

Toward an Empirical Definition of the Thinking Skills

RICHARD B. MILLER

East Carolina University

Professor McPeck has gained notoriety as an outspoken critic of critical thinking. *Critical Thinking and Education* has provoked a heated debate. [1] One of McPeck's central claims was that critical reasoning was a chimera. The "general reasoning ability," which theorists attempted to study, measure and define and practitioners in the classroom attempted to improve, simply did not exist. One conclusion which he draws from this is that courses in critical reasoning or informal logic ought to be rejected. McPeck has chosen the valiant course of trying to prove a negative existence claim. As is well known such claims are exceedingly difficult to substantiate. One of the major lines of criticism is that he is advancing empirical hypotheses without empirical evidence. [2]

McPeck responds in "The Evaluation of Critical Thinking Programs: Dangers and Dogmas" (*Informal Logic* vol. vi, no.2, pp. 9-13; hereafter, "Dangers and Dogmas"). His response represents a significant modification of his position. He no longer claims to prove that critical thinking ability does not exist; he now seeks to put the burden of proof on proponents of critical thinking and points out difficulties in proving this. I accept the challenge and hope that the discipline as a whole will accept the burden of proof. In this paper I discuss the difficulties which McPeck has raised for proving the existence of critical reasoning ability and a *fortiori* for

justifying courses which purportedly improve this ability, and I summarize the results of my initial attempts to build on the currently existing evidence.

I

In his new paper McPeck cheerfully admits that he has not provided empirical evidence and finds the demand for such evidence "at the very least contentious" ("Dangers and Dogmas," p. 9). He has reason to find it so. There are, roughly speaking, two ways to prove the non-existence of x's. There is the tedious route of examining everything that could conceivably be an x and showing that it is not. This McPeck has never claimed to have done. The other is to demonstrate that there are no married bachelors from the concepts alone. This was his strategy in *Critical Thinking and Education*. From McPeck's point of view, the demand for evidence of his non-existence claim misses the point. He had given an *a priori* argument proving, to his own satisfaction at least, the conceptual incoherence of the concept of critical thinking. This objection to his book must surely have struck him as like demanding that he testify that he has looked for married bachelors and not found any. Further, if McPeck has proven that there is no such thing as critical reasoning ability then he has, to put

it mildly, put courses which attempt to improve critical reasoning ability in a bad light.

I understand why the objection appears contentious from McPeck's point of view, but I do not agree with him. I have not, and it seems that very few have, been convinced by McPeck's *a priori* arguments. For me and those of like mind, McPeck's book stands as a challenge to prove the existence of that which he tried to argue away. To accept the burden of proof and show that critical thinking skills exist will surely require more than a *a priori* argument.

McPeck's argument that, since thinking is always thinking about *x*, there can be no such thing as critical thinking simpliciter has been directly and non-contentiously attacked. (See, for example, Perry Weddle's review of *Critical Thinking and Education*, in *Informal Logic*, July 1984). Two major lines of criticism seem clear. McPeck is in no position to put forth his claim that "general reasoning ability" is conceptually incoherent, hence non-existent, for two reasons 1) his argument is defective (see below) and 2) there is no unitary concept of "general reasoning ability" for McPeck to analyze. Concerning the latter, McPeck himself admits that there is no consensus on what critical thinking is. Indeed one of his objections to the use of tests to measure critical reasoning [see below objection (3)] depends on the fact that there is no consensus on this issue. But if this is the case, then an attempt to show that the concept of general reasoning ability is incoherent faces the problem that "the" concept cannot be refuted because "the" concept does not exist. The most that can be expected are separate criticisms directed to each of the many concepts actually employed. Even if two philosophers both use the term "critical thinking," surely nothing can be deduced from the nature of the expression unless one assumes, contentiously, that the same concept

must be represented by both uses of the expression.

