St.D.
F96/S97 (n=152)	15.14	+/-4.46	18.49	+/-4.30	+3.35	0.75
F97/S98 (n=192)	14.50	+/-3.84	17.17	+/-4.40	+2.67	0.60
F98/S99 (n=171)	15.81	+/-4.60	17.90	+/-4.72	+2.09	0.46
F99/S00 (n=153)	15.91	+/-4.20	18.28	+/-4.30	+2.50	0.53
F00/S01 (n=184)	16.00	+/-4.20	18.52	+/-4.23	+2.37	0.51
F01/S02 (n=198)	15.30	+/-4.11	17.47	+/-4.44	+2.17	0.48
F02/S03 (n=221)	15.60	+/-4.1	18.2	+/-4.40	+2.60	0.57
F03/S04 (n=169)	15.40	+/-4.1	18.1	+/-4.60	+2.70	0.60
Mean (n=1447)	15.10	+/-4.2	18.0	+/-4.30	+2.60	0.56

Comparison Group ¹⁴	Pre	St.D.	Post	St.D.	Diff.	Mean Gain in St.D
1990 Test Validation Study (n=262)	15.94	+/-4.50	17.38	+/-4.7	+1.44	0.32
2000 University of Melbourne (n=50)	19.50	+/-4.74	23.46	+/-4.36	+3.96	0.88
2001 McMaster University (n=278)	17.03	+/-4.45	19.22	+/-4.92	+2.19	0.49
2001 Monash University (n=174)	19.07	+/-4.72	20.35	+/-5.05	+1.28	0.28
2002 University of Melbourne (n=117)	18.85	+/-4.54	22.10	+/-4.66	+3.35	0.73
Mean (n=831)	18.08	+/-4.59	20.50	+/-4.73	+2.42	0.54

*The standard deviation used is always 4.52.

That was the standard deviation used when the test was validated.

Comparison of Freshmen Scores to Senior Scores on the CCTST: Fall 2000-Spring 2004

	Seniors	Freshmen	Seniors	Diff.	Mean Gain in St.D.
Grads 2000 (n=102)		15.2	19.4	+4.2	0.93
Grads 2001 (n=79)		14.3	18.3	+4.0	0.88
Grads 2002 (n=86)		15.8	19.2	+3.4	0.75
Grads 2003 (n=65)		15.8	19.7	+3.9	0.87
Grads 2004 (n=88)		15.9	20.2	+4.3	0.95
Mean (n=396)		15.6	19.3	+4.0	0.88

Table 2: Freshmen pre- and post-test scores using the California Critical Thinking Skills Test, fall 1996 to spring 2004

The CCTST is a professionally normed test. It is used to assess critical thinking course outcomes and gives users a clearer sense of what student scores mean relative to other schools’ performances than that provided by the E-W. With the average gain of +2.6 points or 0.56 of a standard deviation for the freshmen year, we did better than the mean gain of 2.42 points, or 0.54 of a standard deviation, for the comparison groups. Again, it is generally understood that any effect-size gain over 0.50 of a standard deviation for one course is a strong performance, even though the gains were much smaller than those on the E-W. Most heartening are our mean scores (0.56), which were always higher than the mean of the test validation study (0.32). The McMaster and University of Melbourne courses both employed computer-assisted exercises, something our timeline for teaching the basic

logic and critical thinking material prohibits. Because the justification for the freshmen sequence is to prepare students to write strong critical papers, we spend minimal time on textbook logic and critical thinking exercises.

The average gain on the CCTST from the freshmen to senior year has been +4.0 points, +1.4 points better than the +2.6 point average gain during the freshman year. This is a reasonable gain on a very challenging test with only 34 points. Studies show that students' critical thinking skills usually do not increase over 0.55 of a standard deviation over three years of college.

Obviously, the pre-test scores for McMaster University (17.03) and the University of Melbourne (18.85) were much higher than the Baker scores. This may be a function of three things: first, the students at those schools were taking the critical thinking courses as an elective or a course serving a major; if so, they may have been better equipped or more inclined to do well in such a course. Second, they were older than the entering freshmen at Baker with more college courses completed, and one might assume that experience with college-level course work would in itself enhance critical thinking skills (although I have no way of knowing whether this is so). Third, unlike the students at McMaster University and the University of Melbourne, those in the Baker program were not allowed to drop the course, which may have meant that weaker students stayed in the courses and lowered the post-test mean.

Some Thoughts about the E-W and the CCTST

What can we say about the different outcomes from the two tests we used to measure the effectiveness of our courses and our program? The differences in mean gain in standard deviation between the E-W and the CCTST are obvious.

