

## Enhancing Critical Thinking Ability Through Academic Debate

KENT R. COLBERT

In a recent issue of the *CEDA Yearbook*, Hill (1993) reviewed debate-critical thinking research. He suggested data are insufficient to demonstrate a debate-critical thinking relationship, concluding the literature does not "demonstrate that participating in competitive debate promotes the development of critical thinking ability to any significant degree" (p. 18). Certainly, additional research to consummate sound educational objectives and critical thinking is well advised. The research exploring the debate-critical thinking relationship is not complete. It does, however, establish a relationship. Like other social sciences, debate-critical thinking research operates on the basis of probability, not a deterministic model asserting causation. Hill's (1993) claim suggested forensic educators are "ill-prepared to use development of critical thinking ability to meet educational accountability demands" (p. 3) may be attributed to misunderstanding social science research, specific debate-critical thinking research, and the nature of psycho-metric instruments.

Underlying Hill's (1993) assessment is what Johnson (1943) termed a "hypercritical attitude" or the "tendency to reject all conclusions based upon probabilities; an unwillingness to make a tentative choice; a tendency to demand 'all' the evidence . . . to the point of intellectual fence-sitting" (p. 86). This hypercritical attitude driven by a crisis mentality is reflected by some members of the forensics community who are troubled by (what they see) as a groundswell of unfounded panic about the basic worth of intercollegiate debate. Unfortunately, the resulting economic-survival rationale attempts to influence the direction of future research. Contrary to this crisis mentality, debate-critical thinking research establishes presumption favoring positive results. Objective analyses of the defensible studies indicate academic debating consistently enhances participant critical thinking abilities. And under certain conditions and instructional approaches, debating can significantly increase critical thinking abilities.

This paper will: 1) classify and review debate-critical thinking research; 2) evaluate the Watson-Glaser Critical Thinking Appraisal (WGCTA) with respect to Hill's (1993) recommendations of clarification and diversification; and finally 3) contend future debate-critical thinking researchers should concentrate their efforts using the principles of-replication and cooperation.

It is important to initially examine the complexity of the critical thinking process before rejecting previous attempts to conceptualize and measure it. Critical thinking remains a difficult and challenging construct to define, operationalize, measure, learn, and teach. As Meyers (1986) explained, "The process of modifying old, or creating new, mental structures is often uncomfortable and at times even painful. . . . Teaching critical thinking involves intentionally creating an atmosphere of disequilibrium, so that students can change, rework, or reconstruct

## Colbert: Debate-Critical Thinking

their thinking processes" (p. 13-4). The interactive experience of the debating process appears to present a catalyst creating the disequilibrium, motivation, and framework needed to facilitate the acquisition of critical thinking abilities.

The ability to think critically has been an imperative of scholars for many years. From Plato's rationale analysis, and Aristotle's empiricism, to Guilford's (1967) intellectual operations of information, and Piaget's (1972) writings on primary structures, educators have long deliberated about critical thinking. The origin of measuring critical thinking can be traced to Dewey's (1910) formation of "reflective thinking" that inspired a permutation of scientific inquiry concerning the thought process. The need to move beyond elocution was subsequently articulated by Johnson (1942). She stated, "Experimental studies in discussion, debate, and persuasion, though numerous and valuable, have been confined largely to rhetorical considerations . . ." (p. 84). The early debate-critical thinking researchers acknowledged Johnson's call by attempting to *measure* the effects of debating, an activity sharing many similar elements associated with the process of critical thought.

The major studies of a debate-critical thinking relationship (Howell, 1943; Brembeck, 1949; Williams, 1951; Beckman, 1954; Jackson, 1961; Cross, 1971; Colbert, 1987) have provided valuable pedagogical data for researchers looking beyond the acceptance or rejection of a null hypothesis. However, other studies display several fatal methodological flaws making their conclusions ill-suited for earnest comparisons. (Luck & Gruner, 1970; Huseman, Ware, and Gruner, 1971; Gruner, Huseman & Luck, 1971; Follert & Colbert, 1983; and Whalen, 1991). A superior explanation of debate-critical thinking research is possible by critically examining the specific experimental treatments, rather than simply reporting their results in chronological order. The three categories that will be used to explain the debate-critical thinking relationship are: 1) Classroom Training Studies; 2) Competitive Experience Studies; and 3) Incidental Studies.

### CLASSROOM TRAINING STUDIES

In 1947, Winston L. Brembeck attempted to determine if a semester's course in argumentation improves critical thinking. He had these conclusions:

- 1) The argumentation students as a whole out gained control students in critical thinking scores over a one semester experimental period. The critical ratio of the difference in battery mean gains between the two groups was found to be 2.56. There is approximately only one time in a hundred that this difference could occur by chance. Therefore, it may be concluded that, on the whole, the argumentation courses studied in this experiment improved critical thinking ability. . . . 2) In all but one of the schools the experimental students had an average test score higher than the control students. This may mean that the students taking argumentation courses were more capable.

Even though this pretest advantage served to narrow the range for improvement, the experimental groups still out gained the controls. . . . 3) The participating schools differed widely in terms of changes in critical thinking ability. . . . 4) Experimental students with one or more years of debate training (high school and/or college) made critical thinking pretest scores which are significantly higher than scores made by those with no previous debate training.

The second major study of classroom debate training and critical thinking was conducted by Vernon E. Beckman during the 1954-5 school year at five colleges or universities. He concluded:

- 1) It cannot be concluded from the present study that there is a significant difference in critical thinking ability, as measured by the WGCTA, between students in college argumentation courses, discussion courses, and other courses of the type used as controls in this investigation. Analysis of variance of gain scores and of covariance of adjusted post test scores showed an  $F$  ratio that was below the five per cent level of significance. . . . 2) The conclusion can be drawn that there are statistically significant differences in gains between schools. . . . 3) Students who score high on the pretest make relatively smaller gains than those with lower pretest scores. . . .

Both studies (Howell, 1947 and Beckman 1954) comparing traditional classroom debate training reported an overall increase in critical thinking scores. Beckman's critical thinking differences between the debate and control groups were not significant, although the direction favored the experimental group. Does this mean, as Hill (1993) speculated, that one study cancels the other? Given the methodologies were not as closely replicated as Hill (1993) presumed, cross comparisons are severely limited. First, the specific training methods were not reported or manipulated by either researcher. Both studies reported significantly different critical thinking scores between schools, Beckman's (1954) lack of significance likely reflects differences in instruction and course design. Course content, instructors, teaching methods, curriculum requirements, and other factors suggest argumentation courses (the experimental treatment) were not equivalent.

