

The Evaluation of Critical Thinking Programs: Dangers and Dogmas

John E. McPeck
University of Western Ontario

For the past few years I have defended a view about the nature of critical thinking (and how to teach it) that runs counter to the dominant view in North America.¹ I have argued that the major ingredient of critical thinking is context-specific, field-dependent **knowledge** and **information**. And, contrary to the received opinion, critical thinking has little (if anything) to do with so-called "general reasoning skills" or the like. My view of critical thinking, moreover, has led me to reject courses in informal logic, and to advocate approaches to critical thinking which attempt to increase one's capacity for understanding complex concepts, information, and problems — as the traditional disciplines try to do. Thus, the differences between the standard approach and my own view are twofold: we disagree over both the ingredients of critical thinking, and how to teach it.

One of the recurrent criticisms leveled against my view is the charge that I am advancing an empirical thesis without presenting data.² For example, Robert Ennis, in rebutting my rejection of informal logic and critical thinking courses, says "He has offered no empirical evidence on this matter ...", and concludes quite unequivocally, "I think that the basic issue is now an empirical one and should be dealt with empirically."³ While I have always thought this claim at the very least to be contentious, I have subsequently discovered quite a few studies which share the belief that the evaluation of critical thinking programs is essentially an empirical question.⁴ It is this claim, then, which provides the motivation and central focus of the present essay. If I cannot get you to share my scepticism about empirically evaluating critical thinking, I hope to at least expose for you some of the serious difficulties that confront such an evaluation.

Burden of Proof

There is a long and respected tradition in philosophy which requires that when someone makes an existence claim, be it for a God, a ghost, or a unicorn, the burden of proof is on the person making the existence claim to justify the belief — the onus is **not** on the doubter to disprove it. This tradition should extend to the evaluation of critical thinking programs. When test-and-measurement enthusiasts talk about **measuring** such things as "critical reasoning ability" or "general reasoning skills," the onus is

on them to be very clear about what they mean by these terms, and to prove that such "general abilities" really exist. Certainly, if someone claimed to be measuring a person's ESP power with an electroencephalograph, for example, we would be properly sceptical about just what was meant by "ESP", and how it was known that ESP was being measured. In both cases, the onus is on the measurer to make the case, not on the doubter to disprove it.

Elsewhere I have argued that the very conception of a "general reasoning ability", or the like, is conceptually incoherent. It is incoherent in the same way that saying being "generally speedy" is incoherent. That is, we do not posit a single skill called "speed" or "general speediness" to an individual because we realize that there are just too many different ways that a person can be slow or speedy (e.g., running, typing, or changing mufflers). No one, to my knowledge, has ever established that there exists anything which might legitimately be called "critical thinking ability" or "general reasoning skill". Even the giants of psychology and psychometrics have come up empty-handed. For example, J.P. Guilford, in his paper entitled "The Nature of the General Reasoning Factor" concludes that:

... a common, unique, psychological core for all problem solving does not exist. Problems are simply too varied, and each type seems to call upon its own pattern of abilities — perceptual abilities as well as thinking abilities...

In conclusion, we may say that it has been much easier to decide what **general reasoning** is not than to say what it is.⁵

In a very recent and lengthy review of the psychological literature on deductive reasoning, published in 1982, Jonathan Evans reports that:

From consideration of the material reviewed in parts I to III of this book, it appears that there is little evidence for the influence of a general system of logical competence, and that the thought processes involved are highly **content dependent** (p. 6)...

Later, he adds:

We are forced to the conclusion that people manifest little ability for general deductive reasoning in these experiments. Very little behaviour can be attributed to an **a priori** system that is independent of the particular task content and structure.

