

THE ENNIS-WEIR CRITICAL THINKING ESSAY TEST

TEST•MANUAL•
CRITERIA•SCORING SHEET

AN INSTRUMENT FOR
TEACHING AND TESTING

ROBERT H. ENNIS
ERIC WEIR

MIDWEST PUBLICATIONS

**THE ENNIS-WEIR CRITICAL THINKING
ESSAY TEST
TEST • MANUAL • CRITERIA • SCORING SHEET
AN INSTRUMENT FOR TEACHING AND TESTING**

by
**ROBERT H. ENNIS
ERIC WEIR**

85 86 87 88 89 90 91 — 9 8 7 6 5 4 3 2 1 .

© 1985
MIDWEST PUBLICATIONS
P.O. BOX 448
PACIFIC GROVE, CA 93950
ISBN 0-89455-290-2

CONTENTS

Introduction	1
Part I: Information about the Test as a Test	
What the Test Measures	1
Characteristics of the Test	2
Uses of the Test	2
Level of Students	2
Administering the Test	2
Grading the Test	3
Validity	3
Reliability	3
Information about Some Students	4
Part II: Comments on the Argument of the Letter and Suggestions for Scoring	
Paragraph One	4
Paragraph Two	4
Paragraph Three	5
Paragraph Four	5
Paragraph Five	6
Paragraph Six	7
Paragraph Seven	8
Paragraph Eight	8
Paragraph Nine	9
Overall	10
Part III: Using the Test as a Teaching Instrument	
Prerequisite Study	10
Socratic and Didactic Approaches to Teaching	10
Summary	11
Bibliography	11
Appendix: The Ennis-Weir Critical Thinking Essay Test: An Instrument for Teaching and Testing	
Directions	12
The Moorburg Letter	13
Criteria and Scoring Sheet	14
Table 1: Means and Standard Deviations	4

INTRODUCTION

Although originally conceived as a critical thinking test, *The Ennis-Weir Critical Thinking Essay Test: An Instrument for Testing and Teaching* (henceforth "*The Ennis-Weir*") can also be used as the primary teaching material in a very short course in critical thinking or as an integral part of a longer course intended to teach critical thinking. In this booklet, we shall first present it as a test, discussing its qualities and the task of grading it. Then we shall consider its use as a teaching device. It can of course be used both ways, first as a test, then as a teaching device.

PART I: INFORMATION ABOUT THE TEST AS A TEST WHAT THE TEST MEASURES

The Ennis-Weir is a general test of critical thinking ability in the context of argumentation. This type of context is one in which someone is trying to defend a point, and in which the defense is usually preceded and succeeded by other argumentation on the point or aspects of it. In this test, a complex argument is presented to the test taker, who is asked to formulate another complex argument in response to the first. The test is intended to help evaluate a person's ability to appraise an argument and to formulate in writing an argument in response, thus recognizing a creative dimension in critical thinking ability.¹

Even though an organizational structure is imposed on the test taker, the test is an open-ended test. Hence it is not possible to provide a fine-grained analysis of the test in accord with a prespecified list of aspects of critical thinking ability. Instead we can roughly indicate some major areas of critical thinking competence that the test covers. Although the logical and psychological dimensions of critical thinking are not completely separable, this test with its scoring system emphasizes

the logical dimension of critical thinking (using "logical" broadly).

Here is a rough, somewhat overlapping list of areas of critical thinking competence covered by *The Ennis-Weir*:

- Getting the Point
- Seeing the Reasons and Assumptions
- Stating One's Point
- Offering Good Reasons
- Seeing Other Possibilities (Including Other Possible Explanations)
- Responding Appropriately to and/or Avoiding:
 - Equivocation
 - Irrelevance
 - Circularity
 - Reversal of an If-Then (or Other Conditional) Relationship
 - The Straw Person Fallacy
 - Overgeneralization
 - Excessive Skepticism
 - Credibility Problems
 - The Use of Emotive Language to Persuade

This is not a test of formal or deductive argument, nor does it require technical knowledge of such. In comparison with arguments considered in many deductive logic tests, arguments in the real world require considerable interpretation (in context), require evaluation of content as well as form, often have value dimensions, and do not have mechanical decision procedures.² This is a real-world test.

In order to accommodate the importance of context, a context that is familiar to many—a parking problem—has been provided. But for test takers who are not familiar with parking and driving, rush hours, etc., there will no doubt be unforeseen difficulties, invalidating the test to at least some degree for such people.

Although *The Ennis-Weir's* open-endedness and content specificity make it difficult to detail what it tests for, it has the corresponding virtues of having the test taker

¹Some readers of this manual will note the extension of "critical thinking" beyond previous delineations by Ennis (1962, 1980, 1981). This shift is primarily to accommodate contemporary usage.

²The informal dimensions of reasoning have come under increasing recognition recently in logic courses and textbooks, and by the philosophers who teach and write them. For representative examples of this trend, see Blair & Johnson (1977), Hitchcock (1983), Schwartz (1980), and Scriven (1976).

develop an argument in a realistic context while providing for reliability of grading. The prospective user must decide whether the advantages of realism and controlled qualitative evaluation of the creative side of critical thinking outweigh the disadvantages of lack of precision in description of the construct being evaluated and the extra amount of time and expertise required for grading.

