

Assessing EFL Student Progress in Critical Thinking With the Ennis-Weir Critical Thinking Essay Test¹

Bruce W. Davidson
Hokusei Gakuen University

Rodney A. Dunham
Tezukayama College

Recent trends in EFL/ESL have emphasized the importance of promoting thinking as an integral part of English language pedagogy; however, empirical research has not established that training in thinking skills can be effectively combined with EFL/ESL instruction. This study made use of the Ennis-Weir Critical Thinking Essay Test to assess progress in critical thinking after a year of intensive academic English instruction among Japanese students ($N=36$). A control group received only content-based intensive English instruction, while a treatment group received additional training in critical thinking. The treatment group scored significantly higher on the test ($p<.001$). The results imply that critical thinking skills can indeed be taught as part of academic EFL/ESL instruction.

近年、第二言語および外国語としての英語教育では思考力の育成を英語教育の一部として扱うことの重要性が強調されている。しかしながら思考技術のトレーニングが英語教育と効果的に統合できるという結果を示す実証的研究はいまだない。本研究では、Ennis-Weir 批判的思考作文テストを用いて、36人の日本人学生が一年間の集中的アカデミックな目的のための英語教育を受けた後にどの程度批判的思考力が発達したかを測定した。統制群は内容中心の英語教育のみを受け、実験群はそれに加えて批判的思考法のトレーニングを受けた。実験群の成績は統制群より有意に高かった。この結果は、批判的思考技術がアカデミックな目的のための英語教育の中で教えることができるのを示している。

Since the advent of research into cognitive development, language teachers and linguists generally have recognized the close connection between language learning and thinking processes. In particular, ESL reading research has shown some correlation between ESL reading comprehension and familiarity with the formal or content schemata of English texts (Carrell, 1987). Furthermore, noting the

unreflective character of many language-teaching approaches that only encourage verbal output or passive input, Tarvin and Al-arishi (1991) and Al-arishi (1994) have explored some methods to make language teaching more thoughtful. Similarly, Chamot (1995) has argued from current educational trends promoting higher-order thinking that EFL/ESL teachers also need to turn the classroom into a community of thinkers. Informal observations may indicate that thinking skills can indeed be taught in an EFL/ESL context (Davidson, 1994, 1995). Without formal testing, however, it is difficult to establish this concretely. Though there has been a lot of thought and research devoted to the development of critical thinking skills in native English speaker educational programs, there has been little research in the area of combining critical thinking with EFL/ESL instruction.

Content-based intensive English instruction has also proven to have many advantages and possibilities (Snow & Brinton, 1988). Is one of them the promotion of critical thinking skills through thought-provoking content? It might be expected that such abilities will develop through discussion, reading, or composition about subjects requiring some serious analytical attention; however, Chance's (1986) survey concluded that critical thinking skills do not develop simply as a by-product of the study of specific subjects. In addition, Halpern (1993) cites evidence from various sources that critical thinking skills can be inculcated through explicit instruction.

These issues and findings inspired a pilot study to discover whether or not critical thinking could be taught to Japanese students of English in a content-based EFL program. After defining what we mean by "critical thinking," we will describe the intensive English program the subjects were enrolled in along with the specifics of the current study. Two research questions guided us:

1. On a critical thinking test task, will English learners exposed to critical thinking skills-training do significantly better than similar students who have not received such training?
2. Can a critical thinking test designed for native English speakers be used as an instrument for evaluating critical thinking skills among non-native English learners?