Second, despite the positive direction favoring the potential of classroom training to enhance critical thinking, it is unlikely any single course can completely isolate an effect on students also exposed to a variety of courses and experiences related to critical thinking. Meyers (1983) explained, "teachers must be realistic about what can be accomplished in the way of critical thinking development in a typical ten-week college course. Most students' previous thinking processes are not going to be radically altered in this length of time" (p. 23). Researchers in several disciplines agree. Frank's (1969) research revealed only certain speech communication courses improve critical thinking. (p. 296-302) Similar claims are reported in other fields. Henkel (1968), Yoesting and Renner (1969), Seymour and Sutman (1973), George (1968), Brakken (1965), and Ness (1967) all found a particular course topic is not a sufficient

guarantee of critical thinking improvement. Annis and Annis (1979) concluded, "This study provides some initial evidence regarding the impact of philosophy on critical thinking. Future research should . . . [determine] the effects of specific kind's of course goals, content, organization, and teaching strategies have on critical thinking" (p. 225). Debate-critical thinking researchers should also pursue investigations documenting and manipulating the precise experimental treatment (course), especially since different programs consistently produce statistically different results.

#### COMPETITIVE DEBATING STUDIES

The second category of debate-critical thinking research encompasses competitive tournament debating. William S. Howell, who's 1943 dissertation founded this specific locus of research concluded:

- 1) Considering the entire experimental and control groups, the debaters out gained non-debaters in critical thinking scores over the experimental period of six months. The critical ratio of the difference in mean gains is 1.04. Since a minimum critical ratio of 2 is required for significance, we cannot concluded that high school debaters are certain to out gain nondebaters. There are 85 chances in 100 that this difference is real. . . . 2) When the experimental and control groups are equated on I. Q. scores the debaters again out gain the non-debaters, but not significantly. . . . 3) Both debaters and non-debaters show significant gains in critical thinking scores over one debate season of approximately six months. . . . 4) Even though the debaters and non-debaters are closely matched on I. Q. scores, the debaters show significant superiority on both pretest and post test of critical thinking. . . . 5) The high school debater's advantage in scores on the WG tests carries over to the college level. The evidence included on this point is not conclusive, but it indicates that college students with high school debate experience score consistently higher on these critical thinking tests than do those who have not debated. . . . 6) Great differences in mean gains of debaters over non-debaters were found among participating schools. . . .

The second major study is a master's thesis of Donald E. Williams at Purdue University during the 1950-51 debate season and involved competing college debaters. Williams concluded:

- 1) Debaters did make significant gains in critical thinking, as measured by the WGCTA, but that their gains were not significantly greater than the gains of a similar group of non-debaters. . . . 2) There was a slight suggestion that those college students who have previous experience in debating in high school and in college may have greater critical thinking ability as measured by this test than those who had one year or less of such experience. . . . 3) Those students who have more than one year of experience in debating did not have greater critical thinking ability as measured by this

test than those students who had one year or less of such experience. . . . 4) There was some indication that those students who are rated as better debaters by their coaches will have greater critical thinking ability than those who are rated as having less skill in debating.

In 1961, Jackson, under the supervision of Brembeck, tested the debate-critical thinking relationship observing 100 debaters and 147 non-debaters from nine colleges and universities. Jackson reported:

The difference between gains was statistically significant between the .05 and .01 level of confidence. An analysis of the variance yielded an  $F$  ratio of 8.20. . . . 2) There were considerable differences among schools in the sample. Pretest total sample means ranged from 66.82 to 79.29 (99 maximums). Debater pretests ranged from 70.64 to 81.50. All of the schools, with one exception, had debater pretest mean advantage over the non-debaters. Post test means for the debaters at all schools were higher. . . . 3) Previous debating experience gave the students in this study a definite advantage. Those with previous debating experience, either in high school and/or college, had a mean pretest advantage of 5.43. A  $X^2$  of 13.86 was significant at the .001 level. 4) There was no significant relationship between success at winning debate contests and gain in critical thinking ability. Those who won over 80 per cent of their debates have a slightly higher gain, but it was not statistically significant. 5) There was significant relationship between the amount of participation during the experimental period and gain in critical thinking ability. Those who participated in over 20 debates registered a slightly higher mean, but it was not statistically significant.

In 1971, Cross studied 136 students at nine high schools to investigate the attainment of a "specific educational goal" believed to be related to competitive debating. An analysis of covariance (ANCOVA) produced significant  $F$  values using the WGCTA. He concluded:

1) Those who are drawn to competitive debate, low and high participants, and continue for one academic year have greater thinking facilities than those who are not attracted to debate. . . . 2) High participation in competitive debate accelerates debaters' capacity in critical thinking while low participation may not enhance critical thinking beyond the normal improvement in an academic year. . . . 3) Debaters with abstract belief systems significantly out gained debaters with concrete belief systems in critical thinking. . . . 4) Abstract subjects with higher belief systems significantly out gained others in critical thinking while concrete subjects with debater experience only out gained nondebaters. . . .

The most recent major study of the competitive debate-critical thinking relationship was reported by Colbert (1987) who investigated 275 CEDA and NDT debaters and nondebaters at nine universities for one complete academic year. Unlike previous research, his study design and use of the pretest as a covariant better controlled for extraneous variables, institutional

differences, the instrument's ceiling effect, and self-selection. He reported a sample size estimate calculated with a moderate (.5 sd.) to large (.8 sd.) effect size illustrating the magnitude of the differences found. The data represented several geographical regions, small and large institutions, and private and publicly funded universities. He reported that:

1) CEDA and NDT debaters scored 61.18 on the WGCTA pretest, 64.53 on the posttest, and a mean difference between the pre and post test of 3.35. Nondebaters had a pretest score of 52.67 on the WGCTA, 49.14 on the posttest, and a negative mean difference of -3.53. CEDA, NDT, and high school debaters outscored nondebaters on the WGCTA on the pretest, post test, and the gain from pretest to post test. An analysis of covariance was made to test for significant differences between debaters and nondebaters. . . . 2) CEDA debaters scored 62.67 on the pretest of the WGCTA, 63.53 on the post test, and scored a mean difference between the pretest and post tests were .87. NDT debaters scored 63.49 on the pretest, 67.91 on the post test, and scored a mean difference between the pretest and post test of 4.42. ANCOVA produced an  $F$  value of 21.88 significant at the .001 level. . . . 3) CEDA debaters scored 62.67 on the pretest, 63.53 on the post test, and a mean difference between pretest and post test of .87. The control group scored 52.67 on the pretest, 49.14 on the post test, and a mean difference of -3.53. CEDA debaters out scored nondebaters by a mean of 10 on the pretest, 14.39 on the post test, and scored a mean difference of 4.39. ANCOVA produced an  $F$  value of 5.368 significant at the .005 levels. . . . 4) NDT debaters scored 63.49 on the pretest on the WGCTA 67.91 on the post test, and scored a mean increase of 4.42. The control group scored 52.67 on the pretest, 49.14 on the post test, and a negative mean difference of -3.53. NDT debaters outscored the controls by a mean of 10.82 on the pretest and 18.77 on the post test for a mean difference from the pretest to the post test of 7.95. ANCOVA found differences between NDT debaters and nondebaters produced an  $F$  value of 31.77 significant at the .001 level.

While the competitive debate-critical thinking studies had similarities, variations in the version and modification of the WGCTA, sampling, population differences, length of experimental treatment, and statistical procedures were substantially different. These differences likely explain why some debate-critical thinking studies report significant findings, while others do not. For example, the WGCTA was administered in each study, however, it was revised by Watson and Glaser (1980) several times during this period. Howell (1947) independently modified the WGCTA in his study. Generalizations equating Howell's self-modified WGCTA with studies using the original and the subsequently refined WGCTA instrument make many comparisons inappropriate.