CHARACTERISTICS OF THE TEST

The Ennis-Weir is an essay test of critical thinking ability. It takes the form of a letter to the editor of a fictional newspaper. In the letter, the writer makes a proposal and offers a variety of arguments in support of it. Each argument appears in a separate numbered paragraph. There are eight paragraphs in all; each paragraph exemplifies at least one of the errors or types of reasoning listed in the previous section. Two of the paragraphs give probable support to the writer's proposal. Examinees read the letter and then write an essay evaluating the argument of each paragraph and the letter as a whole.

The form of the test is intended to minimize the artificiality of the testing situation. While it is somewhat contrived, in that an effort has been made to build examples of each of a set of important errors and types of reasoning into the letter, the test does reproduce in a fairly realistic fashion a common situation in which skill at critical thinking ought to manifest itself.

Graders are provided with a set of suggested criteria and scoring instructions. They are encouraged to use judgment in applying the criteria, and to add or subtract points for unspecified insights or errors. The manual contains a detailed discussion, paragraph by paragraph, of the argument of the letter and the criticisms that might be directed against it. The intent of this discussion is to sensitize graders to the relevant aspects of the variety of responses that examinees may produce and to secure greater uniformity in assigning scores.

USES OF THE TEST

As a test *The Ennis-Weir* has both instructional and research uses. It can be used as a diagnostic device to identify specific areas of reasoning or argumentation with which groups of students may need help. Furthermore, the test can be used as a device for evaluating effectiveness of instruction in informal logic, critical thinking, or reasoning. However, since there are only a few sets of group scores available now, it cannot easily be used to compare a tested group with group norms.

For research purposes, the test could be used as a basis for comparing control groups and experimental groups in an experimental study. One might want to investigate, for example, the effects of instruction in informal logic, science, social studies, or literature on critical thinking ability. Finally, the test could be used in an exploratory pretest-posttest design, providing educated guesses about the effects of a "specific" curriculum.

LEVEL OF STUDENTS

The Ennis-Weir is probably most appropriate for use with high school and college students. However, we have checked its suitability for use with junior high school and sixth-grade students. The students were able to understand the directions without difficulty, and all were able to paraphrase the contents of the letter accurately. In addition, the students enjoyed the exercise.

ADMINISTERING THE TEST

Instructions to examinees are simple and explicit. If the test is being given to a group, it may help to read the directions aloud. It is permissible to answer questions about the directions for taking the test, but in doing so it should not be necessary to refer to anything other than the directions that are supplied with each copy of the test.

The test takes about 40 minutes—10

minutes to read and think about the letter, and 30 minutes to write the nine paragraphs evaluating the argument of the letter.

GRADING THE TEST

The criteria and scoring suggestions for grading the test are to be used flexibly and with judgment. The reason for this is not lack of confidence in the criteria, but rather that adequate responses may be expressed in different ways (even when they are making the same point), and that critical thinking is an open-ended activity. Graders must use their own judgment interpreting responses and determining whether they satisfy the relevant criteria. The second part of the manual contains a detailed discussion, paragraph by paragraph, of the problems of the argument of the letter. This should help graders make reasonably sophisticated judgments.

Two special precautions should be observed in grading the test. The first has to do with the fact that this is primarily a test of critical thinking ability, not writing ability. One should focus on the quality of thinking in the written responses, rather than on mode of expression. One should understand what the examinee has written and whether it does or does not satisfy the criteria, but be generous in doing so. If it is reasonable to judge that the respondent understood a particular problem in the letter, or that he or she intended to make a certain point, one should give credit for doing so. One should not withhold credit because a point could have been expressed more clearly or succinctly.

The second precaution concerns the handling of logical jargon in the written responses. It should not be used superficially. Then technical terms from logic appear in a response, graders should satisfy themselves that the respondent is using them in an appropriate fashion. Statements containing technical terms should make sense, and the point made should be relevant to the argument being evaluated. The problem of superficial use of technical terminology is especially likely to arise when the test is used to evaluate student learning in logic courses.

This may be because the student has not understood the principles underlying the technical terminology.

VALIDITY

In the testing literature, one commonly finds discussion of four types of validity: content, predictive, concurrent, and construct. The type of validity most appropriately claimed for *The Ennis-Weir* is content validity (in the old-fashioned sense).³ The situation that the test presents to examinees is a common type of situation in which skill at appraising and formulating arguments is manifested, and the problems presented provide opportunities for assessing the important areas of critical thinking competence. We have attempted in Part II of the manual to direct the user's attention to the relevant aspects of examinee responses. The aspects of critical thinking competence we believe to be important have been roughly characterized earlier and have been analyzed in more detail in two papers by Ennis (1962, 1980, 1982). Those who are interested in the specific rationale for including the particular problems built into the test should consult these papers.

