

Thirty Great Ways to Mess up a Critical Thinking Test

By

Peter A. Facione
Professor of Philosophy and Education
California State University, Fullerton

ABSTRACT

This humorous guide offers quick, practical advice on assessment, especially CT testing. Eleven rules, like "Set no instructional priorities," and "Emphasize the trivial," apply to all educational testing. Fourteen apply to multiple-choice strategies. "No stems should avoid stating things negatively," and "Tip off wrong answers by category mistakes," rival PDQ Bach. Five rules apply directly to CT assessment. To ruin your next CT test write questions which: target information recall about CT but don't require CT itself; ignore differences in gender interests, domain-specific knowledge, etc.; are boring, time-consuming and yet entirely uncomplicated; require official CT vocabulary; and presume students think like trained experts.

Finals week already! At Michigan State in 1967 the general education course was called "Logic." In 1989 at California State University Fullerton we call it "Logic." Although I would now describe my goal in both cases as teaching critical thinking (CT), so much is different: my conceptualization of CT and its relationship to logic has changed, my curricular and pedagogical ideas about how CT is best taught, and even my sense of why it is important to teach CT. Yet one thing remains: then and now at MSU and at CSUF we give final exams. And exams are what this paper is about. Specifically CT exams.

Like anyone's, my CT course boldly assumes (1) CT can be learned, (2) CT can be assessed, and (3) CT can be taught. I sleep easier these days knowing that (1) is probably true in spite of anything I might do in the classroom. I pray (3) is true, and I most certainly conduct myself on campus as if it were. Imagine the budgetary bloodshed should the campus curriculum committee demand empirical evidence! And, I worry about how to defend (3). One sure fire defense would show how much students have improved in CT as a result of taking my course. But that leads right to assumption (2) -- that CT can be assessed.

This little paper is a tongue-in-cheek look at the practical aspects of framing a learning outcomes assessment tool -- a chore each of us undertakes whenever we prepare a final examination. In this case, fostering students' development of CT skills is taken as the chief learning goal of the course. If you follow the advise given here -- advise hard won through personal experience, trial and error, especially error -- you will be able to construct a truly dreadful test, regardless of your subject matter. Make no mistakes. This abysmal goal, I'm confident, is achievable. I know because I've hit pay dirt a few times myself.

Having elsewhere defended multiple-choice CT assessment strategies, this vulnerable mode of assessment will come in for special attention. But to say that it can be done badly is not to say that it cannot be done well. In fact, more than ever, I am convinced that multiple choice testing is not the enemy of learning, but a great aid. If done well. If done well. That is, if done well.

But, before we have fun at the expense of this really quite fine outcomes assessment strategy, let's start with basic seven rules which apply to both essay testing (ET) and multiple-choice testing (MCT).

1. Never plan ahead! The paradigm essay-tester (PET) will delay writing questions until actually en route to the exam room. In long sloping lines on the side board, words obscured by the play of sun and shadow, he will write three or more non-equivalent questions and tell students to select one. The ultimate multiple-choice-tester (UMCT) will procrastinate at least until the night before the exam. Then, starting from scratch, he will gin out forty or fifty nifty MC-items. Rushed for time, he will duplicate the MC-test using those blurry purple ditto masters that blob closed letters and fade out the final lines on a page.

2. Avoid assessment research. Both PET and UMCT roll their eyes when nerdy educationists mention "construct-validity" "criterion-referenced," "learning outcomes," or "test-reliability." It is wise to shun colleagues who want to discuss assessment. Never bring up pedagogy either, except when trying to impress administrators. Boycott department meetings cluttered with such lowbrow topics.

3. One mode of assessment is plenty. PET never gives an objective test and has profound doubts about the scholarly judgment and professional integrity of colleagues who do. UMCT blindly trusts machine-scored tools and has serious concerns about the reactionary attitudes of closed-minded colleagues who don't.

4. Never evaluate a test after it has been used. To achieve PET-hood drag out those old exam questions, if you can find them, which were out of date during the Reagan administration. If you cannot find them, contact a fraternity for copies. To move to the highest level of UMCT find reasons for not doing a computerized item-analysis. When you are able to hand back the Scantrons without revealing the correct answers, you are making progress. Above all, neither PET nor UMCT can abide student gripes about exams. Don't waste valuable class time going over dead exam questions.