Sampling procedures also differed across studies. Howell (1943) randomly assigned students to debate and control groups. Williams (1951) and Jackson (1961) stratified their

control groups. They matched a control group with the experimental group on differing dimensions including age, sex, educational background, and educational performance. In addition, Jackson (1961) stratified for college major and matched the groups I. Q.'s. Cross (1971) randomly selected the members of his control group. Colbert (1987) matched nondebating controls at each experimental group's institution. Some of these studies routinely engaged speech courses as control groups and research indicates speech courses also develop critical thinking abilities. (Smith, 1942; Ness, 1967; and Frank, 1969) Claims that debating produces small positive or regressive effects on critical thinking (Whalen, 1991; Hill, 1993) are based on studies using no control groups, ones receiving similar experience (treatment) as the debaters, and small unrepresentative samples.

The population samples for each of the competitive debate studies were dissimilar. There were differences between college and high school students, instructors, and quality of instruction, and competition level. Teachers used the instructional methods they deemed appropriate. Thus, no attempt was made to control or manipulate the instructional method, content, or conduct of instruction because the researchers were more interested in testing the way debating was actually being taught. Obtaining institutional and individual support to participate in time consuming research, the inability to administer an appraisal instrument at tournaments (40 minutes to an hour and a half), and the need for long term training and experience makes it unusually impractical for individual researchers to conduct this type of research alone.

The length of exposure to the experimental treatment also varied in the debate-critical thinking research. Howell's (1943) experiment was conducted over six months. Williams (1951) observed one semester of debate competition. Jackson's (1961) experimental period lasted six to seven months. Cross (1971) waited for one school year to pass, as did Colbert (1987). It appears one full year of competitive debate experience consistently produces significantly higher critical thinking scores. Perhaps a period of "Reflective Thinking" (Dewey, 1910) is needed before measurable difference's surface. Findings that students with prior debate experience consistently out gained controls also suggest a minimum threshold of training and experience may be required over time for researchers to detect the debate-critical thinking relationship. Consequently, more experience compressed into shorter durations of time may not simulate exposure to debating over longer periods of time. Attending three tournaments a semester in four years, for example, is not comparable to debating in twenty-four tournaments in one year. The time needed to learn, think, discover, reflect, mature, interact with others, and consider several different topics may require more than a semester or two of debating.

The different statistical procedures each researcher chose should also be considered. In analyzing the results of the research, critical ratios (Howell, 1943), t-test (Jackson, 1961), ANCOVA and Scheffe's post comparison statistics (Cross, 1971), and ANCOVA (Colbert, 1987) were used. More robust statistical procedures could account for different results. Given

the nature of the instrument and the need to control for extraneous variances, the ANCOVA appears most appropriate. Studies using ANCOVA in a pretest post test design consistently rejected the null and supported the alternative hypothesis that debate training significantly promotes critical thinking scores.

Another issue involves the purported "attraction" or "self-selection" effect. Basically, the theory suggests those with greater critical thinking abilities seek out debate activities. Howell (1948), Williams (1951), and Cross (1971) found debaters with previous experience had higher critical thinking scores on their pretest. Jackson (1961) and Colbert (1987) reported the same relationship among college students. Why would critical thinkers be attracted to debating? It is likely debating stimulates critical thinking. If competitive debating was not a critical thinking activity, it is unlikely critical thinkers would elect or continue to participate. Those with prior debate experience indeed test higher on the pretest, but research also shows experienced debaters can and often do improve during the experimental period. If individuals with greater critical thinking abilities are attracted, they could benefit more from debate than novice participants. (Howell, 1948) Additionally, there is no reason to believe critical thinkers would gain, develop, or maintain critical thinking abilities without activities like debating that exercise and hone them. The key issue is not whether debating causes critical thinking, but to discover how debating can cultivate it for those with varying levels of critical thinking ability.

Of the five major competitive debate-critical thinking studies, only Howell (1948) and Williams (1951) failed to reject the null hypothesis. At least three explanations are available for studies without significant differences between experimental and comparison groups. First, Howell (1948) modified an early version of the WGCTA to suit his research interest. Second, Howell (1948) randomly assigned students to debate and non-debate conditions. While randomization reduces the risk of motivational side effects, it also increases the probability that reactance effects will undermine the research efforts. Students forced to engage in the activity may not be receptive to the stimulus being offered. Similarly, coaches with unwilling subjects may also exhibit reactance effects. Reactance effects could diminish the benefits derived from the debating. Thus, studies randomly assigning students to debate groups were testing debating in a way that doesn't accurately reflect reality. Third, those with previous debate students had higher critical thinking scores to begin with. The ceiling effect identified by Crites (1965) could have compressed the gains made by the debaters--diminishing the measured effects. If the students begin high on the scale, they will have less room for improvement. This factor could explain why debaters consistently gain in critical thinking, but not always significantly more than control groups. One study (Lucas, 1972) suggested individuals with low scores systematically improve more than high scoring students. (p. 381-7) Despite limited range of improvement, debater WGCTA scores were consistently in the predicted (positive) direction.

Another important finding was the significant differences observed between schools participating in the debate-critical thinking studies. Howell (1948) and Jackson (1961) both

reported significant (WGCTA) differences between debate groups attending different institutions. Their findings implied instructional techniques, methods, and/or content probably influenced the acquisition of critical thinking skills. If some teaching methods fail to produce significant results, we should not presume it cancels out methods that are successful. Williams (1951) thesis, for example, only studied one institution despite the knowledge, most previous research suggested significant differences between schools existed. The inability of this study to observe significant results is likely from observing one atypical competitive debate program.

The preponderance of defensible evidence suggests competitive debate experience can indeed improve critical thinking skills. The lack of significance in some studies is reflected in design limitations, instrument ceiling, sampling, teaching methods, or statistical procedures. Rationalizing that one insignificant study cancels a significant one is parsimonious. There are no compelling reasons why competitive debating, when properly taught, is unlikely to improve critical thinking skills. Some convincingly argue that debating practiced correctly is a method of critical thinking. (Perella, 1988) All of the major debate critical-thinking researchers generally concluded debating improved critical thinking scores.