Predictive and concurrent validity have not been examined, since there is no established outside criterion for the ability the test was designed to measure. Whether a person is skillful at appraising and formulating arguments is a matter of judgment, based on a variety of indicators. While questions of construct validity, a fourth type of validity, are certainly appropriate for a test purporting to measure critical thinking ability, claims about the construct validity of *The Ennis-Weir* are premature.

RELIABILITY

Essays written by 27 students midway through a college-level introductory informal

³We are not claiming that the test provides a "representative sample" of behaviors constituting the "performance domain" of critical thinking ability (APA, 1974, 28-29). Tomko (1981) has given us reason for not employing the methodology or language of criterion-referenced testing, which has crept into contemporary discussions of content validity. This reluctance exists especially when dealing with an exceedingly complex trait like critical thinking ability. It is our opinion that there is an ineliminable element of judgment by experts in the field—after careful consideration—in assessing content validity.

logic course and 28 gifted eighth-grade students of English were each graded by two different graders (see below for means and standard deviations). Inter-rater reliabilities of .86 and .82, respectively, were obtained. These are high correlations for an essay test.⁴

**INFORMATION ABOUT
SOME STUDENTS**

Table 1 below contains data and group descriptions for two sets of essays. We seek further information of this sort for future editions of this manual.

**PART II: COMMENTS ON THE
ARGUMENT OF THE LETTER AND
SUGGESTIONS FOR SCORING**

In the following discussion of the argument of the letter and its scoring, the grader should interpret instructions for assigning and removing points as guides to be tempered by the grader's judgment. The test and scoring sheet are in the appendix.

PARAGRAPH ONE

The argument of this paragraph is a weak one. The analogy between parking overnight on the streets and having a garage in the streets is not very plausible. A related way of putting this criticism would be to say that an unconventional or incorrect meaning has been offered for the word "garage." Pointing

out specific differences between parking in the streets and having a garage is a stronger criticism than merely claiming an inappropriate analogy or definition, but all are worth three points.

A more sophisticated criticism is that there is an equivocation or shift in meaning in the use of "garage" in the argument. In the first sentence, it means simply "a place to park." But in the law referred to, it most likely means (we may assume) a structure. Pointing out that the conclusion depends on this equivocation or shift in meaning is a sophisticated criticism, also worth three points.

In the light of these faults, the letter writer's failure to say where people would park their cars at night if they did not park them in the streets is a comparatively unimportant defect of the argument of Paragraph One.

It is conceivable, though unlikely, that a respondent might argue effectively that there are important or relevant similarities between parking in the streets and having a garage in the streets (for example, occupying land). Because the ways in which they are similar are, presumably, not against the law, only partial credit (up to two points) should be given to someone defending this aspect of the paragraph's argument.

PARAGRAPH TWO

The defect of this paragraph is obvious: prohibiting parking on the streets at night will not relieve traffic congestion in the afternoon

Table 1. Means and Standard Deviations

Group	1	2
N	27	28
Mean	23.8	18.6
S.D.	4.0	5.9

- 1. Undergraduates in an introductory informal logic course.
- 2. Gifted eighth-grade students of English in a suburban Chicago-area school system. Students had received some critical thinking instruction.

⁴Discussing the level of reliability that can be achieved with essay examinations, Coffman reports finding rating coefficients ranging from .35 to .98 in a survey of the research literature on the subject. In an experimental test

of the reliability of the essay portion of an Advanced Placement Examination in American History, for example, a correlation of .74 was obtained between two different ratings of 200 essays by several different raters (Coffman, 1971).

But respondents should do more than just say that an argument is defective—they should identify the specific defect.

Most respondents will deserve full credit for their responses to this paragraph, since the error is fairly obvious. Finer distinctions can be made by taking off credit for bad judgment or for introducing irrelevant material into a response that essentially contains a correct evaluation of the argument. For example, a respondent might accuse the writer of being biased since he obviously wants to get home in 10 minutes rather than 35. Merely having wants relevant to the conclusion does not necessarily bias one's argument; only if the wants interfered with the reasoning would his argument be biased. So, not only is the charge of bias in the argument relatively unimportant, in this case it seems to be mistaken.

Respondents may be misled by the obviousness and simplicity of the defect of this paragraph. They may be motivated to attribute defects that it does not have. If they show bad judgment in this, this fact should not be overlooked. Take off credit. Criticism should point out the real or important defects of an argument; it should not strain to find unimportant problems or to create problems that are not actually in the argument.

PARAGRAPH THREE

The argument of this paragraph is strong—for the streets mentioned. People on their way to work the 6 a.m. shift would be on the streets during hours Raywift proposes that parking be prohibited, and if there are no cars parked on the streets the flow of traffic will be eased. However, the argument does not fully support the specific proposal being argued for, which is to ban parking on all city streets from 2 a.m. to 6 a.m. The problem of crowded streets could be remedied by prohibiting parking on just those streets that are crowded with factory workers trying to make the 6 a.m. shift. If this limitation is not mentioned here, however, do not take off points. Do so for the summary paragraph if the limitation is not mentioned anywhere.