Freshmen	Pre	St.D.	Post	St.D.	Diff.	Mean Gain in St.D.
E-W Mean (n=977)	7.5	+/-5.3	12.8	+/-5.7	+5.3	+0.97
CCTST Mean (n=1447)	15.1	+/-4.2	18.0	+/-4.3	+2.6	+0.56
BU Freshmen to Seniors						
E-W Mean (n=387)	7.9		16.0		+8.1	+1.47
CCTST Mean (n=396)	15.6		19.3		+4.0	+0.88

Table 3: Comparing the E-W and the CCTST

The effect-size gains on the E-W are nearly double those on the CCTST, even though all students who took each test have gone through the same program, using the same text, doing the same assignments. This raises an obvious question: Which one is more accurately measuring students’ abilities as critical thinkers? The answer to this question may depend on how one conceptualizes critical thinking. If we think that a fairly deep understanding of deductive logic and the ability to test scientific hypotheses are both essential skills of any student who claims to be a critical thinker, then I would say that the CCTST is a more accurate measure. This is because numerous questions on the test require that students have a clear understanding of deductive validity (and much of what that concept entails) or how to test for the acceptability of a hypothesis or to falsify one. I cannot imagine students doing very well on the CCTST without a clear understanding of both deductive and inductive logic.

But many informal approaches to critical thinking adopt a conception which does not emphasize formal logic. If one adopts this kind of conception, then the E-W might be a better tool for assessing student progress in critical thinking. In deciding which instrument to favour, it is important to remember that the ultimate purpose of assessment is not only to measure students’ performance against that of others or some pre-established norm,

but also to see how well students are achieving the educational goals of specific programs, or reaching course objectives.

Beyond the differences in the scope of the two tests (differences one can see more clearly after reading Groarke and Johnson in this volume), one could argue that the act of taking E-W more closely resembles what we want our students as critical thinkers to do in real life: read extended arguments, evaluate their merit, and then articulate them in writing with a cogent critique. The act of taking the E-W is a more natural experience for students than meticulously working through the thirty-four questions on the CCTST, some of which are highly artificial (e.g., the question that asks one to “Consider the `krendalog’ relationship”). Yet the CCTST has the sort of questions, as Hitchcock (2003) and van Gelder, Cumming, and Bisset (2004) have shown, that complement computer-assisted exercises, exercises that can significantly enhance students’ performances. Students can prepare by practising discrete logical skills that can be applied to the CCTST. Yet, because of its resemblance to real-life situations that call for critical thinking, one might argue that the E-W is in fact a better gauge of a student’s ability to think critically in real situations.

Conclusion

No matter which test more accurately measures students’ critical thinking abilities, it is important that more teachers of critical thinking choose a standardized test that has been professionally normed or used so widely that norms are available. Only when a large number of teachers do pre- and post-testing in their courses will it be possible to determine systematically which approaches to teaching critical thinking work and which do not.

Many teachers prefer to use personalized “in-house” assessment tests or portfolios, but they are problematic. To the extent that teachers rely on instruments of this sort, they will not be able to determine how their students are doing relative to other students in similar situations in other institutions. Research reports that use

such individualized, and hence unfamiliar, tests and approaches cannot tell the wider circles of academe what works and what should be avoided. Creating one's own assessment test or grading portfolios is time consuming, in any case, and there is no way to know, without a lot of professional help, whether the test or portfolio approach is valid.

The data on two standardized tests collected by Hitchcock, van Gelder, and me allows us to establish what sort of an effect-size gain can be expected from a one-semester critical thinking course, or a two-semester sequence that combines critical thinking and composition. If the results are better than those reported in the current research, this is good news that should be shared with all. If the results are lower than the norm, this is a useful sign that one should begin to address deficiencies in an attempt to achieve better student outcomes. That is what assessment is really all about: improving student learning by finding out in a systematic way what students know or can do at the end of a course or program and responding conscientiously to the outcomes.

In my case, a project begun in 1988 that grew out of my experience as a student simultaneously enrolled in a course in logic and a seminar in Plato produced a unique approach to teaching critical thinking and writing, and probably provides the largest pool of assessment data available using two well-known standardized critical thinking tests. I hope that our approach at Baker to teaching critical thinking and our ongoing attempts to assess it will be of use to others faced with the challenges of teaching and assessing critical thinking.