Critical Thinking: Concept and Inventory of Component Skills

Critical thinking involves rational judgment and discernment of the elements of reasoning. Various definitions of critical thinking reflect

this. Norris and Ennis (1989) explain critical thinking as "reasonable and reflective thinking that is focused upon deciding what to believe and do" (p. 3), a definition also stated somewhat differently by Lippman (1991), who defines it as healthy skepticism, and Siegel (1988), who considers the critical thinker to be one who "is appropriately moved by reasons" (p. 2). In contrast to rote memorization or simple information recall, methods for encouraging critical thinking have as their goal the stimulation of the analytical and evaluative processes of the mind (Paul, 1992). Norris and Ennis (1989) have listed a number of critical thinking abilities to develop:

Elementary Clarification

1. Focusing on a question
2. Analyzing arguments
3. Asking and answering questions that clarify and challenge

Basic Support

4. Judging the credibility of a source
5. Making and judging observations

Inference

6. Making and judging deductions
7. Making and judging inductions
8. Making and judging value judgments

Advanced Clarification

9. Defining terms and judging definitions
10. Identifying assumptions

Strategies and Tactics

11. Deciding on an action
12. Interacting with others (p. 14)

This concept of critical thinking and inventory of skills inspired both the instructional treatment and the selection of the Ennis-Weir test in this study.

The Study

Method

Subjects: All participants in the study ($N=36$) were first-year students enrolled in a private women's junior college in Osaka, Japan. The college's curriculum consisted mainly of an intensive academic English program. Weekly English courses included Oral Discussion (3 hours), Composition (2 hours), Reading (3 hours), Pronunciation (3 hours),

and Grammar/Listening (2 hours), totaling 13 hours a week, considering each 50-minute class session as an hour. Oral Discussion, Reading, and Composition followed a topical syllabus of six units of instruction, which included such themes as Prejudice/ Human Rights, Advertising/ Consumerism, and Women's Issues/Child-raising. The integrated, content-based aspect of the program was meant to involve students in in-depth analysis and expression concerning subjects significant in their own lives and in Japanese society. This course of study would seem well-suited to encouraging the development of critical thinking skills as a by-product, since the topics all necessitate thought. Along with the topic, each unit also introduced a rhetorical mode: Illustration, Process, Definition, Classification, Comparison/Contrast, and Persuasion. The first three composition units required students to write a paragraph using each mode, and the last three progressed to multi-paragraph essays. The persuasion essay was written in a mini-term paper format, with references.

In addition to these integrated intensive English courses, Ss took a weekly one-hour seminar course, also conducted in English and concerning some interesting topic or theme such as *American Holidays* or *Traditional Folk Songs*. The treatment group ($n=17$) was composed of students from a seminar on Critical Thinking.² Volunteers not enrolled in the critical thinking seminar ($n=19$) served as a control group.

Ss had varying degrees of English proficiency as measured by an in-house proficiency test. At the beginning of the year, this test divided all students into five levels of classes according to scores: A, B, C, D, and E. The A classes had the highest level of proficiency and included students returning from a year or longer of study abroad, whereas the E classes were much less proficient. Regardless of proficiency, however, all classes received similar instruction based on the same content and rhetorical modes noted previously. These Ss represented a broad range of English proficiency levels, as measured by the in-house test, with both groups containing a similar range. Because of the small number of Ss at each level, with the exception of level C ($n=14$), levels were grouped into A+B, C, and D+E for later comparison. The distribution of Ss among the groups was fairly even (control $n=5, 10, 4$; treatment $n=7, 4, 6$). No pre-test was given, in line with the advice of Ennis and Weir (1985), who state that a pre-test is not necessary in research using the test as long as a control group exists. Babbie (1983) has noted that a post-test-only control group design is quite acceptable as long as group assignment is random. Since students enrolled in seminars through a semi-lottery system, the authors consider that in this case group as-

signment generally embodied the spirit of randomness, although it was not completely random.