An example of bad judgment in criticizing

this paragraph is the accusation that “some” or “bad” are vague terms. They are undoubtedly vague terms, but (provided the limitation of the argument to “some streets” is noted) their vagueness does not seem to interfere with the strength of the argument. Unless an advantage of making them more precise is shown, such criticisms should count as bad judgment, and one point should be removed.

PARAGRAPH FOUR

As an argument for the proposition advanced by the writer, this paragraph accomplishes nothing. It cites as a reason for being opposed to overnight parking the (alleged) fact that overnight parking is undesirable. Holding something to be undesirable is not exactly the same thing as being opposed to it, but it is very nearly so; barring indications to the contrary, it can be assumed that anyone who finds a thing undesirable is opposed to it and vice versa. In other words, in this paragraph the writer has merely offered the proposition he is arguing for as a reason for itself. The argument is circular.

Another way of describing the defect of the argument of Paragraph Four is to say that it offers no additional reasons in support of the proposition being argued for.

Few respondents are likely to fail to notice the defects of the argument of this paragraph. They may, however, be tempted to attribute defects to it that it does not have. In particular, it may be asserted that the writer has simply stated his own personal opinion or preference. There is nothing wrong with stating one's personal opinions or preferences in an argument. It is to be expected.

Another mistaken objection is claiming that Raywift has not shown that overnight parking is not desired by the residents of Moorburg. There is a difference between claiming that something is undesirable and claiming that it is not desired. Raywift has not, on the face of it, made any claim about what is not desired by the residents; he has made a claim about what is not desirable. Respondents should be penalized one point for accusing Raywift of a

fault in arguing he did not commit.

PARAGRAPH FIVE

The argument of this paragraph is not very strong, though it is the most complex of any in the letter. There are several points on which the argument can be criticized, and different respondents will either notice or give higher priority to different defects. As a result, graders might vary in their assessments of responses to this paragraph. This problem can be alleviated somewhat if graders are aware of, and give appropriate consideration to, any challenged defect of the argument.

There are three general types of defect in the argument. The first derives from the fact that the type of accidents that would be eliminated if the recommendation were adopted is a very special and restricted type—accidents between vehicles parked on the streets overnight (or more narrowly, between the hours of 2 a.m. and 6 a.m.) and moving vehicles. Put another way, there are other, more significant types of accidents. For example, there are accidents between moving vehicles, and between moving vehicles and vehicles parked on the streets at other hours.

Another way of putting this defect is that the number of moving vehicles on the streets during the hours when parking is to be prohibited is not likely to be very great. Therefore, the number of accidents eliminated, even of this special, restricted type, is not likely to be very great. A criticism closely related to this one is that no evidence is given in Paragraph Five (or anywhere else in the letter) that accidents of the kind that would be eliminated actually do occur (or that they occur in significant numbers).

Noting this first type of defect is a full-credit criticism.

A second type of defect has to do with the specifically evaluative aspects of the argument. It is quite possible, for example, that other things might be judged more important than eliminating accidents of the type referred to in the argument. The inconvenience and economic costs to residents and others resulting from being unable to leave

their cars on the street overnight is a consideration that is neglected by the argument. Similarly, many persons might judge the class of accidents that would be eliminated if the recommendation were adopted to be relatively unimportant compared to those that would remain unaffected.

Finally, there is a defect that arises from the form of the argument. There are several ways in which this type of defect might be described. In ordinary, nontechnical language, the defect is that the argument describes only one possible way of eliminating accidents of the type it claims would be eliminated. Since there may be other ways of eliminating such accidents, it is not incumbent even on those who agree that it is important to eliminate them to support adoption of this specific remedy. In logical terms, the letter writer has shown the recommendation to be a probable sufficient condition for eliminating one kind of accident, but not a necessary condition. Another way of putting this is to say that (roughly speaking) the argument commits the fallacy of affirming the consequent: from the fact that accidents will be prevented if parking is eliminated, and the desirability of eliminating accidents, it does not necessarily follow that parking should be prohibited; there may be other ways of eliminating accidents. Noticing this defect, however it is described, indicates greater sophistication than noticing the other types of defect.

Each of these three types of defect is an “other-possibilities” defect. That is, in each case there is another reasonable possibility. So Paragraph Five is an other-possibility paragraph.

Although the argument of this paragraph is properly regarded as not very strong, it is possible for a response that judges the argument to be good to deserve full credit. This will generally be when the respondent qualifies the judgment that the argument is good by taking note of any of the features we have just identified as defects. For example, a respondent might say, “The argument gives a reason for prohibiting parking, but probably few accidents would actually be prevented, and it

isn't shown that we should necessarily be in favor of the proposal." It is important to preserve the possibility of judging the argument to be good because many good arguments will have the same structure as this weak one. For example, if there was general agreement that some outcome should be avoided, and if other ways of avoiding it were either unavailable or undesirable, then an argument of this form would provide a reason for supporting or adopting the means referred to in the antecedent of the argument.