This does not mean that people cannot reason correctly in contexts where they have no relevant and appropriate experience — indeed some evidence suggests that they can. It does mean, however, that adults' reasoning ability is far more concrete and context-dependent than has been generally believed (p. 254).⁶

And David Ausubel, after a lengthy analysis of the research on psychological transfer, concludes straight off:

Hence critical thinking cannot be taught as a generalized ability. In practice it can be enhanced only by adopting a precise, logical, analytic, and critical approach to the teaching of a particular discipline, an approach that fosters appreciation of scientific method in that discipline. Also, from a purely theoretical standpoint alone, it hardly seems plausible that a strategy of inquiry that must necessarily be broad enough to be applicable to a wide range of disciplines and problems can ever have, at the same time, sufficient **particular** relevance to be helpful in the solution of the **specific** problem at hand.⁷

Likewise, Gagné rejects the transfer of general abilities of this new type also, as did Thorndike 35 years before him when he decisively discredited "faculty psychology" and the "mental discipline" approach to education. Save for I.Q., I know of no reputable psychometric researcher who supports the existence of something which might properly be termed "general reasoning ability" or "critical thinking ability". Thus, considered as a general ability, the burden of proof remains, as it does with ESP and UFO's, on the shoulders of the proponents of "critical thinking ability". In the meantime, serious scepticism is surely justified.

Alternatively, when critical thinking is considered as some **set** of abilities, such as those described in the Watson and Glazer test, or in Robert Ennis' "general aspects", the situation is not measurably improved. All of the tests which purport to measure these "abilities" do two things:

(1) they merely **assume** that the phenomena being tested are in fact **useful to or productive of** real critical thinking (i.e.: that the tests have external validity);

and

(2) because the tests postulate certain singular, requisite "abilities" (e.g., "the ability to evaluate evidence", "the ability to recognize underlying assumptions") it is then **assumed** that there **exist** such unitary underlying abilities corresponding to these descriptions.

In the first instance they are **assuming** what needs to be proven (known to fallacy buffs as "begging the question"); and in the second instance they are **reifying the existence** of a pervasive "ability" from its description. Harold Berlak, for example, was absolutely correct in criticizing Robert Ennis's "general aspects" for merely assuming their usefulness. He says:

The value of any set of intellectual skills (Ennis calls them aspects) rests on whether they have demonstrated value to persons who have dealt successfully with some problem or issue. This is the "ultimate" test of any set of intellectual operations ... [At] some point it must be shown that the aspects selected are **of major importance to persons who are attempting to deal with issues or problems**. Certainly, if a reading expert proposed that knowledge of certain aspects of linguistics is important to the learning of reading, we would expect the proposition to be defended by argument and, if possible, with data. Similarly, if knowledge of the aspects of thinking selected by Ennis or anyone else is of major importance to the process of engaging in critical discourse, then we should expect a justification for selection in terms of argument and data. Ennis does not do this, and rarely does anyone else. In most of the writings in this area, the value of operations is assumed to be *prima facie*.⁸

Moreover, these same **assumed** "aspects" by Ennis are what the Cornell Critical Thinking Tests (Level X and Z purport to be measuring. (**Note:** Ennis is the major author of the Cornell tests.)

The widely used **Watson-Glaser Critical Thinking Appraisal** purports to measure five distinct "abilities". Let us look at the sole justification provided for the belief that there are five unitary abilities underlying critical thinking:

Dressel and Mayhew (1954) have listed the following abilities that appear to be related to the concept of critical thinking:

- The ability to define a problem
- The ability to select pertinent information for the solution of a problem
- The ability to recognize stated and unstated assumptions
- The ability to formulate and select relevant and promising hypotheses
- The ability to draw valid conclusions and judge the validity of inferences

Judgments of qualified persons and results of research studies (Houle, 1943; Morse & McCune, 1957) support the author's belief that the items in the **Critical Thinking Appraisal** represent an adequate sample of the above five abilities and that the total score yielded by the test represents a valid estimate of the proficiency of individuals with respect to these aspects of critical thinking.⁹

There you have it. How do Watson-Glaser know that there are true "abilities" at work here? **Answer:** Because they took them from a list provided by Dressel and Mayhew in a Government document. But how do Dressel and Mayhew know that there are "abilities" corresponding to these descriptions? **Answer:** Because they "**appear** to be related to the concept of critical thinking" (**ibid.**). Thus we have one person's "appearance" serving as the next person's "reality", which have subsequently served as the basis of hundreds of "empirical" studies in the area.