Treatment: The treatment group took part in a course designed to train them in basic elements of critical thinking: source credibility, inductive reasoning, informal deductive logic, and assumption-identification. These broad categories encompass most of Norris and Ennis's (1989) list of critical thinking skills, so they were adopted as a framework for the seminar course. During the first semester, instruction dealt with inductive reasoning and source credibility; during the second semester, the emphasis was deductive reasoning and assumption-identification. The course began with an introduction to the concept of critical thinking. Sessions were devoted to exploring various kinds of reasoning fallacies and misuse of evidence, such as over-generalization and the false dilemma (Chaffee, 1990; Damer, 1995). Students were given lists of brief fallacious arguments and asked to explain the problems of each in their own words. In the second half of the semester, the focus shifted to source credibility. Students did exercises in which they evaluated varying accounts of the same event according to differing viewpoints. For example, in groups they discussed and ranked accounts of the results of an international conference, keeping in mind a list of question-criteria: Does the news presenter have a reason to be biased? Is the source an expert in the field? (Beyer, 1991). Students brought in similar examples to present and evaluate.

In the second semester the emphasis shifted to basic argument analysis. First, students did exercises to help them distinguish real arguments from bare claims offering no reason (Engel, 1994). Then they identified the claims and supporting reasons. Later in the semester, the instructor introduced less-obvious aspects of deductive reasoning: unstated assumptions and implications (Scriven, 1976). Using magazine advertisements and other material, students practiced identifying assumptions and implications. As a result of the Hanshin earthquake and other circumstances, a total of only 18 class hours was actually devoted to the course's content.

Instrument: The Ennis-Weir Critical Thinking Essay Test

Test Description: The Ennis-Weir test was chosen for various reasons. One is that it is one of the most generally well-accepted measuring instruments among educators in the critical thinking movement (Walsh & Paul, n.d.), and inter-rater reliabilities have been very high when it has been used (Ennis & Weir, 1985; Hatcher, 1996). Another is that, in contrast to multiple-choice tests, it allows students to justify varying responses, and the test

itself presents a realistic critical evaluation task. Other critical thinking tests are available, but almost all of these are multiple-choice instruments that suffer from various weaknesses such as background bias and the impossibility of knowing the reasoning behind an examinee's answer-choice (Ennis, Millman, & Tomko, 1985; Norris & Ennis, 1989). Furthermore, the relatively simple subject-matter and language of the Ennis-Weir test make it suitable for non-native speakers. It has been used successfully with first-year junior high school native English speakers in the U.S.

The test itself contains a simple set of instructions and a letter to a newspaper editor containing 10 brief paragraphs. The fictional writer, Raywift, recommends that overnight parking be prohibited on all the streets of his town, Moorburg. After a brief introduction, eight numbered paragraphs elaborate the argument. Most are weak and commit various common reasoning fallacies such as equivocation, irrelevancy, poor statistical sampling, and circular reasoning, but some contain legitimate support, consisting in the use of qualified experts or a relevant reason. Point-by-point, the examinee's task is to judge the thinking of each of these numbered paragraphs and to evaluate the strength of the letter's argument as a whole in a final summary-paragraph. For example, in responding to Paragraph 3, examinees are expected to notice that a relevant reason is offered to support Raywift's argument. Similarly, test-takers are supposed to show some indication that they comprehend the flaws of the experiment in Paragraph 6. (See Appendix for copies of the introduction and paragraphs 3, 6, and 8.)

A clear and specific scoring protocol accompanies the test indicating various possible answers and how each is to be scored. Points are awarded both for judging correctly and for indicating a valid reason for

Table 1: Critical Thinking Skills Addressed on the Ennis-Weir Test

Paragraph	Skill
1.	Noticing misuse of analogy and/or shift in meaning
2.	Recognizing irrelevant reasoning
3.*	Recognizing relevant reasoning
4.	Recognizing circularity and/or the lack of a reason
5.	Recognizing defective reasoning
6.	Recognizing insufficient sampling
7.	Recognizing equivocation and/or the use of an arbitrary definition
8.*	Evaluating the credibility of expert testimony

*Paragraphs that exhibit sound reasoning

one's judgment, and a penalty point of -1 can be deducted for poor reasoning. Each answer can receive a maximum of 3 points and a minimum of -1, except for the summary paragraph, where a maximum of 5 points can be awarded. Therefore, the overall score can range from -9 to +29. In general, the protocol gives latitude to raters to award points whenever an examinee can give a credible reason in support of this evaluative judgment, even when it differs from that of the protocol writers. Brief answers are acceptable as long as they indicate a valid judgment, backed up with a sound reason for that judgment.