Concerning the former, I offer the following reductio: if McPeck's argument were valid then we could prove the non-existence of any skill or ability. A skill or ability is always general; that is, one does not possess a skill or ability unless successful performance is repeatable. Further specification of the conditions under which the ability is exercised can always be described as a "content." Seeing is always seeing some object. Walking is always walking in some particular place. If McPeck were correct, we would have to chide the proud parent who tells us that the baby has learned to walk by reminding him that there is no such thing as the general ability to walk. There are abilities to walk in the living room, walk in the playground, walk on the lawn etc. Which of these abilities is it that he meant? It cannot be seriously doubted that there is a generalized ability to walk. The general ability to walk is never employed and logically could never be employed without some "content." But once one has learned to walk, one can walk on any non-slippery, flat surface. There are no content-specific walking skills. Walking is a general skill even though it must always be exercised on some particular surface. McPeck's skeptical conclusion about "general reasoning ability" does not follow from the fact that there is always a content to thought. It is the very least possible that recognizing the validity of instances of *modus ponens* is an ability which may be exercised independently of the subject matter of the sentences.

It is similar with the ability to add. Addition is not an unanalyzable ability. It probably consists of separate abilities to add the integers from 0 to 9 and the rule for carrying. But addition is a truly general ability. One also exercises the ability on some content. One adds $3 + 9$ or $115 + 878$. But the ability to add is not the ability to add particular numbers. If it were, there

would be little point to teaching arithmetic, since one cannot predict which numbers the child will someday encounter. Luckily, one does not have to relearn addition whenever one is presented with a new pair of numbers. So when a grammar school teacher says that he teaches addition, it would be foolish to ask "of which numbers?" To teach addition is to teach someone how to add any numbers. [3]

My counter-argument is this: if the concept of "general reasoning ability" is "conceptually incoherent," then so is the concept of "the ability to add." But the later concept is not incoherent. [4] Therefore McPeck's argument is defective. At least the non-existence of general reasoning ability has not been proven.

II

In "Dangers and Dogmas," McPeck has adopted a new, and in my view wiser, strategy. Rather than trying to prove the non-existence of "general reasoning ability," he now argues that it has not yet been proven that it does exist and that there are grave difficulties in the way of those who would try to prove this. He concludes:

Thus, despite the prevalent jargon, there are insufficient grounds for believing 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 skeptic. ("Dangers and Dogmas", p. 10)

Most, with good reason, do not find his Oxford, ordinary language style argument convincing. But this is a weaker and more reasonable claim than conceptual incoherence; it represents a retreat to a more defensible position. I think that the discipline should accept the burden of proof.

In "Dangers and Dogmas," McPeck gives some insightful criticisms of the assumptions which lie behind the use of objective tests to measure critical reasoning ability. If

his criticisms provoke some critical reflection on the whole process of assessing these skills, then he will have made a major contribution to the maturity of the field. I believe that McPeck's criticisms are valuable but that they show only the difficulty, not the impossibility of measuring these elusive skills.

McPeck's new argument is best restated as a dilemma.[5]

1. If general reasoning ability exists, it is either a single skill or a small set of skills.
 2. There is no single reasoning skill.
 3. There is no small set of reasoning skills.
-
4. General reasoning ability does not exist.

A few comments on premise 2 are in order here. At times McPeck sounds as though the field of critical reasoning abounds with people who believe that premise 2 is false. He writes "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 skepticism is surely justified" (p. 10). Who are these proponents? In *Critical Thinking and Education* Ennis's account of the concept of critical thinking is said to have "formed a general framework around which subsequent efforts have tended to build." [6] Does Professor Ennis believe that a unitary "critical thinking ability" exists? Ennis addresses this question directly in "Problems in Testing Informal Logic Critical Thinking Reasoning Ability" (*Informal Logic*, Jan. 1984).