We have here in microcosm the chronology of how a casual phrase ("critical thinking abilities") can become a recurrent piece of educational jargon, which is eventually **reified** into a cognitive ability—in this case, a latent trait.

Specifically, what I think has gone wrong in this instance is that educators and measurement-types have mistakenly taken the **description** of an **achievement** as indicative of an **ability**. Notice, for example, that such things as "defining a problem", or "recognizing underlying assumptions", or "correctly evaluating evidence" are all descriptions of **achievements**—in each case something has been successfully accomplished. Notice further that **achievements** do not necessarily **describe** corresponding abilities. For example the statements "He reached the summit of the mountain" and "He crossed the finish-line" both describe achievements, but in neither case do you know **how** it was done. The summit could have been reached by helicopter, or the finish-line could have been sailed across, walked across, or driven across. In neither case do we know what actual "abilities" were involved in the achievement. Similarly for such achievements as "defining a problem" or "correctly evaluating evidence", one cannot assume that a unitary "ability" is indicated, nor be certain what that "ability" might be like. In such cases, literally hundreds of separate **abilities** might have been involved, or conversely, nothing recognizable as an ability might have been involved. Thus, despite the prevalent jargon, there are insufficient grounds for believing both that such abilities actually exist, or that standardized tests are measuring them. To repeat, the burden of proof remains with the claimant in this case, not with the sceptic.

The Definition of "Critical Thinking" and Empiricism

Yet another obstacle in the path of measuring the effectiveness of various critical thinking programs is that different

definitions of "critical thinking" will require different criteria of measurement. That is, for different meanings of "critical thinking," different kinds of behaviour will count as evidence for it. Thus tests of critical thinking are not empirically neutral, but are decidedly theory-laden with their own specific notions of 'critical thinking'. When Robert Ennis and others assert that "the issue is now an empirical one and should be dealt with empirically",¹⁰ it's not at all clear how this can be fairly done. Where there are competing conceptions of critical thinking it is unlikely that any neutral test can arbitrate among them.

Indeed, the evaluation of critical thinking programs is not unlike the difficulty of evaluating "therapy" for some neurosis in psychiatry. The problem is exacerbated when you have therapists from different theoretical orientations. For example, a Freudian and a radical behaviourist cannot even agree on the **type** of evidence which should count as relevant to such an evaluation. The Freudian is likely to require only that the patient integrate the neurotic behaviour into his personality so that he is no longer troubled or uncomfortable with the problem. Thus, when trauma and discord have been reduced, and psychic harmony has been achieved, the Freudian declares the therapy a "success". The radical behaviourist, on the other hand, is not concerned with how the neurotic **feels**, but is interested only in what he or she **does**. For the behaviourist, therapy is successful when, and only when, the overt behaviour stops. On this view, the patient's oral reports of harmony or discord are quite beside the point. Thus, these different conceptions of a "cure" require correspondingly different kinds of evidence to support them; there is no neutral test for a cure that can decide between these therapeutic orientations. Thus, it is not at all that we are dealing with an empirical question here, since neither side accepts the other's "evidence" as evidence.

The standard conception of an **empirical question** is: "that which is decidable by appealing to objective experience." But, in this case, what kind of experience should count as objective experience? This question, notice, is not itself an empirical question; yet it lies at the heart of the dispute between our two competing therapists. It follows from these considerations that the original question of "**which** therapeutic method is most effective," is not in fact an empirical question. The reason that it is not an empirical question is because there is a distinctly **normative** (i.e., valuational) component to the conception of "cure". And normative questions are neither true nor false, but more related to such things as **approval** or **disapproval**, and likes or dislikes. For similar reasons, then, my dispute with Ennis and standard courses in critical thinking is **not** just an empirical issue. It is not just an empirical issue because we have different conceptions of what critical thinking **is**, and therefore different standards for admissible evidence.