PARAGRAPH SIX

Paragraph Six can be regarded as an other-possible-explanation paragraph. The most important defect of the argument is that the results of the one-day experiment do not adequately support the causal claims implicit in it. The argument implies that the lack of accidents in the four-hour period the day of the experiment was due to the installation of no-parking signs. It also suggests that, if parking were prohibited on other streets, accidents would again be prevented. There are, however, a variety of plausible alternative explanations for the lack of accidents. The existence of these other possible explanations also undermines the generalization of the experimental results. Since the argument does not present sufficient information to rule out these other explanations, it does not provide adequate support for the claims being argued for.

For example, it might be claimed that the lack of accidents during the period parking was prohibited could easily have been simply a chance occurrence and not really due to the parking ban itself. The fact that there were no accidents during such a brief period would not be a surprising occurrence. The inference that installing the no-parking signs was effective in eliminating accidents would have been stronger if observation had been extended over a significantly longer period.

Another possibility is that the day on which the experiment was conducted could easily

have been atypical in some way. We are not told that the experiment was conducted on a normal work day. For all we know, it was conducted on a weekend or holiday, or perhaps there was a bad snowstorm that day. If any of these propositions were true, the reduced volume of traffic the day of the experiment would be a plausible alternate explanation of the results. Still another potential explanation arises from the possibility that there were no accidents on the streets where signs were not installed. If that were the case, not only would it be reasonable to suspect that something other than the no-parking signs was responsible for the lack of accidents on Marquand, we would also have reason to doubt that extending the parking prohibition to other streets would eliminate any more accidents.

A final alternative arises from the fact that we are not told how many of the more than 400 accidents occurring on Marquand involved parked vehicles. If few of them did, we would have to seek an explanation for the lack of accidents in some other factor, since the parking prohibition would probably not be responsible for eliminating accidents that did not involve parked vehicles in the first place. For the same reason, it would be unreasonable to expect that prohibiting parking on other streets would prevent accidents.

It is important to note that these are all explanations that are reasonable to propose. There are many other explanations that could conceivably be offered, but they would not necessarily constitute valid criticisms of the experiment or the conclusions drawn from it. For example, it is conceivable to suggest that alien beings hovering in a nearby spaceship intervened in some way to prevent the accidents. But clearly we would not take such a suggestion seriously. Only when the alternative explanation is a reasonable one to propose does it constitute a significant criticism of the experiment.

Some respondents may fault the argument for its use of the expression "everyone knows, of course," on the ground that this is an attempt to exercise unwarranted influence on

the reader. This is a weak criticism and should be faulted for exhibiting bad judgment. If there were this many accidents in one year, it would not be unreasonable to assume that most people knew about it. And in any case, it is an easily checked factual claim. It does not appear that Raywift has attempted to gain unwarranted assent in asserting that "everyone knows."

An adequate response to the paragraph would judge the argument to be weak and would indicate in some way that the reason for this is the tenuousness of the implied causal claims. A criticism that is justified and is at least fairly specific receives full credit. For example, even if the response says simply, "The experiment does not prove that prohibiting parking caused the lack of accidents," it should be given full credit. As the criteria indicate, the same fundamental criticism may take a variety of forms and be expressed in many different ways.

A problem that might not be noticed by unsophisticated respondents is that the claim being argued for, a value claim (i.e., that parking should be prohibited) has not been adequately supported, the problems about causation aside. When this problem is noticed, the appropriate criticism would be similar to those applied to Paragraph Five (e.g., that the inconvenience of the parking ban would be too great, or that there are other ways of preventing accidents). Respondents who make such a criticism should not be penalized for not mentioning the problems about causation. They should receive full credit.

PARAGRAPH SEVEN

The defect in Paragraph Seven can be put several different ways. It might be claimed that the definition that is stipulated is simply incorrect, that this is not what "safe" really means. Another way of challenging the defect is that the proposed definition actually renders the word useless, since not every chance of an accident can be eliminated. Not only would present conditions be unsafe if the proposed definition were accepted, they would

remain so even if Raywift's proposal were adopted. Thus his definition, though framed to suit his purpose, is actually self-defeating. A third way is to note that the meaning of the word "safe" has been shifted in mid-argument, making the argument a case of equivocation.

An adequate response will at least judge the argument weak. If the justification is that the definition is incorrect, the response should be given three points. This is a reasonable criticism, though not as incisive as pointing out that the word has been rendered useless, or that equivocation has occurred. Either of the latter criticisms are also worth three points.

An example of poor judgment in justifying a correct evaluation of the argument would be a claim that Raywift's definition is vague or that it is unclear what he means by "safe." His definition is very clear; it just cannot actually be satisfied. This criticism should cost the test taker one point.

If the respondent correctly judges the argument, but justifies the judgment only by claiming that Raywift has "slandered" his opponents in accusing them of "not knowing what safe really means," give credit only for the correct judgment (one point). While Raywift's claim about his opponents' knowledge of the meaning of the word may be intended to have rhetorical effect, this is a trivial criticism compared with those mentioned above.

PARAGRAPH EIGHT

The argument of this paragraph is one of the better ones in the letter. The authorities cited could reasonably be expected to be knowledgeable about the subject being discussed. Their recommendation is directly relevant to Raywift's proposal. Further, there seems to be no good reason to doubt their expertise or to doubt that they actually made the recommendation he claims they made. There is, however, a crucial qualification that weakens the support provided for Raywift's proposal: the authorities' recommendation applies only to busy streets. Again, do not take off credit if this limitation is not noted here.