My inclination is to say that critical thinking ability is fairly heterogeneous, consisting of such diverse elements as open-mindedness, ability to see other alternatives, experience and background knowledge, knowledge of criteria to apply in thinking critically, ability to handle complexity in an orderly fashion, and some others. (p. 6)

If Ennis has set the general framework for discussions of the nature of critical

thinking and Ennis clearly rejects the simplistic unitary ability concept, then the identity of these proponents becomes a matter of some curiosity. Yet McPeck gives no examples of anyone who believes this. I suspect he is demolishing a strawman. Surely there is disagreement over the nature of critical reasoning, but to my knowledge no one has argued that there is a general unitary ability to think critically. The controversy in the nascent discipline is over how many and what kinds of basic reasoning skills there are. Premise 2 is a truism, not a controversial insight. I would ask Professor McPeck to name a major figure, or even a minor figure, in the field who believes that such an ability exists.

Concerning premise 3 the argument gets a great deal more interesting. McPeck has three main points to make in support of premise 3. The first two points are directed primarily at the Cornell Critical Thinking Tests and the Watson-Glaser. He complains that tests which attempt to measure a small set of abilities suffer from the following defects:

(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. ("Dangers and Dogmas," p. 10)

These criticisms are intended to torpedo two of the best known instruments in the field. If the tests beg important questions, then, it is alleged, using the tests to prove the effectiveness of critical reasoning programs is questionable. One might respond that it is preferable to use the tests, no matter how imperfect, rather than relying on some vague intuition. I am not happy with such an answer however and offer what I hope is a stronger rebuttal below.

A third point (I have added the number for convenience) makes no reference to any particular instrument. It is intended to destroy not only the Cornell and Watson-Glaser but all such tests.

[(3)] 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'.... Where there are competing conceptions of critical thinking it is unlikely that any neutral test can arbitrate among them. ("Dangers and Dogmas," p.11)

I will deal with (3), the most general point, first. McPeck has hit upon an important problem here, but he draws the wrong conclusion. All tests and measures of critical thinking are theory-laden. Indeed all tests and measures are theory-laden! However, the implied conclusion that theory-laden tests cannot provide evidence to arbitrate among theories is false, as reflection on the relation between theory and measurement techniques makes clear.

One must distinguish between two stages in the history of a measuring technique. In its mature stage, the technique is used uncritically. Astronomers look through telescopes, chemists apply litmus tests, biologists employ staining techniques to learn about the objects that they subject to these tests. They do not question the instruments or techniques employed, but this is not because they are begging any questions. The techniques themselves have gone through a period of critical evaluation. Even the telescope had to prove itself. In their first few decades telescopes were used, but the evidence gained from them was not universally accepted. Nor was it, as those with a superficial view of history suggest, irrational to be skeptical of the new instrument. Only after the results of those tests turned out to confirm the value of telescopes could telescopes be used uncritically. This is typical of the history of a mea-

asuring instrument or technique. One constructs measurement devices to test theories and one uses theories to construct measurement devices, but whether the theories will be confirmed or disconfirmed cannot be determined in advance. Theories and measuring instruments evolve in tandem. We are at present in the "earth, fire, water and air" stage with respect to reasoning abilities. Our theories are no doubt far cruder than reality and our measuring instruments reflect the state of our theories. [7] We are looking through very primitive "telescopes" and should not be shocked to find a McPeck refusing to look through them.

The fact that the tests are theory-laden does not make them useless and question-begging. On the contrary, it makes them useful for refuting and revising theories. The test must produce results that are (1) self-consistent, (2) consistent with other independent measures, (3) consistent with other beliefs widely held in the scientific community. For example, mercury thermometers are theory-laden. Thermometers measure temperature, but they also presuppose that mercury in contact with glass will reach temperature equilibrium with a surrounding substance in a few seconds and that the volume of a quantity of mercury increases predictably with increases in temperature. Using thermometers allows us to test these theories. If mercury thermometers did not give consistent readings or if the readings of mercury thermometers were inconsistent with thermocouples, resistance thermometers, and bimetallic thermometers or if they told us that boiling water was really colder than ice, then we would know that our theory had gone wrong somehow and mercury thermometers would not be used.