To be perhaps excessively brief about it, the major differences between Ennis' view (which I take to be the standard view) and my own might be summarized as follows. Ennis's definition of "critical thinking" is "the correct assessment of statements", and he believes the proper training for this consists in what might normally be included in an informal logic course. Also, evidence for the successful achievement of the relevant skills can be measured by performance on standardized tests such as the Cornell Critical Thinking Tests or the Watson-Glaser test. These are the major features of the standard view. In my view, by contrast, critical thinking has little or nothing to do with performance on these standardized tests because, for me, critical thinking has to do with "engaging in activity with reflective scepticism", and there are almost as many ways of doing this as there are kinds of activities. For

me, there is no denumerable set of skills which demarcates critical thinking, so no single test could ever hope to capture it. Moreover, the normative difference between the standard approach to critical thinking and my own is that the standard approach takes its criteria for **good performance** from the field of informal logic, whereas I take my criteria from the different fields of study and activity. For these reasons then, the dispute between the standard approach (e.g., Ennis, **et al**) and my own view is not an empirical issue: we do not agree on the definition of "critical thinking", nor on the criteria for judging good performance, nor even on what constitutes evidence for critical thinking.

This latter point, that we do not agree on what constitutes evidence for critical thinking, is of sufficient interest and importance to warrant separate treatment here.

The Conflation of "Empiricism" with Tests and Measurement

En route to charging me with having made an empirical claim without substantiation, Ennis cites a study by David and Linda Annis "Does Philosophy Improve Critical Thinking?" as an example of the general type of evidence needed, not only to settle the dispute between him and myself, but to establish the effectiveness of critical thinking programs generally. Briefly, the Annis study is a typical statistical analysis of **pre** and **post** test results for different groups of students undergoing different course treatments. That is, it attempts to measure quantitatively the "impact" of several undergraduate courses on critical thinking ability. However, quite apart from any of the study's results, which were paltry, the study is of interest because it is an excellent example of some confusions which underlie much of the educational evaluation literature in general and the critical thinking evaluation literature in particular.

The confusions I have in mind are twofold. They tend to occur together in practice, so I will try to separate them here for purposes of analysis. Both confusions, however, stem from a rather virulent form of instrumentalism, or **test-measure-aphilia**, where the measurement tail wags the educational dog. The first such confusion is the tendency to equate "empiricism" with tests and statistical measurement—the assumption that for something to be "empirically known" it must be test-measurable. As Elliot Eisner observed.

Becoming familiar with correlation procedures too often leads simply to questions about what one can correlate: the existence of statistically reliable achievement tests too often leads to a conception of achievement that is educationally eviscerated. Our tools, as useful as they might be initially, often become our masters. Indeed, what it means to do any type of research at all in education is defined, stamped, sealed, and approved by utilizing particular premises and procedures. A brief excursion into the pages of the **American Educational Research Journal** will provide living testimony to the range of such premises and procedures. For example, during the past three volume years the AERJ has published over 100 articles. Of these only three were nonstatistical in character.¹²

In the Annis paper "Does Philosophy Improve Critical Thinking?"¹³ the authors strongly intimate that if you cannot statistically measure the effects of your courses then your belief in their value rests on dogmatism. In response to a statement by Bertrand Russell, where Russell touted the educational virtues of philosophy, the Annises say:

Although we may believe that philosophy has such an impact on our students, what evidence do we have for this belief? It is noteworthy that philosophers are quick to criticize others

for unsupported views, but when it comes to the issue of why philosophy is valuable, we ourselves rely on dogma (p. 145)...

A measure of the impact of philosophy on critical reasoning would be a comparison of the amount of improvement on the Watson-Glaser between philosophy and the control group. If the difference in the improvement on the total score is statistically significant, that is, if there is a low probability that the difference is due to chance alone, then we may conclude there is a differential impact. The statistical technique of analysis of variance is a measure of this differential impact. Analysis of variance applied to the subtests of the Watson-Glaser provides information on the specific impact of philosophy and the various courses (p. 148).

It is clear the Annises think that Russell's faith in the value of philosophy borders on dogmatism since he never statistically measured its effects with standardized instruments such as they recommend. Indeed, throughout the Annis study it is assumed that the value of any course of study cannot be known unless it is test-measured. No other type of evidence will seem to do.