Generally, an adequate response would

judge the argument to be reasonably good. If the citing of these particular authorities is judged to be relevant and appropriate, and to lend force to the argument, the response should be given three points. This judgment should be explicit, however. If the respondent supports a positive judgment only by stating that authorities are cited, without commenting on the relevance or appropriateness of these particular authorities, he or she should be given credit for a semi-adequate justification only (two points). On the other hand, "The author appeals to two different legitimate authorities" or "This is all right, if the authorities are qualified" would be marginally worth three points, because they indicate concern for the authorities' qualifications.

Raywift is not required to give evidence that the authorities actually said what he claims they have said, or to produce the reasons they gave for making their recommendation. Respondents who criticize him for not doing either of these things should be faulted for bad judgment in justifying. The important point is that their recommendation is relevant to his. The fact that they made it can be checked if necessary. If the reasons offered in arguments were generally judged suspect when not themselves positively justified, almost all reasons in real-life arguments would be under suspicion. Claims made in the course of an argument, if they can be easily checked, should generally be granted credibility—unless there are reasons for doubting their truth or relevance.

Some respondents, however, may wonder whether the authorities' recommendation would still appear relevant to Raywift's purpose if their reasons were known; they may point out that his argument would have been stronger if he had indicated their reasons and shown them to be relevant to his purpose. Where possible, this should be distinguished from demanding positive justification for the authorities' recommendation or evidence that they actually made it, and respondents should be given credit for observing that the strength of the argument could be affected by knowledge of the authorities' reasons.

A respondent might judge the argument in Paragraph Eight weak on the ground that it is dangerous to infer from a loose general recommendation to this particular city without knowing that Moorborg is typical of cities this size, or that the facts about Moorborg do not disqualify it from fitting this general recommendation, because Moorborg might well be different from other cities its size. Give full credit for this sort of sensitive skepticism.

Some respondents may judge the argument weak because they take it to be advancing a trivial, tautologous claim. It may be argued that it is "obvious" that the best way to prevent overnight parking is to prohibit parking from 2 a.m. to 6 a.m.—as if the two were equivalent. They are not equivalent: prohibition of parking between 2 a.m. and 6 a.m. is offered as a means of discouraging overnight parking, and Raywift's proposal makes this clear. Take off one point when this criticism is made.

If a respondent rejects the argument as an "appeal to authority," apparently believing that any appeal to authority is fallacious, then give a minus one.

PARAGRAPH NINE

Responses in this paragraph are among the most difficult to rate in the test. What is desired is a judgment of the overall strength of Raywift's argument that gives specific reasons but does not simply recapitulate the responses of the first eight paragraphs.

The minimum requirement of an adequate response is that the argument of the letter be judged faulty (worth one point). It is difficult to imagine that a plausible case could be made for a judgment to the contrary. To receive more than one point, however, the response should do more than just condemn the argument (by merely calling it "fallacious," for example). If, in addition, the response says that six of the eight paragraphs contain faulty arguments, if it correctly identifies the two paragraphs with reasonably good arguments, or if it simply summarizes the judgments made in the preceding paragraphs, give it one more point.

If, and only if, the mistake of inferring from some streets to all streets is mentioned somewhere, even if it is not mentioned here, the respondent should receive two more points.

The use of emotive language in the introductory paragraph ("any intelligent citizen") is an attempt to get people to agree by illicit means. Noting this sort of thing (somewhere) is good for one point (here).

So, five points are available from Paragraph Nine.

OVERALL

The top possible score is 29. Because responses are solicited without offering a great deal of time to think and write, this top score is unlikely, even for accomplished critical thinkers.

Although judgment is required, a trained grader can do these at the rate of one every five or six minutes. Most student responses can be handled without a great deal of thought, but the grader must still remain continuously alert, watching for the idiosyncratic response.

PART III: USING THE TEST AS A TEACHING INSTRUMENT

The Ennis-Weir can be used as a teaching instrument, either as a follow-up to testing or as its primary use. Because each teaching situation is so unique, we can offer here only general advice.

PREREQUISITE STUDY

Before using this test as a teaching instrument, it is essential that the teacher take the test, grade herself or himself (after reading these guidelines), then reread the advice on grading just before launching into a class discussion of the test. Otherwise there is a good chance that the teacher will be caught flat-footed and unready for questions and problems that arise. This is not to assume that one should tell one's students how to respond to the test. But unless one is familiar with the content and its possibilities, one could miss opportunities to ask illuminating questions

and might not understand some of the students' comments.

It is also important for the teacher to have studied critical thinking elsewhere—at least a college-level course in informal logic, critical thinking, or the equivalent. Otherwise there is a danger of being trapped into saying things that might be regretted later. This is so even if a strongly Socratic approach is adopted. Among the many useful texts in this area are those by Blair & Johnson (1977), Hitchcock (1983), Schwartz (1980), and Scriven (1976).