We considered the possibility that cultural differences between America and Japan might bias the results, but we felt that in this case culture would not be a significant issue. For one thing, the Moorburg letter concerns street parking, which is not a point of significant cultural difference between the U.S. and Japan. Street-parking laws also exist in Japan and are stringently enforced. If anything, parking is more of a problem in Japan, so Japanese readers might be even more likely to identify with a parking-related issue than Americans. Furthermore, writing a letter to an editor complaining about a public problem, as in the Raywift letter, is a common practice in Japan as well. The directness and abrasiveness of Raywift's style are perhaps the only aspects of the letter that might seem strange or unsettling to a Japanese reader. However, it should also be noted that abrasive political rhetoric is not unknown in Japan.

A limited amount of research has been done in the U.S. using the test. The largest study to date has been Hatcher's (1996) at Baker University. Over a period of four years (1990-1994), American freshmen scored an average of 11.8 to 13.8 on the Ennis-Weir test after a year-long compulsory critical thinking course. They had scored from 5.8 to 9.4 on a pretest and registered gains of 2.8, 5.8, 5.8, and 6.0 points. Interestingly, a number of Chinese and Japanese students at Baker University also took part in the study, but their scores were eliminated from it because they consistently scored poorly. Hatcher (1994) speculated that their low scores may be due to Oriental politeness and accordingly a hesitancy to criticize the Moorburg letter.

Test Administration: In the last week of second semester classes, the Ennis-Weir Critical Thinking Essay Test (1985) was administered, with the control group given the test within the same week. Both groups had 80 minutes to read the test and write nine brief paragraphs in response, twice the amount of time recommended by the test-makers. Since the subjects were non-native English speakers, it was felt that

more time would be necessary for them to comprehend the material and compose answers. To help them with the language aspects of the test, they were allowed to use dictionaries. Furthermore, before taking the test, all subjects received two sample test items with model answers to make sure the Ss understood two things: (a) that they had to make a clear evaluative judgment as to whether the argument in each paragraph was a good one or not and (b) that they had to give a clear reason or explanation for their judgment. Without such explicit direction, the subjects might not have done either of these two things. However, students in the present study were used to doing peer-evaluation of essays in composition classes, so the idea of writing comments or criticism about a piece of writing was already somewhat familiar. Basic information about the Ss, including English level and overseas experiences, was collected on the answer form.

Results

Tests were scored blindly and independently by two raters. The test-raters in this study found little difficulty in using the protocol to judge student answers. Grammatical or vocabulary problems were overlooked unless they made an answer incomprehensible. Inter-rater reliability was found to be adequate ($r = .72$). The scores and information collected were examined. Therefore, the average scores given to each student for each of the 10 scores on the test, one score for each of the nine paragraphs and a total score, were used for all subsequent analysis.

The small number of Ss in the study (control $n = 19$ and treatment $n = 17$) makes relationship detection difficult unless it is very strong. Because this is the first study of this type, we were interested in detecting moderate relationships as well as strong ones. Consequently, we decided that the risk of committing a Type I error would be less important than missing moderate relationships. Therefore, the significance level of .10 was chosen as the cutoff for accepting or rejecting relationships. Nevertheless, we have reported here the exact probability for all results that indicated statistically significant relationships.