It is arguable, however, that the Annis's unflinching reliance on such psychometric procedures is equally 'dogmatic' since as in this case, the validity of the testing instrument (i.e., Watson-Glaser) and the soundness of the research design, are usually open to serious challenge.¹⁴ One is here reminded of G.E. Moore's proof of the external world where holding his hand up in front of him he declared "this is a hand before me"; and this, he argued, is more certain than any principle of scepticism upon which one's doubt might be based. In the social sciences, as in daily life, it is likely that there are more phenomena which **resist** accurate and valid measurement than there are those which submit to it; and it is always open to question which kind one has before one. Thus, as in Moore's argument, there is often as much reason to question indirect statistical evidence about the efficacy of critical thinking programs, or educational programs, as there is to accept one's own evidence based on direct inspection.

I would like to make it clear that I am not opposed to the broad use of statistical research procedures in education, and I would be among the first to defend their usefulness in many situations. However, I **would** suggest that when they are used to assess programs which are intended to have wide-ranging outcomes, such as a liberal education program, or a critical thinking program, they pose very serious validity problems. So much so that there is more reason to question their validity (as in Moore's proof) than there is to accept it at face value. In the Annis study, for example, the entire second half of the paper is spent offering methodological alibis for why they failed to find any statistically significant results in this study: maybe we should have taught them longer, maybe we should have more items in the subscales, maybe we should have tried the Cornell, maybe we didn't teach the philosophy correctly that we did teach, etc., etc. It never occurs to the Annises that maybe, like measuring sweetness with a yardstick, there is something wrong-headed in what they were trying to do, that there is a bad fit between what they want to know and how they are trying to find out. It is no accident that complex courses or programs which are intended to have diverse outcomes, are not normally evaluated psychometrically, let alone in a pre-test/post-test format. We seem to realize intuitively that our methods of direct inspection of these programs are usually more valid than psychometric test instruments. To demand statistical rigour where it is not likely to be forthcoming is an instance of what A.N. Whitehead called "the fallacy of misplaced concreteness". We should not feel lacking in academic "integrity", as the Annises actually suggest at one point, because we do not share their enthusiasm for the

psychometric evaluation of critical thinking programs. Rather, we should be acutely aware, as they are not, of the very real limitations of these procedures.

A second confusion pervading the Annis study is the assumption that the sole purpose of education is to develop skills, as such, which have **instrumental** value; moreover, these skills are always considered to be psychometrically test-measurable. There is a total failure to recognize that much of our educational knowledge and understanding does not involve skills of any kind, yet has an **intrinsic** value to us. That the Annises value the various educational subject areas solely in terms of their **instrumental value** can be seen from the fact that on every page of their article there is at least one reference to the "impact" which subject area X has on critical thinking. Indeed, the word "impact", used as a noun, is the single most recurrent word in the paper. It is clear that for them subjects are not studied for their own power and interest, but rather for what **impact** these subject make upon critical thinking ability **per se**. Here is a sample of the Annis's instrumental interpretation of educational value:

Although we may believe that philosophy has such an **impact** on our students, what evidence do we have for this belief? It is noteworthy that philosophers are quick to criticize others for unsupported views, but when it comes to the issue of why philosophy is valuable, we ourselves rely on dogma. The same principles of rational belief that commit philosophers to the careful and critical assessment of the reasons for some philosophical views, require us to be concerned with empirical support for claims made about the **impact** philosophy has on students. The present study is an initial step at empirically assessing the claim that the study of philosophy improves a person's ability to think critically (p. 145).

Elsewhere:

At present, however, there is practically no direct empirical evidence of what **impact**, if any, philosophy has. Furthermore, even if philosophy does have an **impact**, we need to know more specifically what effects it has ... What specific abilities are affected by these courses? Since we do not know what **impact** philosophy has, we also are ignorant of instructional factors affecting critical thinking in philosophy (italics are mine, p. 147).

The Annises never even entertain the possibility that the value and purpose of philosophy does not reside in its capacity to "impact" (the new verb) **skills** or **abilities** as such, but rather, to **do** philosophy just **is** to engage in critical thought. That is what the discipline **is**. Philosophy does not try to develop instrumental **skills**, as such, but rather to provide insight and understanding about the frailty of the human condition. Having this insight and understanding just is to be thinking critically. Herein lies its power and its purpose; its purpose is not to develop **skills**, such as those the Annises want to test for. The purpose and value of philosophy, as with most academic disciplines, is to provide a perspective through knowledge and understanding. And this perspective, I would argue, is the most important ingredient in any situation requiring critical thinking.