SOCRATIC AND DIDACTIC APPROACHES TO TEACHING

A Socratic teacher does not set out with the goal of conveying a body of content, but rather asks questions, attempts to focus the discussion, reinterprets a student's comments or simply reflects them back to the student or class, wonders what is meant by something, suggests possible implications, keeps the discussion on track, asks "Why?", and exhibits a respect for each student and for a search for the truth—or at least for trying to get things right.

A didactic teacher has in mind a set of goals, a plan for pursuing these goals in some reasonable order, and an idea of what would constitute success in achieving these goals. This is not to say that the Socratic teacher does not have a set of goals; there might be broad goals, some of which the Socratic teacher realizes might never be achieved. And it is not to say that the didactic teacher does not ask questions, or suggest implications, etc.

The Ennis-Weir can be used with either approach to teaching, or a mixture of the two. A Socratic teacher could ask the students to take the test, ask students to grade each other's or their own tests, and invite discussion about how to apply the proposed criteria on the scoring sheet. In an even less structured manner, the Socratic teacher could invite students to react to the letter, either orally or in writing, and to react to each other's reactions (without use of the scoring sheet), again either orally or in writing, ending up with a discus-

sion that leads into areas that students agree need further investigation.

A didactic teacher could ask students to write out a response to the letter, grade the students' responses (or have them grade each other's), explain the grading to the students, diagnose the students' weaknesses, and organize further instruction based upon this diagnosis. In this further instruction the teacher should arrange for students to apply the insights in other situations.

The possible combinations of the two approaches are many. The important thing is to focus students' attention on the issue and the arguments offered; to involve them in an attempt to grasp, state, or develop principles for thinking critically and to apply these principles to significant-appearing issues. The discussion of the paragraphs in Part II states or implies many of the relevant principles. For further elaboration of these principles, see one or more of the texts mentioned earlier, or Ennis' (1962, 1981) discussions of critical thinking.

If a teacher asks students to read about principles of critical thinking, we recommend that they be told (in advance and with examples) the principle they should learn, and otherwise be provided with motivation for their going to that piece of material. This seems superior to telling them later what they were supposed to have read, a procedure that is often insulting. It is also self-defeating since students often then decide to wait to hear from you and don't bother to read at all.

SUMMARY

Since attempting to teach critical thinking is occasionally perilous, we urge the teacher to review the letter and the discussion in this manual just before using *The Ennis-Weir* as a teaching device, and to have done some study of critical thinking, either in a course or independently.

The Socratic and didactic approaches to teaching critical thinking can each be viable, depending on the circumstances, as can combinations of the two. In the Socratic approach,

the students often challenge each other; in the didactic approach, the teacher methodically pursues some clear goals. More specific advice is not possible since so much depends on the specific situation and the opportunities that develop.

BIBLIOGRAPHY

- American Psychological Association. *Standards for educational and psychological tests*. Washington, D.C.: American Psychological Association, 1974.
- Blair, J. Anthony, and Johnson, Ralph. *Logical self-defense*. Toronto: McGraw-Hill-Ryerson, 1977.
- Coffman, William E. Essay examinations. In Robert L. Thorndike (ed.), *Educational measurement* (2nd ed.). Washington, D.C.: American Council on Education, 1971.
- Ennis, Robert H. A concept of critical thinking, *Harvard Educational Review*, 1962 32, 81-111.
- Ennis, Robert H. A conception of rational thinking. In Jerrold Coombs (ed.) *Philosophy of education 1979*. Bloomington, Ill: Philosophy of Education Society, 1980.
- Ennis, Robert H., Rational thinking and educational practice. In Jonas F. Soltiss, (ed.), *Philosophy and education* (Eightieth Yearbook of the National Society for the Study of Education, Vol. 1). Chicago: The National Society for the Study of Education, 1981.
- Hitchcock, David. *Critical thinking: A guide to evaluating information*. Toronto: Methuen, 1983.
- Schwartz, Thomas. *The art of logical reasoning*. New York: Random House, 1980.
- Scriven, Michael. *Reasoning*. New York: McGraw-Hill, 1976.
- Tomko, Thomas N., and Ennis, Robert H. Evaluation of informal logic competence. In J. Anthony Blair & Ralph H. Johnson (eds.), *Informal logic: The first international symposium*. Inverness, Calif. Edgepress, 1980.
- Tomko, Thomas N. The logic of criterion-referenced testing. Unpublished Ph.D. dissertation, University of Illinois at Urbana-Champaign, 1981.

THE ENNIS-WEIR CRITICAL THINKING ESSAY TEST AN INSTRUMENT FOR TESTING AND TEACHING

DIRECTIONS

Read the letter to the editor of the Moorburg newspaper. Consider it paragraph by paragraph and as a total argument. Then write a letter to the editor in response to this one. For each paragraph in the letter you are about to read, write a paragraph in reply telling whether you believe the thinking good or bad. Also write a closing paragraph about the total argument. *Defend your judgments with reasons.*

Your answer should have nine numbered paragraphs. Numbers one through eight should give your reactions to paragraphs one through eight in the letter. Your paragraph number nine should give your overall evalua-

tion of the letter considered as one total argument. Each paragraph, including the last, should contain your reason(s).