The most important analysis, of course, dealt with the effect of critical thinking training on test scores. Therefore, we conducted a t test to compare the scores of the two groups. The treatment group's mean score on the Ennis-Weir Test was statistically 6.6, significantly higher than the control group's mean score of 0.6 ($t(27.73) = -4.99, p < .001$). Table 2 shows the range of scores for each group and details the differences. As the table shows, 10 Ss in the treatment group scored 7 or

Table 2: Group Comparative Scores on the Ennis-Weir Test

Score Range	Control Group	Treatment Group
-4.0 to 0.0	9	1
1.0 to 2.0	6	1
2.5 to 6.5	4	5
7.0 to 13.5	0	10
Total <i>N</i>	19	17
<i>Mean</i>	0.6	6.6
<i>Median</i>	1.0	7.5
<i>Mode</i>	-1.5	3.0

higher, while only one scored 0 or lower. In contrast, in the control group, nine Ss scored 0 or lower, with no score higher than 6.5.

Next, the individual paragraph scores of the control and treatment groups were compared. There were statistically significant differences between the mean scores of the control group and the treatment group on two paragraphs: the third paragraph ($M = -0.18$ and 0.50 , respectively, $t(33.87) = -2.10$, $p = .043$) and the sixth paragraph ($M = -0.55$ and 0.15 , respectively, $t(27.10) = -2.59$, $p = .015$). The difference in scores on the eighth paragraph approached statistical significance ($t(33.69) = -1.71$, $p = .096$), with the treatment group scoring higher ($M = 0.47$) than the control group ($M = -0.52$). However, the scores on the remaining paragraphs showed no statistically significant difference between the control and treatment groups ($p > .10$).

Since the test was in English, a foreign language for the Ss, proficiency may have affected scores. As mentioned, since the number of Ss at each proficiency level, except C ($n = 14$), was quite small (A level $n = 7$, B level $n = 5$, D level $n = 5$, and E level $n = 5$), A level was grouped with B, and D level with E. The C level was left intact, creating a three-level variable. The distribution for this ordinal variable was compared for levels A + B, C, and D + E between the control group ($n = 5, 10, 4$) and the treatment group ($n = 7, 4$, and 6). As determined by a chi square test, the distribution indicates that there was no relationship between English proficiency and the type of group ($\chi^2(2, N = 36) = 3.204$, $p = .202$). An analysis of variance was also run to examine the relationship between English level and test scores. There was no statistically significant relationship between the two variables ($F(2, 35) = 1.57$, $p = .224$). Table 3 shows that the range of scores was comparable for each proficiency level. Judging by this analysis as well as the phrasing of student answers on the test, we believe that students gen-

erally did not do poorly simply as a result of an inability to understand the English of the test. The sample test items appeared to succeed in helping students to grasp the kind of test task they were engaged in, and the wording of the test did not appear to present an insurmountable problem even for lower-level Ss.

Table 3: Scores on Ennis-Weir Test by English Level

Score Range	A-B Ss	C Ss	D-E Ss
-4.0 to 0.0	4	4	2
1.0 to 6.5	4	8	4
7.0 to 13.5	4	2	4
Total <i>N</i>	12	14	10
<i>Mean</i>	4.3	1.8	4.7
<i>Median</i>	3.8	1.5	4.3
<i>Mode</i>	10.5	1.0	3.0

The next variable examined was overseas experience, since a number of Ss had lived a year or longer in an English-speaking country. Using a *t* test, scores of Ss who had traveled overseas and those who had not were compared. The differences in scores between the two groups was not significant for total scores or for any individual paragraph score ($p > .10$).

Because the Ennis-Weir test deals with parking problems, each of the subjects was asked to report on the test form whether or not she possessed a driver's license. Japanese students often first learn to drive at the age of 19 or 20, the age of the Ss in this study, and familiarity with driving an automobile may have helped some Ss do better on the test, which concerns a parking problem. Scores for the two groups, those with licenses and those without, were compared. Total scores were statistically the same for both groups; however, students without driver's licenses ($M = -0.66$) scored statistically significantly lower on the seventh paragraph than Ss with ($M = -0.17$) [$t(21.14) = -1.84, p = .079$]. On the eighth paragraph, those without licenses ($M = -0.24$) also scored significantly lower than Ss with ($M = 0.70$) [$t(31.50) = -3.31, p = .002$]. Otherwise, the fact of having a driver's license showed no significant relationship with student scores ($p > .10$). Since the specific issues addressed in the seventh and eighth paragraphs are not directly related to the experience of driving we consider the statistical significance to be unrelated to the current study.