#### INCIDENTAL STUDIES

Three studies Luck & Gruner (1970); Gruner, Huseman, and Luck (1971); and Huseman, Ware, and Gruner (1972) measured high school debaters at a summer workshop, administering the WGCTA (Form YM), and the California *E* scale (authoritarianism). Luck and Gruner (1970) made no claim of a debate critical thinking relationship, only that critical thinking and authoritarianism have a "negative  $r$  of  $-.322$  ( $p < .01$ )" (p. 380). While Gruner, Huseman, and Luck (1971) concluded, "debate ability has been shown once again to be related to critical thinking ability," they did not predict a positive increase in critical thinking as a result of debating. (p. 65) Their data suggested a negative relationship between authoritarianism and critical thinking in the best rated debaters. Presumably, better debaters have higher critical thinking scores, and are less authoritative. Huseman, Ware, and Gruner (1972) concluded, debate "director[s] can best improve their charges' debating performance by attempting to develop in them the abilities measured by the tests in this study [WGCTA]" (p. 265). All three studies (Luck & Gruner, 1970; Gruner, Huseman, and Luck, 1971; and Huseman, Ware, and Gruner, 1972) suffer serious methodological flaws making their conclusions little more than speculation. First, the studies were one-shot design measures. Second, no control or comparison groups were used to isolate the experimental treatment from a host of extraneous variables. (Campbell & Stanley, 1963) Third, the researchers admittedly used "crude" methods of having "coach-critic-judges" rate debaters into "debate ability quartiles" without measuring the reliability or validity of their methods. (Holsti, 1969) Although Luck and Gruner (1970); Gruner, Huseman, and Luck (1971); and Huseman, Ware, and Gruner (1971) provided

interesting questions about debating and authoritarianism, data are insufficient to support or deny a debate-critical thinking relationship.

A meta-analytic study by Follert and Colbert (1983) attempted to analyze debate-critical thinking research collectively. Admittedly, their study suffered methodological shortcomings. The methodological differences of previous studies severely limited statistical generalizations. The "shaking of the educational foundation" did not suggest the foundation does not exist. It sounded a call for additional research. They stated, "since there are some problems with score stabilities and [WGCTA] form compatibility, the more powerful meta-analytic techniques were rejected" (p. 8). The authors further cautioned that using dichotomous binomial distribution estimates would sacrifice part of the variance. These and other limitations prevented using Follert and Colbert's (1983) results to reject the findings of Howell (1948), Jackson (1961), Williams (1951) and Cross (1971). The heuristic value of Follert and Colbert (1983) was to provide future scholars an alternative procedure to examine debate-critical thinking research and stimulate interest for additional study of contemporary competitive debate practices. What is most important, the authors cautioned others against making cross comparisons and generalizations using various debate-critical thinking studies.

More recently, Whalen (1991) attempted to study the self-selection supposition, the amount, and type of debate experience and its effect on critical thinking. The focus on isolating debate and specific treatments raised interesting questions about debate-critical thinking research. However, several fatal flaws make any conclusion from the conference paper untenable. First, he sampled one institution ignoring previous research reporting significant difference in critical thinking scores between schools. Second, the sample size was inadequate. The author reported the "average subject size for these tests was approximately 12 per group." Third, the author did not provide a sample size estimate or rationale to justify an extremely small  $N$ . Fourth, only one form of the WGCTA was employed in a pretest post test design, clearly against the recommendations of Watson-Glaser (1980). Fifth, many speech courses also contain the critical thinking components taught in debate, so there is no way to determine if the comparison (speech) group was receiving similar stimuli as the control group. Finally, the author claimed to control for self-selection, but never reported if students registering for the debate courses were aware of the course requirements before doing so. Whether the course was a curriculum requirement, or how systematically excluding those with previous debate experience produced valid conclusions about those being "attracted was not addressed." Concluding the debate course group without competitive experience out gained the debate course group with a debating requirement was significant at the .10 alpha levels. This level of significance is not consistent with the .05 alpha levels used by previous researchers. If the aforementioned researchers raised their acceptable error level, virtually all of them could have claimed debaters significantly out gained nondebaters on the WGCTA. Ironically, Hill (1993) quoted Whalen's (1991) conclusion because it was stated "more emphatically" (p. 7). Given the

many deficiencies of Whalen's (1991) study, any comparison or conclusion based on it is meaningless.

#### THE WATSON GLASER CRITICAL THINKING APPRAISAL

Hill (1993) offered two recommendations for future debate-critical thinking research--clarification and diversification. The basic proposition underlying Hill's (1993) general principle of "clarification" relies on criticism of the WGCTA. While a superior measuring instrument is always desirable, many of Hill's (1993) indictments concerning the WGCTA are not supported by the preponderance of empirical research. The following section briefly describes the WGCTA and discusses reliability and validity of the instrument as it relates to the "clarification" principle.

The WGCTA measures five dimensions of critical thinking: 1) Inference 2) Recognition of Assumption; 3) Deduction; 4) Interpretation; and 5) Evaluation of Arguments. Sixteen questions (Forms *A* and *B*) are used in each scale to explore the five "subabilities" (Watson and Glaser, 1980). Together the composite score reflects a sampling of overall critical thinking ability, as it measures various ability domains of the critical thinking process. While other definitions of critical thinking are possible, they are "likely" [to produce] considerable overlap among alternative lists of component abilities" (Watson and Glaser, 1980 p. 1). The WGCTA provides an operational definition to measure the critical thinking construct and appears the most promising to satisfy the educational accountability demands Hill (1993) contended "may be imperative to the existence of our [debate] programs" (p. 2).

The individuals demanding "program accountability" require the most reliable and valid instrument available, especially when comparisons are made with competing programs that empirically assess their outcomes. The WGCTA is the most reliable measure of critical thinking that exists. The reliability of the total test score is adequate. Annis and Annis (1979) observed, "In critical reviews the Watson-Glaser [Critical Thinking Appraisal] generally has been evaluated as an effective test instrument" (p. 221). Crites (1965) supported the WGCTA stating, "Watson-Glaser represents an approach to the measurement of ability which is novel, as far as item content and format are concerned, and is a laudable approach. It is also one which data on the test justify as empirically useful . . . [and] its internal consistency is high . . ." (p. 785)

In 1980 the WGCTA changed from Ym and Zm to the currently used A and B forms. In a review of the *A* and *B* forms Helmstadter (1983) writes in a review of the Watson-Glaser Critical Thinking Appraisal:

A wide variety of reliability indexes have been computed using different groups and different method for assessing this characteristic of tests. . . . [I]n suggesting uses of the test, proper professional restraint and scientific caution have been used. All of the

above contributes to a feeling of confidence that this test is a good, solid measure of adequate--but not outstanding--reliability." (p. 1692-3).

Keyser and Sweetland (1985) reported:

In a recent survey of a panel of psychologists, Mofjeski and Micheal (1983) found the WGCTA to meet more of the criteria for a psychological test than did its only competitor, the Cornell Critical Thinking Test (Ennis, Millman, & Tomko, 1979). . . . [T]he WGCTA was rated more highly with respect to having clearly defined the universe of situations and how it was sampled. Overall, the WGCTA rated as superior to the Cornell in terms of the test criteria described as 'essential' in the Standards for Educational and Psychological Test. (American Psychological Association, 1974). (p. 685)

Berger (1983) concurred, "This is a well-constructed test. . . . This reviewer knows of no similar test that is on a par with the WGCTA." Woehlke (1985) concluded, "this reviewer recommends the WGCTA as the best available instrument for measuring critical thinking ability" (p. 685). In short, the WGCTA is reliable for group comparisons of total test scores of critical thinking in educational settings, so long as the subscales are not generalized as independent measures.