Spend about 10 minutes reading the letter and thinking about it. Then write for not more than 30 minutes (about three minutes for each of your short paragraphs). The maximum total time for the test is 40 minutes.

Do not forget to give your reasons in each paragraph. Please write clearly.

Sign your name to your letter. You are a local citizen, and this topic concerns you.

Remember, write *nine numbered paragraphs* and give reasons.

NOTE: Individuals and institutions who have secured this test from MIDWEST PUBLICATIONS are permitted to reproduce the test and the scoring sheet for classroom use only. The test consists of this page of directions and the letter on the next page. For each student a separate scoring sheet (page 14) will be needed for the grader to grade the student's response.

THE MOORBURG LETTER

230 Sycamore Street
Moorburg
April 10

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.

1. For one thing, to park overnight is to have a garage in the streets. Now it is illegal for anyone to have a garage in the city streets. Clearly, then, it should be against the law to park overnight in the streets.

2. Three important streets, Lincoln Avenue, Marquand Avenue, and West Main Street, are very narrow. With cars parked on the streets, there really isn't room for the heavy traffic that passes over them in the afternoon rush hour. When driving home in the afternoon after work, it takes me thirty-five minutes to make a trip that takes ten minutes during the uncrowded time. If there were no cars parked on the side of these streets, they could handle considerably more traffic.

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.

4. Furthermore, there can be no doubt that, in general, overnight parking on the streets is undesirable. It is definitely bad and should be opposed.

5. If parking is prohibited from 2 a.m. to 6 a.m., then accidents between parked and moving vehicles will be nearly eliminated during this period. All intelligent citizens would regard the near elimination of accidents in any period as highly desirable. So, we should be in favor of prohibiting parking from 2 a.m. to 6 a.m.

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** on Marquand. Everyone knows, of course, that there have been over four hundred accidents on Marquand during the past year.

7. The opponents of my suggestions have said that conditions are safe enough now. These people don't know what "safe" really means. **Conditions are not safe if there's even the slightest possible chance for an accident.** That's what "safe" means. So, conditions are not safe the way they are now.

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.

I invite those who disagree, as well as those who agree with me, to react to my letter through the editor of this paper. Let's get this issue out in the open.

Sincerely,

Robert R. Raywift

Student's Name _____ Total Score _____ Graded By _____

CRITERIA AND SCORING SHEET FOR THE ENNIS-WEIR
Robert H. Ennis and Eric Weir

Credit Given
(maximum is 3 points
per line except #9)

See manual for interpretation and qualification of these criteria.

1. Recognition of misuse of analogy, and/or recognition of shift in meaning, and/or claim that incorrect definition has been stipulated.	
2. Recognition of irrelevance.	
3. Recognition that Paragraph Three is OK. (Neglecting the busy-streets limitation is not penalized here.) ^A	
4. Recognition of circularity, and/or recognition that no reason is offered. (Subtract one point from credit for interpreting "undesirable" as "not desired.")	
5. Recognition that there may be other ways of preventing accidents, and/or recognition that other things might be more desirable, and/or recognition that there probably isn't much traffic at that time, and/or recognition that other types of accidents are unaffected, and/or recognition that no evidence has been given that such accidents occur. (Other possibilities)	
6. Recognition of lack of controls, and/or inadequate sampling, and/or "only one case," and/or "post hoc fallacy." (Other possible explanation)	
7. Recognition of winning argument by definition, and/or recognition that a word has been made useless for empirical assertion, and/or claim that an incorrect definition has been asserted.	
8. Recognition that Paragraph Eight is OK. (Neglecting the busy-streets limitation is not penalized here.) ^{A C}	
9. One point for just condemning the overall argument; another point for reviewing or summarizing the responses to the other paragraphs in some reasonable way; two points for recognizing (anywhere) the error of concluding about all streets on the basis of reasons that relate only to busy streets; ^A and one point for noting (anywhere) that Raywift has attempted to push people around with his emotive language. Total possible: 5 points.	

A score of -1, 0, 1, 2, or 3 will be given for each of the first eight numbered paragraphs:^B

- 1 judges incorrectly (good or bad)^C
- 1 shows bad judgment in justifying
- 0 makes no response^D
- +1 judges correctly (good or bad), but does not justify^C
- +2 justifies semi-adequately
- +3 justifies adequately

For Paragraph Nine, the range is -1 to +5.

^ADo not penalize for failure to note busy-streets limitation in Paragraphs Three or Eight. If it is not noted at least somewhere, do not give the allotted 2 points in Paragraph Nine. If the limitation is noted in Paragraphs Three or Eight, credit should be granted at Paragraph Nine.

^BThese criteria are guidelines. The grader should use judgment in awarding points, subtracting for unspecified errors and adding for unspecified insights.

^CSometimes, something judged one way here will be judged another way by the test taker, and so well defended that a positive score (sometimes even +3) is warranted. The grader must use judgment. For example, a good argument could be mounted against Paragraph Eight.

^DIf the examinee makes a response, but the argument of the paragraph is not judged either good or bad and no reasons are given, count it as "no response."