5. Never start with a clear concept of CT. After all, who really knows what CT is anyway? Some people say CT includes everything from multi-modal forms of reasoning to fair-minded motives, others restrict CT to a set of elementary logic micro-skills. All this confusion is to our advantage. If we were clear about CT, then people might want us to be clear about related issues, like why how CT is best taught, and whether we are doing a good job of teaching it. All in all, a nasty can of worms. Best keep the lid on by being mushy about what we mean by CT.

6. Never specify which aspects of CT one's course aims to develop. What, you mean I don't teach all of CT? Can't I do argument analysis, judging the credibility of various kinds of claims, fallacies, formal logic, analogical reasoning, scientific method, diagnostics, probabilities, whatever else, and also cultivate all the proper CT dispositions in fifteen weeks?

7. Set no instructional priorities. Poor tests don't just happen. Planning them is contingent on weak instructional development. Don't fret about priorities like if mastering Venn Diagrams is more vital in a CT program than learning to interpret, analyze and evaluate newspaper editorials. Just start at the front of the text book and crank through it, chapter by chapter. Be democratic, treat every topic equally, then you won't have to proportion the number of test questions or the points a student can earn on those questions to the relative significance of different topics in your curricular plan.

This last point applies particularly well to MC-tests. Which reminds me of what Caesar said about such tests, "When in doubt pick 'C'!" No, that's not right. What Caesar really said was, "Remember, heroic troops, next week's exam will be 800 multiple-choice questions. It covers everything known to Romans or barbarians. You'll have all day to complete the test. Bring swords, shields, two No. 2 pencils, three photo ID's, and a bag lunch. Nobody will be allowed to leave the coliseum for any reason, even if the air conditioning fails."

A sadistic UMCT might thrill to the horror students feel when faced with a two hour multiple-choice final. Even garden variety multiple-choice testers can detect the telltale signs of fatigue: Bright students glare angrily as they start to find every question hopelessly ambiguous; poor students lean back in those hard, too-small tablet armchairs, groan and guess; and stressed-out students tremble, twitch, and squint as they fight stiffness and headaches. I'm know some say "No pain, no gain." But what good can possibly come from such suffering? After all, as every PET knows, the MC-format cannot possibly be valid for assessing anything as slippery and esoteric as CT. PETs dismiss out of hand the objectivity of MC-tests. PETs question the professional commitment of people who argue the superior efficiency of MC-tests. They ignore putative evidence that an MC-test can be made more sophisticated than the mind-dulling memory tests they remember from junior high school. To many, MC-tests are, by definition, rote and moronic. So, here are some ways to take the MC-test, that most abominable and odious of all assessment devices, the stalking horse of the CT movement, and make it even more pernicious.

8. Disregard question order and the frequency of answer choices.

Remembering Caesar's advice, test-wise students expect right answers to cluster in

the B, C, range and not at the A or D extremes. And beside, what difference can it make to a student's motivation and self-confidence if she or he finds the first seven to ten questions massively difficult? So, if you are going to worry about the order of the questions, do what the football coach suggests and hammer them right from the start with your toughest stuff.

9. Emphasize the trivial and make the test tricky. Don't even ask if the test has covered all the crucial things in the course in rough proportion to their relative importance. Even if I plan my course well, I still have the chance to fill my tests with questions about minor exceptions to key principles or with questions about footnotes or about my own esoteric interests which I happened to mention in passing one day when half the class was absent. That will separate the wheat from the chaff, keep 'em honest, and show 'em how well I know the subject.

10. Never pre-test the test nor revise any items. If they're in the instructor's manual or the item bank, they must be good. If they're on the test they have to count. It all shakes out in the end anyway; tinkering around will only disturb natural selection.

11. Never ask students how they understand or interpret an item. If they understand it the way I do, then they're right and there is no point in asking them. If not, they're wrong -- and so, what's the point of asking anyway? What other possibilities are there?