It is, or should be, the same with tests of critical thinking. The tests should be 1) internally consistent; 2) they should correlate with each other; 3) they should enable us to predict such things as success in fields which

are generally believed to require good thinking skills (e.g. business management and the professions). Hence, properly used, theory-laden tests can provide the objective evidence required to decide among rival definitions of critical thinking. Suppose the Cornell Test lacked internal consistency according to some plausible measure: e.g., suppose the variation in test scores of an individual taking the test on successive days was equal to the variation in test scores of two randomly selected individuals. (I don't mean to suggest that small variation in scores of the same individual is a sufficient or even a usual measure of consistency, but I do think that it would be at least a necessary condition.) In that case, the assumptions that lie behind the construction of the Cornell Test have been disconfirmed. Or suppose that, on the contrary, there is found a very high degree of correlation between scores on the Cornell Test and the Watson-Glaser. This would tend to confirm the validity of both tests. Or suppose that while these two tests correlated well with each other, they also correlated negatively with academic achievement or success in management. This would tend to disconfirm both. Contrary to McPeck's intuitions on the matter, I find that the high correlations of critical reasoning tests with IQ and reading ability do support their claims to validity. Surely reasoning skills are essential to the ability to read beyond the "Go, dog, go" level. If there were low correlations between reading comprehension and critical reasoning scores, eye-brows should be raised.

Have the Watson-Glaser and the Cornell tests been subjected to and survived this kind of scrutiny? McPeck's points (1) and (2) display some confusion, but do succeed in raising this important question. Of course the designers of the tests assume that parts of the test measure particular abilities and that these abilities add up to "real critical thinking." But these assumptions are not sacro-

sanct. Independent observers who do not share them can objectively examine the data. *Indeed the construction of the tests provides independent observers with the means to evaluate these assumptions.*

To be fair to McPeck, I believe that what he is driving at is the fact that the tests themselves have not been adequately tested and that reliance on their results prior to such rigorous evaluation is unwarranted. McPeck has raised the highly important question of whether or not they have been subject to such an examination. McPeck believes that they have not. Studies have been done to try to evaluate the tests; see, for example, the information included with the Cornell critical thinking tests. But McPeck does not want to be drawn into an empirical battle. His criticisms have been loftily a prioristic.

One problem which he finds with these tests is the overlooked distinction between achievements and abilities.

Educators and measurement-types have mistakenly taken the description of an achievement as indicative of an ability. ... Achievements do not necessarily describe corresponding abilities. ... Similarly for such achievements as "defining a problem" or "correctly evaluating evidence", one cannot assume that a unitary ability is indicated. In such cases literally hundreds of separate abilities might have been involved, or conversely, nothing recognizable as an ability might have been involved. ("Dangers and Dogmas," p.10)

There is certainly a distinction which needs to be noted here, and perhaps test designers have been insufficiently sensitive to the difference. Surely some reflection on the distinction is called for. We test abilities by achievements. *To have an ability is to be able to repeat the achievement an indefinite number of times under appropriate conditions.* Both repetition and the specification of appropriate conditions are crucial. Of course, when testing the ability to sprint we do not allow one person to use a bicycle while all others have only their

Nikes, but that is because we know that running and bicycling are different activities. Do we know what mental processes our test takers are performing when they decided to blacken a "b" on a test item? Do all test takers who give the same answers perform the same mental calculations? What conditions need to be held constant and which can be varied? These are difficult questions that McPeck is hinting at, and he is quite justified in requiring answers to them. But McPeck is at fault for suggesting that test designers have been unaware of such difficulties. Robert Ennis, for example, has addressed himself to the difficulty of test construction very recently ("Problems in Testing Informal Logic Critical Thinking Reasoning Ability," *Informal Logic*, Jan. 1984). Included in Ennis's recent article is a good discussion of the problem of the variation in background beliefs held by the test takers.