References

- Cederblom, J., and D. Paulsen. 2001. *Critical reasoning*, 5th ed. Belmont, CA: Wadsworth Publishing.
- Groarke, L. 1999. Deductivism within pragma-dialectics. *Argumentation* 13: 1-16.
- Hatcher, D. 2000. Arguments for another definition of critical thinking. *Inquiry: Critical Thinking Across the Disciplines* 20(1): 3-8.
- _____. 1999a. Why formal logic is essential for critical thinking. *Informal Logic* 19(1): 77-89.
- _____. 1999b. Why we should combine critical thinking and written instruction. *Informal Logic* 19(2 and 3): 171-83.
- Hatcher, D., and L. Spencer. 2006. *From critical thinking to composition*, 3d ed. Boston, MA: American Press.
- _____. 2000. *Reasoning and writing: An introduction to critical thinking*, 2d ed. Boston, MA: American Press.
- Hitchcock, D. 2003. The effectiveness of computer-assisted instruction in critical thinking. In *Informal logic at 25: Proceedings of the Windsor conference*, ed. J. Blair, D. Farr, et al. Windsor, ON: OSSA.
- Norman, G., J. Sloan, and K. Wyrwich. 2003. Interpretation of changes in health-related quality of life: The remarkable universality of half a standard deviation. *Medical Care* 41: 582-92.
- Nosich, G. 1982. *Reasons and arguments*. Belmont, CA: Wadsworth.
- Pascaralla, E., and P. Terenzini. 2004. *How college affects students revisited: Research from the 90s*. San Francisco, CA: Jossey-Bass.
- van Gelder, T. 2001. How to improve critical thinking using educational technology. In *Meeting at the crossroads: Proceedings of the 18th annual conference of the Australasian Society for Computers in Learning in Tertiary Education*, ed. G. Kennedy, M. Keppell, C. McNaught, and T. Petrovic, 539-48.

Melbourne: Biomedical Multimedia Unit, University of Melbourne.

van Gelder, T., G. Cumming, and M. Bissett. 2004. Cultivating expertise in informal reasoning. *Canadian Journal of Experimental Psychology* 58(2): 142-52.

Notes

1. After the original FIPSE grants of \$168,000 to plan and set up the freshmen sequence, the Hall Family Foundation has supplemented the program with grants over \$850,000 since 1991.
2. We also assessed the writing outcomes using the Test of Standard Written English, and found that our students did better than students taking courses using more standard approaches to written composition.
3. Of course, it was also a good excuse to try out my theory about the relationship between knowledge of formal logic and good prose.
4. For a defence of the conception we finally agreed to use, see Hatcher (2000).
5. I would be remiss not to give credit to Connie Missimer for her influence on this conception of critical thinking. Connie convinced me years ago that critical thinking, like good scientific investigation, should always include the weighing of alternatives, whether theories, explanations, accounts, courses of action, or policies. Note also, that although we distinguished critical from creative thinking in Baker program, the part of our definition which includes getting clear on and honestly evaluating alternatives does not conflict with much of what is said in Part Two of this volume about the nature of creative thinking. If one is to evaluate alternatives, one must first "imagine" them.
6. Of course, in a complex argument, the reasons A, B, C, and D might themselves have reasons to support them.
7. While the use of Deductive Reconstruction dates back to my college days mentioned at the beginning of the chapter, this approach to critical thinking is also present in Nosich (1982) and Cederblom and Paulson (2001). For a defence of Deductive Reconstruction, see Groarke (1999).
8. By expository paper, I mean any paper where the student must have a

thesis and support it with evidence and arguments. The techniques we teach would be of little use to students whose writing assignments do not involve such a task, e.g., creative writing, reports, surveys of the literature, or accounts of historical events.

9. For a more complete description of the program, see Hatcher (1999a, 1999b).
10. Perhaps a better strategy to insure that students take both the pre-and post-test seriously is to tell them at the pre-test that some students do worse on the post-test, albeit not many, and the score used for points on the final exam will be the higher of the pre- or post-test.
11. Pascarella and Terenzini (2004); the three-year estimate for CT gain was +0.55 mean standard deviation.
12. In addition to the work of Pascarella and Terenzini, Norman, Sloan, and Wyrwich (2003) come to the same conclusion.
13. The approach I take with respect to teaching critical thinking is quite similar to that of Nosich (in this volume). That is, we share emphasis on [pb_glossary id="447"]reasoning assessment[/pb_glossary]. Comparing the two approaches and definitions of critical thinking might, then, be a worthwhile exercise. Similarities between the two approaches are not surprising, since previously Nosich (1982), like our program, has taken a Deductive Reconstruction approach to critical thinking.
14. Both the McMaster University and University of Melbourne courses used computer-assisted instruction to supplement in-class work. I think their positive gains indicate that computer exercises have a positive role to play in enhancing critical thinking test scores. It would be interesting to see their gains if they used the E-W. For a full account and analysis of the McMaster course see Hitchcock (2003). See "How to Improve Critical Thinking Using Educational Technology" by Tim van Gelder (2001).