The majority of Hill's (1993) "clarification" issues related to specific validity concerns. For example, he quoted McPeak (1990) who accused Watson-Glaser of taking their definitions from a "list in a government document" (p. 10). This content validity challenge deserves further explanation. Watson-Glaser spent more than forty years developing and enhancing the definition of critical thinking. In constructing the original WGCTA, thirty-five scholars from a wide variety of disciplines and universities contributed to constructing test items to enhance content validity. The five abilities as measured by the WGCTA are consistent with, though not exhaustive, most definitions of critical thinking. The assertion that Watson-Glaser's definitions "have not been demonstrated to be the constituents of critical thinking" (Hill, 1993 p. 10) is not supported by any convincing empirical research. Anastasi (1982) observed, "Content validation involves essentially the systematic examination of the test content to determine whether it covers a representative sample of the behavior domain to be measured" (p. 131). Hill (1970) explained:

The five subtests [of the WGCTA] . . . are clearly pertinent to most definitions of 'critical thinking'. . . . The WGCTA is one of the most useful instruments to understand and appraise critical thinking. (p. 796-7)

Comparisons between the conceptual definitions of Dewey (1910), Dressell and Mayhew (1954), Brembeck (1949), Follman, Brown, and Burg (1970), Ennis (1969), Drake (1976), and the WGCTA subscales correspond to a great degree. The composite of abilities as measured by the WGCTA is a fair representation of overall critical thinking ability, even though it may not include every conceivable rudiment of the critical thinking process.

The ensuing admonishment of the WGCTA pertains to the issue of construct validity. The "more serious problem" is the WGCTA does not "provide a sufficient basis to explain our [debate community] educational mission" (Hill, 1993, p.11). The abilities of the WGCTA are consistent with the objectives of many forensic educators. The ability to define a problem directly relates to the issues of topicality. The ability to select pertinent information for the solution of a problem is reflected in the burden of proof requirements and solvency issues frequently argued in debates. The ability to recognize stated and unstated assumption is found in the practice of debaters analyzing published authoritative proof in support of claims. The ability to formulate and select relevant hypotheses are depicted through interpreting resolutions, writing plans and criteria, and developing cases of advocacy. The ability to draw valid conclusions and judge the validity of inferences are intrinsic components in all debating. The reasons why measuring these five abilities to fail generate a numerical bench mark of overall debater-critical thinking are unclear. The problem solving framework operationalized by the WGCTA and mastering these abilities is a reasonable representation of critical thinking goals of debating. Hills' (1993) criticism of the WGCTA would be applicable to most psychological measures. Do I. Q. tests measure every aspect of intelligence? Do ACT, SAT, or GRE tests measure every ability gained or needed representing a student's education? They do not and cannot. However, their utility in providing information about thinking abilities remains important and influential. Given no measuring device can capture every element of complex thought processes like intelligence, aptitude, or critical thinking, instruments like the WGCTA are preferred because they are far more precise than the "qualitative research alternatives."

Another validity contention of Hill (1993) concerned the correlation of the WGCTA with measures of intelligence. He referenced McPeak (1990), Kurfiss (1988), and Helmstader (1985) contending the WGCTA measures the same ability as I.Q. tests. McPeak (1990), Kurfiss (1988), and Helmstader (1985) supports their conclusion with anecdotal claims about the WGCTA without empirical support. It is inconsistent for Hill (1993) to challenge the debate-critical thinking studies on the basis of "unconvincing empirical proof" and then make factual conclusions without a similar burden. Factor analysis has revealed the WGCTA correlates with general intelligence, but its overlap as a construct is not complete. For example, Landis (1976) factor analyzed the WGCTA with measures drawn from the Guilford Structure of Intellect Model. The WGCTA reflected a dimension of intellectual functioning that is independent of that tapped by the measures of the intellect system. Follman, Miller, and Hernandez (1969) also reported high loadings on a single factor, when analyzed along with achievement and ability measures. Follman, Brown, and Burg (1970) reported a factor analysis of the WGCTA:

It appears that the basic structure of the interrelations of the (WGCTA) is not a general ability but a composite of different groups and specific factors each accounting for a relatively small percent of the variance. (p. 16)

Combining the WGCTA with personality and aptitude scales have produced two factors, (Westbrook and Seller, 1967) five factors, (Singer and Roby, 1967) and even more. (Hunt and Randhawa, 1973) Their factor analyses supported the claims of Watson and Glaser that their test overlaps with intelligence, while retaining an unidimensional quality. Certainly, some intelligence level seems necessary to develop higher level thought processes like critical thinking, but this does not imply the WGCTA only measures intelligence.

Similarly, Hill (1993) also referenced McPeak's (1990) assertion the WGCTA measures "little more than reading comprehension" (p. 15). The *A* and *B* forms of "the WGCTA w[ere] carefully examined for reading difficulty using three indices: the Dale-Chall, the Fry, and the Flesch. Sections of the test that exceeded a ninth-grade reading level were either modified or eliminated" (Watson and Glaser, 1980 p. 1). The test would therefore be appropriate for those having a ninth-grade reading level or above. After postulating the WGCTA simply measures reading comprehension, Hill (1993) reasoned, "much of the critical thinking presumably involved in debate occurs about information that is presented orally" (p. 15). Although speeches during debates are presented orally, much critical thinking occurs when reading, marking, selecting, and processing information into briefs that debaters use to support their claims. This is not to say that critical thinking does not occur during debates, but successful debaters prepare and develop arguments well in advance of the actual presentation. Critical thinking is also likely developed through discovery, information processing, hypothesis testing, interaction, and reflective processes, not exclusively from the oral component of the activity. A compelling logical nexus between the debate process and critical thinking development can be made directly relating to the activities of processing written information.

Overall, the WGCTA represents a reliable, valid, and appropriate measure to quantify the critical thinking abilities of debaters. It is the best available measure of critical thinking ability. According to the American Psychological Corporation, a more refined version is scheduled for release in 1996. While experimenting with other methods of inquiry is interesting and qualitative assessment has been reported in the literature since classical rhetoric, the debate community would be well served at this time to continue to develop the foundation of empirical research based on the WGCTA.

#### FUTURE RESEARCH

Do critical thinkers migrate toward academic debating or does academic debating enhance critical thinking? The two seem intrinsically related. Whether the chicken or the egg came first, may not be as important as considering whether one could develop without the other. Few scholars challenge the importance of developing critical thinking skills and that students participating in debate generally have a higher level of critical thinking than their nondebating counterparts. In this regard, debate educators have a tremendous opportunity and

obligation to maximize a unique educational interaction for these highly talented students. Therefore, how should the forensics community proceed with future critical thinking research?

While Hill's (1993) suggestion for "diversification" in debate-critical thinking research is interesting, it does not address his premise of producing defensible data in the age of educational accountability. First, no other empirical measure is equal or superior to the WGCTA and the vast majority of educational researchers in many domains utilize it. Second, the use of qualitative explanations of "how" argument enhances thinking has been explored throughout history dating back to classical rhetoric. Third, it appears unlikely individual researchers will abandon training and research skills taking many years to acquire. Conducting both types of research does not appear mutually exclusive. The danger in advocating a shift to only one was articulated by Anderson (1987):

Method becomes ideology because it is useful to control membership (you cannot be one of us unless you know our methods); to distinguish nonmembers (you cannot be one of us because you do not use our methods); and to distribute power and resources (you must use our methods to eat at our table) (p. 17).