These statistical analyses appear to indicate that the differences in scores between the treatment and control groups cannot be accounted for by differences in English proficiency levels or other factors such as overseas experience or having a driver's license. Therefore, the differences in scores on the Ennis-Weir Test can probably be attributed to the critical thinking training given the treatment group.

Discussion and Conclusion

Both research questions can be answered in the affirmative, based on the results of this study. The Ennis-Weir test, designed for native English speakers, appears to be usable for non-native English learners. Furthermore, it is encouraging to find that even a small amount of instruction in the basics of critical thinking appeared to result in higher scores for the treatment group. Critical thinking skills can apparently be taught to some extent along with English as a foreign language and can, therefore, enhance a content-based course of study. In view of the relatively small amount of actual instruction, the rather low average score of 6.6 is not surprising, and is much better than the performance of the control group ($M=0.6$). As a point of comparison, Baker University American freshmen registered gains of 2.8, 5.8, 5.8, and 6.0 (Hatcher, 1996) in four successive years. Interestingly, three of those gains approximate the difference of 6.0 that we found in the mean scores of our two groups, though the mean score of the treatment group (6.6) is only half that of the average post-test scores of the Baker freshmen. Looking at the individual test items, differences between the two groups appeared specifically in items which had received instructional attention in the critical thinking class. Paragraph 6 deals with the misuse of statistics, a reasoning problem dealt with in class, while Paragraph 3 featured a relevant reason, another instructional point. The difference in performance on Paragraph 8, which concerned the use of experts and their credibility as sources, also approached statistical significance, and that area also had received attention in the source-credibility component of the critical thinking seminar. In contrast, little difference in scores appeared in the case of Paragraphs 1 and 7, which both concerned inappropriate definitions, an area not dealt with in the course. Furthermore, there was little difference in scores on Paragraph 4, which consists in circular reasoning and was very similar to one of the sample test items. Perhaps because of its similarity, 35 Ss responded correctly to it.

The overall quality of the answers of the two groups differed, but they shared certain tendencies indicating a general weakness in the area

of critical thinking skills. This is not surprising in view of the fact that Japanese education does not seem to encourage debate or the critical evaluation of reasoning (Davidson, 1995). Detailed consideration of the answers themselves is beyond the scope of this study, but it is revealing to explore the kinds of errors consistently made by the participants. All of the subjects had been taught to identify and use definition, illustration, and argumentation as rhetorical modes; however, this training apparently did not prepare them to recognize reasoning errors related to these modes. For example, in the case of the 24 Ss who positively but incorrectly evaluated Paragraph 2, all gave as their justification the fact that Raywift provided a reason grounded in reality or else that he gave a concrete example. They missed the fact that both the reason and the example were irrelevant to his argument. Similarly, students accepted the definition-arguments in Paragraphs 1 and 7, even though the definitions offered by Raywift were inappropriate. For instance, he argues in Paragraph 7 that his opponents "don't know what 'safe' really means. *Conditions are not safe if there's even the slightest possible chance for an accident*" [italics added] (Ennis & Weir, 1985, p. 13). Only 2 Ss found fault with this impractical definition of the concept of safety; the others credited him with giving a clear definition. Likewise, 25 of 36 Ss accepted the false analogy used in Paragraph 1. Though the treatment group fared better on some paragraphs and in their overall scores, these common tendencies seem to point to a general need for critical thinking training among these particular Japanese EFL students that perhaps is not being addressed adequately by practice in English rhetorical modes or content-based study. It is even possible that exposure to rhetorical modes such as definition, illustration, and argumentation may only predispose students to accepting weak ideas simply because they are presented in the proper rhetorical format. Without concurrent attention to reasoning fallacies and the pitfalls related to each mode, teachers may discover that for their EFL/ESL students, a little bit of knowledge of rhetorical modes is a dangerous thing. Such students may one day find themselves struggling with the reasoning tasks required in an English academic setting, regardless of their general English language proficiency or familiarity with English modes of expression.