Further reflection on the distinction between achievement and ability does seem called for. On the one hand, one cannot fault McPeck's distinction, but on the other, surely achievements are the most convenient measure of abilities. But different sorts of abilities are related to achievements in different ways. I will fall back on simple physical analogies once again, as I did with the case of walking.

The difference between a good hitter in baseball and a poor one is not that the good hitter hits every pitch and the poor hitter misses every pitch. The difference is a matter of how often the good performer achieves the desired result. Consequently it is impossible to test batting skills by monitoring the performance of a batter on a single pitched ball. If the analogy to reasoning skills is apt, then it would be necessary to make discriminations based on performance in several trials. And the analogy may be more apt than it appears at first. We all know of cases of trained logicians affirming the consequent or illicitly converting A propositions. The only difference is that they do so

much less often than other people and recognize their own mistakes faster.

In the case of some skills, like hitting a baseball, only a large number of trials will show what degree of skill the individual possesses. This is not the case for all skills. Performance in the hundred yard dash is very consistent. It is possible to distinguish great sprinters from mediocre sprinters from poor sprinters in a single race because the variation in the times of the same sprinter is so slight. However this knowledge comes from observation of many sprinters in many races. A priori we could not tell that sprinting ability could be measured so well in a single race.

If we are testing the average ability of a group, the difference between sprinting and batting is not important. If we throw one pitch to each individual, and there are enough individuals in the group, then the group batting average will probably reflect the average skill in the group. If we are concerned with testing the effectiveness of a critical reasoning course, then group averages are appropriate. But if we are concerned with individual performance on particular items, e.g., whether this individual does or does not possess a given skill, then it becomes necessary to repeat test items. Until it is proven that reasoning ability is more like sprinting than hitting a baseball (which seems very unlikely), repetition is necessary.

III

Such questions lead me to consider how well the currently existing tests measure the particular thinking skills of individuals. Internal consistency is generally measured by mathematical formulas such as the Kuder-Richardson which measure the difficulty of individual items on the test against the test as a whole. The basic idea is that individuals who do well on the test as a whole should do better on individual items than individuals who do not do well on the test. Ennis has

argued cogently that Kuder-Richardson formulas are of limited use in evaluating the consistency of critical thinking tests because of the heterogeneous nature of critical thinking ("Problems," *Informal Logic*, Jan 1984).

The idea is this. If critical thinking is really several abilities not one, and as I have said before there is no serious doubt that it is really several abilities, then it is not only possible but likely that individuals will have quite different patterns of strengths and weaknesses. So that a given individual may do quite well on the test as a whole, but have some very definite weakness. Her weakness need not correspond to what is the most frequent weakness in the population taking the test. That is, it need not be the "hardest" question in general for it to be the "hardest" for her. If the test purported to test a single unitary ability, such results would be incongruous, but not if the test is intended to measure several possibly unrelated abilities which together add up to critical reasoning ability.

If this is correct, then the standard measures of internal consistency are not reliable for reasoning skill tests. It seemed to me unfortunate to give up on this point and simply admit that the internal consistency of these tests could not be measured. I then tried to construct my own test of internal consistency. It would have to be the consistency of subtests, however, because of the heterogeneity of critical reasoning skill itself. The internal consistency of the whole test would have to be found in the consistency of the subtests.

I decided to use the New Jersey Test of Reasoning Skills, because I had already done some work with it and because it possessed highly specific subtests. This was a distinct advantage for the kind of research I wanted to do.

I had first to get clear about predictions I could make about the distribution of right and wrong answers on the hypothesis that the questions tested the same skill as opposed to the hypothesis that they did not. Initially, I

thought that iff two or more questions tested the same skill then the ratio of right/wrong answers should be the same. However, a little reflection caused me to abandon this thought. Questions which tested different skills could have the same ratio. This fact indicates only that the questions are equally difficult, not that the questions test the same skill. Moreover, questions with different degrees of difficulty could test the same skills, as an 85 mph fastball tests the same skill as a 95 mph fastball.