Qualitative and descriptive research concerning debate-critical thinking research should *not* be discouraged. However, it should *not* be considered a substitute for or in place of well-designed empirical study.

Improving debate-critical thinking research will require longitudinal trend, cohort, and panel studies. Empirical replications should explicate the precise methods, instructional content, and practices found to be most effective. These studies should report the specific components of the debate experience resulting in optimal critical thinking development. Advancing a debate curriculum based on cumulative empirical research offers greater potential benefits than adopting a "crisis mentality" and starting over. In this regard, two principles should be considered by future researchers--replication and cooperation.

#### REPLICATION

Previous debate-critical thinking research focused on epistemology. It has consistently shown some methods of debate training and competitive experience can elevate critical thinking, while others have not. The ontology and praxiology issues in the debate-critical thinking relationship remain unexplored. Some instructors, for example, may approach the task by teaching debaters game strategies. Others emphasize classical argumentation and rhetorical theories, or subscribe to problem-solving, hypothesis testing, judicial, and legislative paradigms. There are many different perspectives available for teaching competitive debate, but little or no research to discern which method(s) works best regarding critical thinking ability development. The specific characteristics, methods, and practices of tournament debating need additional investigation.

Replication of debate-critical thinking research should focus on the specific nature of the debating process. Classroom studies need to collect and report specific course characteristics. Some courses may have the goal of obtaining critical thinking, while others may not. What teaching methods are being used? Lecture formats may not be as effective as mastery learning strategies or those promoting interactive learning. Teaching philosophies may also play a significant role in the outcomes of a debate course. Instructional content, performance activities, evaluation procedures, course organization, and requirements can all play a role in developing critical thinking. The availability and afford ability of video recording can make subsequent analysis of the treatment far more precise and meaningful. Replicating classroom debate-critical thinking studies could provide valuable information about "how to" teach critical thinking using debate. Replicating studies with better control, reporting, and observation of various strategies could provide guidance for debate educators in developing their programs. As Babbie (1995) observed, "Replication can be a general solution to problems of validity in social research" (p. 326).

Beyond teaching methods and strategies, the process or framework used by debaters needs examination. As Meyers (1986) stated, "No matter what specific approach is used, a teacher must present some explicit perspective or framework for disciplinary analysis--a structure for making sense of the materials, issues, and methodologies of the discipline being taught" (p. 6). The debate activity through its framework for analysis may be a catalyst and motivator for critical thinking, rather than a specific cause of it. Myers (1986) continued, "Teaching a framework for analysis will be in vain unless students have the motivation to engage in critical thinking" (p. 8). Research should isolate the demand characteristics of preparing to debate (research) and the optimal amount debating (number of rounds, tournaments, topics, years, etc. . . .) enhancing critical thinking short of diminishing returns. Is it beneficial to hold tournament competitions lasting ten to twelve hours a day over three to four days? Should students attend five, ten, fifteen, or more tournaments during an academic year? Past research supports a debate-critical thinking relationship, the challenge now is to discover why some programs produce significant positive outcomes and others do not. Building on existing data appears more productive, than reinventing the wheel or focusing on less defensible methodology.

Improving the replication of debate-critical thinking research requires greater precision. Researchers should follow the recommendations and instructions of those designing measuring instruments. Inferring from subscales, using one nonequivalent form, unrepresentative sampling, not reporting effect size, and inappropriate instrument administration practices reduce detecting valid findings using the WGCTA or any measuring instrument. Building on past designs by using the pretest as a covariant is an effective method to control for extraneous variations (i.e., self-selection, I. Q., past experience, differing experimental treatments . . .) that otherwise complicate interpreting results. Multiple classification analyses can also be used to

determine the extent of extraneous variation. More over, researchers using the WGCTA should carefully read and follow the instructions found in the appraisal manual. And before embarking on new, untested, and underdeveloped instruments, debate-critical thinking researchers should continue using the WGCTA until a better instrument is available. Many other educational domains use the WGCTA to justify their pedagogical methods. If accountability becomes relevant to program existence, forensic educators would want to report the "best" available measuring device was used to demonstrate its results. As Anderson (1987) observed, "The proof of the method is its utility in solving the problem" (p. 17). In short, abandoning this genre of research would strengthen the anecdotal criticism about the debate-critical thinking relationship making educational accountability more difficult.

#### COOPERATION

Few educators in the forensic community deny the advantages of additional debate-critical thinking research. However, few have been willing or able to volunteer the time, effort, and expense needed to conduct long term studies producing incontrovertible data. All of the major studies reported with adequate sample size were doctoral dissertations and most were subsequently published in academic journals. Given the apparent pragmatic difficulties inherent in debate-critical thinking research, it is understandable why few of these researchers reported subsequent studies. Forensic educators often have similar teaching, research, and service responsibilities as their colleagues. In addition, travel to debate tournaments typically consumes anywhere from five to twenty-weekends out of town, significant practice and preparation time, and substantial administrative tasks. In short, even the most dedicated forensics scholars have insufficient resources to individually conduct well-designed research projects measuring the effects of debating on critical thinking.

Cooperation is needed by the debate community and its organizations (NDT, CEDA, NFL, ADA . . . ) to coordinate, finance, support, and sponsor longitudinal studies. Debate-critical thinking research is expensive, time consuming, and dependent on highly specialized experts. Stronger research designs requiring many subjects contend with high attrition rates. The administration of the WGCTA requires between forty minutes and an hour making administration between debate rounds impractical at tournaments. The pragmatic difficulties of test administration, data collection, and processing are generally too difficult for individuals traveling as coaches, while also serving as full-time faculty.

Debate-critical thinking research, requires debate organizations to sponsor, coordinate, and support ongoing longitudinal research programs to determine not only the effects of debate on critical thinking, but the effects of debating on a variety of behavioral outcomes believed to be associated with participating in debate. As the academic debate community continues to evolve and it's debate organizations establish and perpetuate their educational philosophies, they

should follow a course proven to accommodate important educational outcomes like critical thinking. Debate organizations adopting value orientations had no empirical proof that discarding the problem solving framework practiced by policy debaters would be equal to or accelerate critical thinking skill development. Despite the organizational implications, forensic educators should promote and embrace research that identifies the best educational experience for participants. The fragmentation of the academic debate community necessitates cooperation to advance comprehensive critical thinking longitudinal studies.

In summary, the research spanning five decades supporting a debate-critical thinking relationship is under attack. Hill (1993) has suggested a radical shift in debate-critical thinking and essentially advocated redefining the concept itself. While clarification and diversity should be considered, the data and methods of measurements of existing studies should not be abandoned. A preponderance of defensible research already suggests debating, when properly taught, can enhance the critical thinking abilities as measured by the WGCTA. Critical thinking is difficult to measure, but these difficulties should not discourage scholars from developing an important field of research. The collective efforts of the forensics community could strengthen empirical investigations and contribute to maximizing critical thinking improvement.