The Annis's emphasis upon skills, and their consequent de-emphasis of knowledge and understanding as provided by the traditional disciplines, is symptomatic of a wider trend in the critical thinking literature. This literature all but ignores the traditional liberal arts disciplines, or dismisses them as though they were irrelevant relics in an academic museum. The disciplines, the thinking goes, exist merely for the enjoyment of academic specialists, and a few artsy eccentrics. In critical thinking text books, the power of a liberal arts education has been either brushed aside or forgotten in a kind of mass amnesia suggestive of a new dark age. This represents either a

loss of faith or a failure to remember that the origin and justification of the liberal arts has always been that it liberates people from everyday ignorance. Moreover, it liberates people from ignorance about everyday problems. That is, in precisely those situations that the so-called critical thinking skills purport to be so useful. It seems we need to be reminded that history, literature, philosophy, and science **are about this everyday world**. They are not museum pieces, but rather they provide the perspective from which rational beliefs and decisions are made, and from which they can be judged. Indeed, Paul Hirst argues that to use these traditional forms of thought is synonymous with having a rational mind. If this view is even partially correct then it has already demonstrated its centrality to situations requiring critical thought. For these reasons, I favour improving our methods for teaching the disciplines in trying to develop critical thinkers. Whichever way one resolves this pedagogic issue, however, I hope I have made it clear that the question is anything but a straightforward empirical one.

Notes

1. See my **Critical Thinking and Education**. New York: St. Martin's Press, 1981; Oxford: Martin Robertson Publishers, 1981.
2. See, in particular, Robert Ennis, "Logic and Critical Thinking", **Proceedings of the Philosophy of Education Society**, 1981, pp. 228-232.
3. **Ibid.**, p. 229 and p. 231.
4. For a representative sample of this literature see: David and Linda Annis, "Does Philosophy Improve Critical Thinking?", **Teaching Philosophy** 3:2, Fall, 1979; Daryl G. Smith, "College Classroom Interaction and Critical Thinking", **Journal of Educational Psychology** Vol. 69, 1977, p. 180; Bruce L. Stewart, "Testing for Critical Thinking: A Review of the Resources," Illinois Rational Thinking Project, Report 2, 1979; Thomas N. Tomko and Robert H. Ennis, "Evaluation of Informal Logic Competence," **Informal Logic: The First International Symposium** (eds.) J.A. Blair and R.H. Johnson (1980).
5. Originally printed in **Psychological Review**, 1956.
6. **The Psychology of Deductive Reasoning** (1982), London: Routledge and Kegan, pp. 6 and 254.
7. David P. Ausubel, **Educational Psychology: A Cognitive View**, Second Edition, New York: Holt Rinehart, p. 544.
8. "The Teaching of Thinking," first appeared in **The School Review**, 1965, vol. 73, p. 1-13; reprinted and quoted here from **Democracy, Pluralism and the Social Studies** (eds.) J.P. Shaver and H. Berlak (Boston: Houghton Mifflin, 1968), pp. 384-392.
9. The Manual for **Watson-Glaser Critical Thinking Appraisal** (New York: Harcourt Brace Jovanovich, 1980), p. 1.
10. **Op. cit.**, p. 231.
11. **Teaching Philosophy** 3:2, Fall, 1979, pp. 145-152.
12. "On the uses of educational connoisseurship and criticism for evaluating classroom life," **Teacher College Record**, Feb. 1977, Vol. 78, No. 3, p. 349.
13. **Op. cit.**, p. 145.
14. For criticisms of this test see Bruce L. Stewart, **Testing for Critical Thinking: A Review of the Resources**, Rational Thinking Report II, Bureau of Educational Research, Urbana, Illinois, 1979; see also my **Critical Thinking and Education** (New York: St. Martin's Press, 1981). ●

John E. McPeck, Faculty of Education, University of Western Ontario, London, Ontario, Canada N6G 1G7