Having a skill or ability is a matter of being able to repeat successful performances under controlled conditions. So what mattered was not how many got the questions right, but whether or not *the same individuals* got all three questions right (or wrong). Ideally (if reasoning skill was like sprinting) everyone who possessed the skill would get all the questions right and everyone who did not possess the skill would get all the questions wrong. None would get the easier questions wrong and the harder question right. Only a few would get the harder question wrong and the easier question right.

I did not expect it to actually work out so neatly, of course. For one thing, even if there were a clear distinction between those who had the skill and those who did not, those who did not would guess and in some instances guess correctly. Moreover, I suspect that having skill at reasoning is more like skill at hitting a baseball. The skillful succeed *more often* than the less skillful.

It is necessary to distinguish between using the results of the test 1) to measure the skill of an individual, 2) to provide evidence that a program designed to improve thinking skills is working, 3) to confirm or disconfirm the existence of various hypothesized general reasoning skills. Most attention has been directed to purpose 2). Quite naturally, those who are attempting to design and implement educational programs want to be able to prove that they work. My concern was with 3). A test may be well-designed for pur-

pose 2) and yet be inadequate for 1) or 3). Let me emphasize that the sort of evidence of internal consistency that I was looking for is at best necessary but not sufficient for proving the existence of generalized skills. As in other serious empirical questions, there are no crucial experiments. The judgement as to whether the existence of these skills is "proven" or not is really a very complex judgement based on weighing evidence of various kinds.

The result of my reflections on the nature of critical reasoning lead me to invent a new test for internal consistency. I felt that I could not depend on the familiar Kuder-Richardson formulas. The method I used was as follows: I administered the New Jersey Test of Reasoning Skills to 132 introductory level philosophy students in five different classes.[8] But I was not interested in the whole test. Instead, out of the fifty items on the test I selected five sets of three questions. Each set could plausibly—that is, until I got a look at the results of the test!—be interpreted as testing for the presence or absence of a unitary skill. The skills were 1) recognizing ambiguities, 2) standardizing A propositions, 3) part/whole reasoning {the fallacies of composition and division}, 4) hypothetical reasoning {modus tollens, denying the antecedent and affirming the consequent}, 5) the Barbara syllogism {two valid figures; one invalid figure}. I then computed the probability of the null hypothesis. My null hypothesis was that the three questions did not test the same skill, i.e., that a correct answer on one of the three questions did not affect the probability that a correct answer would be given on the other two. In other words, the null hypothesis treated the questions as independent coin tosses. The probability of a correct answer on any question was simply taken to be its frequency. The probability of getting all three questions in any set correct was simply the product of the probabilities of each correct answer. I then counted the actual number of individuals who answered a whole set correctly. I reasoned that if the three ques-

tions did test a single skill, the actual number would be greater than the number predicted by the null hypothesis. The difference was then tested for significance and the results are recorded in Table 1.

Table 1

SKILLS	NULL HYPOTHESIS
recognize ambiguities	1.00
standardization of A prop.	.88
part/whole reasoning	.73
hypothetical reasoning	.72
Barbara	.15

{The figures under the column represent the probability that the actual results would be obtained if the null hypothesis was correct, i.e., the ability to get one question right was completely unrelated to the ability to get the other two questions in the same group. For example, the data indicate that if part/whole reasoning is completely content dependent, results like those which I obtained would occur three out of four times.}

The results, while based on a fairly small sample, are suggestive. The first four sets of questions do not seem to test general skills. McPeck's skepticism about the existence of general skills is unrefuted and since the burden of proof is on those who believe that there is a generalized skill at, e.g., recognizing ambiguities, skepticism is at present justified. (This is not to say that McPeck is right in rejecting courses in critical thinking!) Being able to recognize one ambiguity does not seem to be correlated with recognizing another ambiguity. The incorrect answers were scattered randomly throughout the population. The standardization of "each," "always," and "only" shows a very minor positive correlation, but this degree of correlation is so slight that it would occur by chance variation almost nine times out of ten. Practically the same situation holds for part/whole reasoning and hypothetical reasoning.