This paper has responded to recent criticism regarding debate-critical thinking research. The appropriate classification and critical review of these studies reveal a wealth of information. Investigating complex thought constructs that apparently interact with behavioral activities present a difficult and interesting challenge. The instruments to measure critical thinking continue to improve and the WGCTA remains the best available measuring device. The academic debate communities innovating this genre of study should further replicate, extend, and develop empirical debate-critical thinking research. Cooperation among and between the many fragmented debate organizations' membership is needed for long term longitudinal studies assessing, measuring, and reporting the specific characteristics of methods of different forms of academic debating.

Despite many limitations, the collective body of debate-critical thinking research supports at least three important inferences: 1) extended periods of training and experience in academic debating enhance critical thinking abilities as measured by the WGCTA; 2) specific educational strategies and certain academic debating experiences significantly increase critical thinking abilities as measured by the WGCTA; and finally, 3) the debate-critical thinking literature provides presumptive proof favoring a positive debate-critical thinking relationship. Admittedly, not every form of, instruction in, strategy used, or tournament experience in competitive academic debate assures critical thinking abilities for every individual. Research suggests some debate programs develop critical thinking abilities better than others. Cooperation and support by a collective debate community for long-term replicated longitudinal studies offer the best opportunity for educational accountability and improving academic debate.

## WORKS CITED

- Agne and Blick, D. (1972). (1972). A comparison of earth science classes taught by using original data in a research approach technique versus classes taught by conventional approaches not using such data. *Journal of Research in Science Teaching*, 9, 83-89.
- Anastasi, A. (1987). *Psychological testing*. New York: MacMillan.
- Anderson, J. A. (1987). *Communication research: Issues and methods*. New York: McGraw-Hill.
- Annis, L. and Annis, D. (1979). The impact of philosophy on students' critical thinking ability. *Contemporary Educational Psychology*, 4, 219-226.
- Babbie, E. (1995). *The practice of social research*. (7th ed.) Belmont, CA: Wadsworth.
- Bartanen, M. and Frank, D. (1994). *Nonpolicy debate*. Scottsdale, AZ: Gorsuch Scarisbrick.
- Beckman, V. (1955). *An investigation of the contributions to critical thinking made by courses in argumentation and discussion in selected colleges*. Unpublished doctoral dissertation. University of Minnesota, Minneapolis, Minnesota.
- Berger, A. (1983). Review of the Watson-Glaser Critical Thinking Appraisal. *Eighth Mental Measurement Yearbook*, p. 1347.
- Brakken, E. (1993). Intellectual factors in PSSC and conventional high school physics. *The Journal of Research and Science Teaching*, 3, 19-25.
- Brembeck, W. (1947). *The effects of a course in argumentation on critical thinking ability*. Doctoral dissertation, University of Wisconsin, Madison, Wisconsin.
- Brembeck, W. (1949). The effects of a course in argumentation on critical thinking ability. *Speech Monographs*, 16, 172-89.
- Brewer, K. (1984). *Introductory statistics for researchers*. Minneapolis, MN: Burgess.
- Burns, R. (1974). The testing of a model of critical thinking ontogeny among Central Connecticut State College undergraduates. (Doctoral dissertation, Central Connecticut State College, 1973) *Dissertation Abstracts*, 5467-A.
- Campbell, D. and Stanley, J. (1963). *Experimental and quasi-experimental designs for research*. American Educational Research Association.
- Colbert, K. R. (1987). The effects of CEDA and NDT debating on critical thinking ability. *The Journal of the American Forensics Association*, 23, 194-201.
- Colbert, K. and Biggers, T. (1985). Why should we debate? *The Journal of the American Forensics Association*, 21, 237-240.
- Cohen, J. (1969). *Statistical power analysis for the behavioral sciences*. New York: Academic Press.
- Crites, J. (1965). Test reviews. *Journal of Counseling Psychological*, 12, 328-30.
- Cross, G. (1971). *The effects of belief systems and the amount of debate experience on the acquisition of critical thinking*. Unpublished doctoral dissertation, University of Utah, Salt Lake City.
- Dewey, J. (1910). *How we think*. New York: D.C. Heath and Co.
- Drake, J. (1976). *Teaching critical thinking*. Dansville, IL: The Interstate Printers & Publishers.
- Dressel, P. and Mayhew, L. (1954). General education: Exploration in evaluation. *Final Report of the Cooperative Study of Evaluation in General Education*. Washington, D.C.: American Council on Education.
- Fogg, C. and Calia, V. (1967). The comparative influence of two testing techniques on achievement in science and critical thinking ability. *The Journal of Experimental Education*, 35, 1-14.
- Follert, V. and Colbert, K. (1983). *An analysis of the research concerning debate training and critical thinking improvements*. (ERIC Document Reproduction Service No. ED 238 058)
- Follman, J., Brown, L. and Burg, E. (1970). Factor analysis of achievement scholastic aptitude, and critical thinking subtests. *The Journal of Experimental Education*, 38, 11-16.

- Follman, J., Miller, W. and Hernandez, D. (1969). Factor analysis of achievement scholastic aptitude, critical thinking subtests. *The Journal of Experimental Education*, 38, 48-53.
- Frank, D. (1969). Teaching high school speech to improve critical thinking. *The Speech Teacher*, 18, 296-302.
- Fritz, P. and Weaver, R. (1984). Teaching critical thinking skills in public speaking courses: A liberal arts perspective. Paper presented at the *Annual Meeting of the Speech Communication Association*. (70th, Chicago, Ill., November 1-4, 1984). (ERIC Document Reproduction Service No. ED 249 556)
- George, K. (1968). The effects of critical thinking ability upon course grades in biology. *Science Education*, 52, 421-6.
- Gruner, C. R., Huseman, R. C. and Luck, J. I. (1971). Debating ability, critical thinking ability, and authoritarianism. *Speaker and Gavel*, 8, 63-4.
- Helmstader, G. (1965). Review of the Watson-Glaser Critical Thinking Appraisal. *The Seventh Mental Measurement Yearbook*, p. 784.
- Helmstader, G. (1983). Review of the Watson-Glaser Critical Thinking Appraisal. *The Eighth Mental Measurement Yearbook*, p. 1347.
- Henkel, T. (1968). Undergraduate physics instruction and critical thinking ability. *The Journal of Research and Science Teaching*, 5, 89-94.
- Hill, B. (1993). The value of competitive debate as a vehicle for promoting the development of critical thinking ability. *CEDA Yearbook*, 14, 1-22.
- Hill, W. (1970). Review of the Watson-Glaser Critical Thinking Appraisal. *The Fifth Mental Measurement Yearbook*. New Brunswick: Turgers.
- Huseman, C. I. (1970). Review of the Watson-Glaser Critical Thinking Appraisal. *The Fifth Mental Measurement Yearbook*, 5.
- Hovland, C. I. (1970). Review of Watson-Glaser Critical Thinking Appraisal. *The Fifth Mental Measurement Yearbook*, 5.
- Howell, W. (1943). The effects of high school debating on critical thinking. *Speech Monographs*, 10, 96-103.
- Hunt, D. and Randhawa, B. (1973). Relationship between and among cognitive variables and achievement in computer science. *Education and Psychological Reports*, 36, 380.
- Jackson, T. (1961). *The effects of intercollegiate debating on critical thinking*. Unpublished doctoral dissertation, University of Wisconsin.
- Johnson, A. (1943). An experimental study in the analysis and measurement of reflective thinking. *Speech Monographs*, 10, 83-96.
- Katula, R. and Martin, C. (1984). Teaching critical thinking in the speech communication classroom. *Communication Education*, 33, 160-167.
- Keyser, D. and Sweetland, R. (1985). *Test Critiques*. Vol. III.
- Kirk, R. (1968). *Experimental design: Procedures for the behavioral sciences*. Belmont, CA: Brooks/Cole.
- Kurfiss, J. G. (1988). *Critical thinking: Theory, practice, and possibilities*. Washington, D. C.: Association for the Study of Higher Education.
- Landis, R. (1976). The psychological dimensions of three measures of critical thinking and twenty-four structures of intellect tests for a sample of ninth-grade students. *Dissertation Abstracts International*, 5705-A.
- Lucas, A. (1972). Inflated posttest scores seven months after pretest. *Science Education*, 56, 381-7.
- Luck, J. L. and Gruner, C. R. (1970). Note on authoritarianism and critical thinking ability. *Psychological Reports*, 27, 380.
- McPeak, J. (1990). *Teaching critical thinking: Dialogue and dialectic*. New York: Routledge.
- Modjeski, R. and Michael, W. (1983). An evaluation by a panel of psychologists of the reliability and validity of two test of critical thinking. *Educational and Psychological Measurement*, 43, 1187-1197.
- Meyers, C. (1986). *Teaching students to think critically*. San Francisco: Jossey-Bass.