Though Hatcher (1994) speculated that politeness and a hesitancy to make negative judgments may have inhibited Japanese and Chinese performance on the Ennis-Weir test at Baker University, answers in our study reveal that they did not err only in positively evaluating weak paragraphs. They also often negatively assessed Raywift's better arguments in Paragraphs 3 and 8. For example, a number of Ss rejected

Raywift's citation of qualified experts in Paragraph 8 as "just opinions." Ss did not appear to suffer from any hesitancy to criticize.

This is only a limited pilot study, and more research of a similar type needs to be done to substantiate these tentative conclusions. Larger student samples are needed. Also, it would be helpful if a translated version of the test could be administered to groups of similar Japanese students to remove completely the possibility that English language deficiencies may to some extent account for the lower scores. For cultural and linguistic reasons, however, such a translated test may be difficult to make and administer. Students in mixed-nationality EFL/ESL programs in other cultural settings could also provide interesting and relevant data about critical thinking abilities and the possibility of developing and testing them in English language programs, since English language-learning problems related to thinking are not confined to Japan. English instructors in other places have noted reasoning weaknesses similar to the ones we have found (Sherman, 1992; Matthews, 1994). Furthermore, it would be informative to experiment with other standard tests of critical thinking in EFL/ESL programs. Finally, it is worth exploring the question of whether training in critical thinking can improve general English language proficiency, especially in writing and reading. Nevertheless, we hope to see the Ennis-Weir test applied by others in studies bearing some similarity to ours. This relatively unexplored area invites further inquiry.

Bruce W. Davidson teaches English at Hokusei Gakuen University in the Department of Social Welfare and researches issues related to critical thinking education. Send correspondence to: z00134@hokusei.ac.jp.

Rodney A. Dunham teaches in the English Department at Tezukayama College and conducts research in mass communication.

Notes

1. Earlier versions of this paper were presented November 5, 1995 at the Twenty-First Annual JALT (Japan Association for Language Teaching) International Conference, Nagoya, and July 29, 1996 at the Sixteenth International Conference on Critical Thinking and Educational Reform, Rohnert Park, CA.
2. The treatment group consisted of 17 out of 22 members from the seminar on critical thinking. Five could not take the Ennis-Weir test due to circumstances arising from the Kobe-Osaka earthquake.

References

- Al-arishi, A. (1994). Practical ways to promote reflection in the EFL/ESL classroom. *English Teaching Forum*, 32(2), 2-5.