Only with the Barbara syllogism is there good reason to believe that all three questions test the same skill. In this instance there is some reason

to believe that a unitary skill has been isolated. Only 15 times out of 100 could this deviation from the null hypothesis have happened by chance. In larger samples I feel confident that the significance would be even greater. The results of the first four skills represent a failure to supply the evidence that would refute McPeck's contention that general reasoning skills exist, but the last suggests that individual reasoning skills can be isolated and their existence verified.

After looking at these results, I now think that it is unlikely that "recognizing ambiguities" or "standardizing A propositions" is a skill. I suspect that there are as many skills of recognizing ambiguities as there are ambiguities. Similarly the ability to standardize "always" is probably one skill and the ability to standardize "only" is another. But perhaps with larger sets of questions designed to test the same skill (four, five, or even ten) statistically significant results would be obtained for all five skills.

Do these results mean that the New Jersey Test of Reasoning Skills does not reliably test reasoning ability? That is not an appropriate conclusion to draw. As indicated above, there is evidence of the external validity of the test as a whole. The problem that I uncovered had to do with the internal consistency of the subtests. This means that we do not know which reasoning skills the NJTRS tests. How is it possible that a test could fail to identify correctly the specific items which it tests and yet have external validity? There seems to me to be a plausible explanation of this. I think it likely that the test questions which I grouped together really represent different skills; e.g., *modus tollens* is not the same skill as avoiding denying the antecedent. The fact that there was only one *modus tollens* on the test means that it is logically impossible to use internal consistency to determine if there is such a skill. Nevertheless, the test does measure the ability of students on a wide variety of different skills, most of which are correlated with other

measures of academic performance. There are simply too few items on the test as a whole, and the subtests are too broadly defined, to allow for a reliable test of internal consistency of the type that I have used. In other words, it might be a very good test if used for the purposes of evaluating how well individual students reason on average, and in consequence of this it might be useful in testing how well a program improved reasoning ability. It does not seem to be a good test for doing research into what the reasoning skills really are. These tests were not designed for research into the nature of reasoning skills. They were designed for, and are primarily used for, the evaluation of programs. Despite weaknesses in the tests, it seems preferable to rely on them at the present time than to give in to "intuition" and anecdotal evidence. There is a clear need for tests designed specifically for research into the theory of critical reasoning ability. I feel confident that such research would show that a very fine-grained approach must be taken to defining the thinking skills.

The results obtained with the Barbara syllogism triplet suggest how the test could be redesigned. The skill isolated here is quite precise. I suspect that a triplet of *modus tollens* would show the internal consistency that the triplet composed of one *modus tollens*, one denying the antecedent and one affirming the consequent, lacked. I plan to construct a test composed of triplets of questions designed to test very narrowly defined skills and to give such a test to a larger number of students. In this second trial, I will try to prove the existence of from 15-20 different skills. The cooperation of other philosophers interested in empirically investigating this issue would be welcome.

The question which concerns me and should concern the discipline is how many reasoning skills there are. How finely must we discriminate skills in order to get at reality? It is obvious that five skills is much too crude a picture. How many then? Twenty? Two

hundred? A thousand? The data I have at present is of course inadequate to support a very good guess, but it seems reasonable to suppose that there are at least hundreds. If Barbara is a separate skill from *modus tollens* and *modus tollens* from recognizing the invalidity of affirming the consequent and recognizing the invalidity of affirming the consequent from standardizing "only" sentences etc., then even a thousand distinct reasoning skills does not sound far fetched.

Supposing for the moment that this estimate is not too far wrong, what consequences would this have for the teaching of critical reasoning? If there are too many of these skills, then teaching critical reasoning will not be feasible in a single course. If there were a thousand such skills, then one could do no more than scratch the surface if skills were, as is usual, taught serially to ordinary-sized classes. New approaches would certainly be called for but it is probably not appropriate to speculate about that in this paper.