- Ness, J. (1967). The effects of a beginning speech course on critical thinking ability. (Doctoral dissertation, University of Minnesota, Minneapolis, Minnesota). *Dissertation Abstract International*, 28, 4826A.
- Nie, N., Hull, C., Jenkins, J., Steinbrenner, K. and Brent, D. (1975). *Statistical package for the social sciences*. (2nd ed.). McGraw-Hill.
- Perella, J. (1983). *The debate method of critical thinking*. Dubuque, IA: Kendall Hunt.
- Rust, V. (1962). A factor-analytic study of critical thinking. *The Journal of Educational Research*, 55, 253-9.
- Seymour, L. and Sutman, F. (1973). Critical thinking ability, open-mindedness, and knowledge of the process of science of chemistry and on chemistry students. *The Journal of Science Teaching*, 10, 159-63.
- Smith, D. (1977). College classroom interactions and critical thinking. *Journal of Educational Psychology*, 69, 180-190.
- Sorenson, L. (1966). Watson-Glaser Critical Thinking Appraisal: Changes in critical thinking associated with two methods of teaching of teaching high school biology. *Test Data Report No. 51*. New York: Harcourt Brace and World.
- Trank, D. (1977). *Secondary forensics programs: Can they survive current pressures?* (ERIC Document Reproduction Service No. 170 824)
- Trank, D. (1978) *Back to basics: A case for teaching forensics*. (ERIC Document Reproduction Service No. 170 824)
- Watson, G. and Glaser, E. (1980). *Critical thinking appraisal manual for forms A and B*. New York: Harcourt Brace Jovanovich.
- Watson, G. and Glaser, E. (1964). *Critical thinking appraisal manual for forms YM and ZM*. New York: Harcourt Brace Jovanovich.
- Westbrook, B. and Sellar, J. (1967). Critical thinking, intelligence, and vocabulary. *Education and Psychology*, 27, 443-6.
- Whalen, S. (1991). Intercollegiate debate as a co-curricular activity: Effects on critical thinking skills. In Donn W. Parson (ed.), *Arguments in Controversy: Proceedings of the Seventh SCA/AFA Conference on Argumentation*, (pp. 391-97). Annandale, VA: Speech Communication Association.
- Williams, D. (1951). *The Effects of Training in College Debating on Critical Thinking Ability*. Unpublished Master's Thesis, Purdue University.
- Wilson, D. and Wagner, E. (1981). The Watson-Glaser Critical Thinking Appraisal as a predictor of performance in a critical thinking course. *Educational and Psychological Measurement*, 41, 1319-1323.

## Counterfactual Possibilities: Constructing Counter-to-fact Causal Claims

KENNETH T. BRODA-BAHM

*Advertising fosters consumption. If advertising was not as pervasive in this society, then the level of consumption would be much less than its current level.*

The preceding is a counterfactual argument.<sup>1</sup> It asserts that certain results would obtain if conditions were different than they presently are. In the field of academic debate, such counterfactual claims recently have been the subject of increasing attention. Both inside and outside of the debate round, students and teachers of debate have confronted issues related to the validity and the meaning of such claims. Scholarship on counterfactual analysis in academic debate to date has focused on issues such as the paradigmatic validity of counterfactuals (Roskoski, 1992; 1994), their relationship to topicality and competition (Korcok, 1994), their applications within recent CEDA topics (Broda-Bahm, 1994; Hoe, 1994; Roskoski, 1994), problems associated with their use (Berube & Pray, 1994; Voight & Stanfield, 1992), and the issue of infinite regression (Broda-Bahm, 1994).

Without a doubt, many technical issues await resolution. These concerns can be most clearly addressed, however, in the context of a clear understanding of the possible meanings of counterfactual analysis and application. Accordingly, after first justifying the the need for a schema for counterfactual analysis, the present essay will consider several such schemata, and ultimately advocate one as a consistent and relatively clear method of constructing counterfactual claims.<sup>2</sup> Such a focus on how we conceive of the counterfactual is not merely an exploration into a single "exotic" argument form. As the next section will show, the

---

<sup>1</sup>A counterfactual claim, also known as a "contrary-to-fact conditional" or a "subjunctive conditional," can be defined as an assertion about matters which are not, at present, believed to exist. Counterfactual claims are made whenever speculation centers on what *would* happen, *if* something were the case. As Richard Creath (1989) notes, "As a first approximation we might say that a counterfactual is any sentence which says what *would* happen under specified conditions, even though those conditions do not in fact obtain. . . . A typical, if somewhat shopworn, example of a counterfactual is: 'If I had struck that match, it would have lit'" (p. 95).

<sup>2</sup>It is important to note that in promoting a *schema* for constructing counterfactual claims, I am not articulating a *test* of their validity as causal statements. There are several well-known tools for evaluating causal statements (e.g., see Mill, 1900, pp. 255-66.). The present essay deals with the question of how to structure, articulate, or set-up the counterfactual claim in the first place. It is a descriptive step: when we make a counterfactual statement, what do we mean?

*Kenneth Broda-Bahm is Assistant Professor and Director of Forensics in the Department of Speech and Mass Communication at Towson State University. Portions of this essay were presented at the 1994 Speech Communication Association Convention in New Orleans.*