- Babbie, E. (1983). *The practice of social research* (3rd ed.). Belmont, CA: Wadsworth.
- Beyer, B. (1991). Practical strategies for the direct teaching of thinking skills. In A. Costa (Ed.), *Developing minds: A resource book for teaching thinking: Vol. 1* (2nd ed., pp. 274-279). Alexandria, VA: Association for Supervision and Curriculum Development.
- Carrell, P. (1987). Content and formal schemata in ESL reading. *TESOL Quarterly*, 21, 461-481.
- Chaffee, J. (1990). *Thinking critically*. New York: Houghton-Mifflin.
- Chamot, A. (1995). Creating a community of thinkers in the ESL/EFL classroom. *TESOL Matters*, 5(5), 1, 16.
- Chance (1986). *Thinking in the classroom: A survey of programs*. New York: Teachers College Press, Columbia University.
- Damer, E. (1995). *Attacking faulty reasoning: A practical guide to fallacy-free arguments*. Belmont: Wadsworth.
- Davidson, B. (1994). Critical thinking: A perspective and prescriptions for language teachers. *The Language Teacher*, 18(4), 20-26.
- Davidson, B. (1995). Critical thinking education faces the challenge of Japan. *Inquiry: Critical Thinking Across the Disciplines*, 14(3), 41-53.
- Engel, M. (1994). *With good reason: An introduction to informal fallacies*. New York: St. Martin's Press.
- Ennis, R., Millman, J., & Tomko, T. (1985). *Cornell Critical Thinking Tests Level X & level Z* [manual]. Pacific Grove, CA: Critical Thinking Press and Software.
- Ennis, R., & Weir, E. (1985). *The Ennis-Weir Critical Thinking Essay Test*. Pacific Grove, CA: Critical Thinking Press and Software.
- Halpern, D. (1993). Assessing the effectiveness of critical-thinking instruction. *JGE: The Journal of General Education*, 42(4), 238-254.
- Hatcher, D. (1994, August). *Grading the Ennis-Weir Critical Thinking Essay Test*. Paper presented at the Fourteenth International Conference on Critical Thinking and Educational Reform, Rohnert Park, CA.
- Hatcher, D. (1996). Combining critical thinking and written composition: The whole is greater than the sum of the parts. *Inquiry: Critical Thinking Across the Disciplines*, 15(2), 20-36.
- Lippman, M. (1991). *Thinking in education*. Cambridge: Cambridge University Press.
- Matthews, C. (1994, March). *Using news stories to develop thinking/reasoning skills*. Paper presented at the TESOL annual convention, Baltimore.
- Norris, S. & Ennis, R. (1989). *Evaluating critical thinking*. Pacific Grove, CA: Critical Thinking Press and Software.
- Paul, R. W. (1992). *Critical thinking: What every person needs to survive in a rapidly changing world*. Santa Rosa, CA: Foundation for Critical Thinking.
- Scriven, M. (1976). *Reasoning*. New York: McGraw-Hill.
- Sherman, J. (1992). Your own thoughts in your own words. *ELT Journal*, 46(2), 190-198.
- Siegel, H. (1988). *Educating reason: Rationality, critical thinking, and educa-*

- tion. New York: Routledge Press.
- Snow, M. & Brinton, D. (1988). Content-based language instruction: Investigating the effectiveness of the adjunct model. *TESOL Quarterly*, 22, 533-574.
- Tarvin, W., & Al-arishi, A. (1991). Rethinking communicative language teaching: Reflection and the EFL classroom. *TESOL Quarterly*, 25, 9-27.
- Walsh, D., & Paul, R. (n.d.). *The goal of critical thinking: From educational ideal to educational reality*. American Federation of Teachers Educational Issues Department.

(Received July 3, 1996; revised Nov. 20, 1996)

Appendix: The Moorburg Letter (Introduction and paragraphs 3, 6, and 8)

Dear Editor:

Overnight parking on all streets in Moorburg should be eliminated. To achieve this goal, parking should be prohibited from 2 a.m. to 6 a.m. There are a number of reasons why any intelligent citizen should agree.

3. Traffic on some streets is also bad in the morning when factory workers are on their way to the 6 a.m. shift. If there were no cars parked on these streets between 2 a.m. and 6 a.m., then there would be more room for this traffic.

6. Last month, the Chief of Police, Burgess Jones, ran an experiment which proves that parking should be prohibited from 2 a.m. to 6 a.m. On one of our busiest streets, Marquand Avenue, he placed experimental signs for one day. The signs prohibited parking from 2 a.m. to 6 a.m. During the four-hour period, there was *not one accident* [italics added] on Marquand. Everyone knows, of course, that there have been over four hundred accidents on Marquand during the past year.

8. Finally, let me point out that the Director of the National Traffic Safety Council, Kenneth O. Taylor, has strongly recommended that overnight street parking be prevented on busy streets in cities the size of Moorburg. The National Association of Police Chiefs has made the same recommendation. Both suggest that prohibiting parking from 2 a.m. to 6 a.m. is the best way to prevent overnight parking.

Sincerely,
Robert R. Raywift

(Ennis & Weir, 1985, p. 13)