Conclusion

We are still in the "earth, fire, water and air" stage with regard to the defining and testing of critical reasoning. There is a difference between constructing tests to 1) measure the critical thinking ability of individuals, 2) measure the effectiveness of educational programs designed to enhance critical thinking ability, and 3) test the reasoning skill taxonomies currently in existence. There is a great deal of disagreement over the nature of critical thinking. Contrary to McPeck, I argue that tests incorporating these differing assumptions can be used to resolve some of these disagreements. However, I find that to do this properly tests will have to be designed explicitly with purpose 3) in mind. Repetition of items designed to test the same skill are necessary if the test's purpose is 3) or probably even 1) (unless most of these skills are more like sprinting than hitting a baseball), but not for

2). I have constructed such a test and will have results from it early in 1986. Doing the job properly will require the construction of many such tests. Once again, I invite interested philosophers to cooperate with me on this project.

Notes

- [1] John McPeck, *Critical Thinking and Education* (New York, St. Martin's Press, 1981).
- [2] McPeck, in "The Evaluation of Critical Thinking Programs: Dangers and Dogmas," *Informal Logic*, vol. vi, no. 2, pp. 9-13, has singled out Robert Ennis, "Logic and Critical Thinking," *Proceedings of the Philosophy of Education Society*, 1981, pp. 228-232, as a particularly prominent example of this kind of criticism of his work. McPeck argues that the "general reasoning skills" which critical reasoning courses seek to enhance do not exist and on this basis that such courses ought to be rejected. Professor Ennis informs me that he did not intend to charge Professor McPeck with alleging the non-existence claim without evidence. To be precise, the objection which Ennis makes in "Logic and Critical Thinking" is "if McPeck is making the empirical claim that courses in logic and critical thinking do not help people to be better critical thinkers in a discipline, then he owes us some empirical evidence that is more substantial and contrary to the evidence that exists." (p. 229) Thus, Ennis objects that the consequence needs empirical evidence. I concur and extend the objection to the premise.
- [3] The matter is more complicated than this when looked at more carefully, as an anonymous referee has pointed out. For example, some children can add numerals arranged vertically but not horizontally. Others can add numerals but not things. Moreover, adding is not really unitary when it is broken down into component abilities as grammar school math teachers must. First, there is the ability to add pairs of single digit numerals, then there is the ability to add pairs of two-digit numerals, then there is the ability to carry, then there is the ability to add columns. This means that the ability to add is not a simple unanalyzable ability. It is composed of subroutines. However, this complexity does not detract from its generality. This sort of fine-grained analysis is undoubtedly correct and necessary to effective teaching of arithmetic, but it does not take away from my point, which was that the ability to add is general. Incidentally, I suspect that the analogy here holds with respect to logical abilities, and that in the future instruction in basic reasoning skills will be informed by the study of the subroutines involved, just as it is in the teaching of arithmetic today.
- [4] With apologies to Kripke/Wittgenstein. Saul Kripke, *Wittgenstein on Rules and Private Language* (Cambridge, Harvard University Press, 1982). The Kripke/Wittgenstein argument to the contrary would not be challenging and intriguing if the rest of us did not firmly believe that there was such an ability.
- [5] This argument is found in "Dangers and Dogmas" pp. 9-10.
- [6] *Critical Thinking and Education*, p. 40
- [7] See Ennis, "Problems in Testing Informal Logic, Critical Thinking, Reasoning Ability," *Informal Logic*, Vol. VI (1984), p. 7.
- [8] Though I was not concerned with the external validity of the test, an accidental condition provided evidence of this. Two of these classes were "Special Studies." Students in these classes were freshmen who had been admitted to college but

just barely met the admissions requirements and were placed in the program because they were judged to be marginal. Interestingly, the New Jersey Test concurred with the admissions office. The two Special Studies classes scored al-

most a full standard deviation below the mean for the test.

Professor Richard Miller, Department of Philosophy, College of Arts and Sciences, East Carolina University, Greenville, North Carolina 27834 □