I'm going to get technical now, so here's some terminology. An MC-item is composed of two parts: the "stem" and the "distractors" (choices). The stem can be expressed either as a question to be answered or a statement to be completed. Typically the stem calls for finding the one correct choice or finding the best choice from among those given. However MC-items can also ask which of several choices is the incorrect or worst, or which combinations of choices is correct, or which ordering of choices represents an optimal ranking. Use these rules when writing MC-item stems:

12. All questions should target at least two different objectives. Keep students on their toes! Make things interesting by not letting students know what the question is really about. I've found a measure of confusion in one's own mind helps with this rule and makes my exam questions far more challenging, even for me!

13. Be vague at crucial points and rely on subtle ambiguity. This also keeps them guessing. If they aren't sure what a question means, then I have the edge during those dreadful post-examination review sessions. I mean, if they misunderstood the question, well, whose fault was that?

14. No stems should avoid stating things negatively. Some people ask "which one is true", but that's a crude and unsophisticated way of asking "which is not untrue." Don't fear non-avoidance of double and even triple negatives.

15. Never emphasize crucial words. In a long string of "which is true" questions

try to sneak by a quick "which is false" one. Hey, if they can't read, too bad for them.

16. Be wordy, unless it clarifies something important.

17. Write so students cannot anticipate the correct answer. Make them read every distractor first, just to understand the question. We don't want them looking for the right answer just because they know it! Make correctly interpreting the question contingent on thinking about all the answers and guessing what the test writer is trying to ask.

18. Keep back information crucial to answering the question at all. This is even better than rule 17 for frustrating students, particularly the better ones -- who are, after all, our bitterest rivals.

To illustrate the points just made, consider this example MC-item stem. Note its artful ambiguity, its vacuous wordiness, and the subtle use of negatives, all designed to make the reader pause and say, "What?"

Q1: Some people think CT is one thing, but others disagree and call it something else. Which of the following false statements is not one Robert Ennis, a professor at the University of Illinois, who once conducted research at Stanford, a private college in California, and who has been known to collaborate with Stephen Norris of the University of Newfoundland, might deny?

Lest you think that the only ways to mess up MC-items has to do with question stems, here are seven ideas for undermining the distractors.

19. Provide no correct answers, or give at least two equal choices. This really keeps the good students in a quandary and gives the weaker students time to catch up. Oh, and to be sure they stick out, use "none of the above" or "all of the above" only when they are the correct choices, otherwise not at all.

20. Make distractors long, repeat words the stems could include. This will force students to take extra time to consider them and reject the wrong choices. We don't want students finishing too soon. What will our colleagues think if our exams do not require the full period?

21. Avoid homogeneous grammatical form. This gives an edge to the test-wise students who otherwise might be forced to learn something.

22. Tip off wrong answers with category mistakes. This helps students rejected wrong choices out of hand even if they are absolutely clueless regarding the right answer. As with rule 17, this is but also a small reward for the test-wise.

23. Make the wrong distractors stupid and implausible. This helps weaker students discard wrong answers and improves their chances of guessing correctly. It's a nice thing to do, considering all the stresses of exam week. Experts say that

it's hard enough to write four distractors (one correct one and three wrong ones) all of which are good enough to attract attention. They say if no students pick a given wrong distractor, then it is so obviously wrong it should be discarded from future versions of that question. Well, phooey on all that!

24. Don't put distractors in any consistent, logical order. It makes things too easy if you arrange events chronologically, sentences sequentially, or names alphabetically. Jumble things up, make students search for the right answer, maybe they'll be in a hurry and miss it.

25. Make the correct answer scholarly, detailed and precise. In case a copy of your exam is accidentally left in the copier and happens to fall into the hands of your department chair, you want to make sure she or he can recognize the quality of your scholarship. From personal experience I can assure you that a brief technical footnote documenting right answers is a sure way to impress the dean. The next example illustrates many different rules.

Q2: The First

A= thing to do is open the book.

B= person on the moon was two-legged.

C= National Savings and Loan will close soon unless the Bush Administration, supported by the Federal Reserve, finds a politically and economically acceptable long term strategy, in cooperation with major US banks and the FSLIC and FDIC, to bail out the S & L industry.

D= step toward litigation, so contact your lawyer.

E= one of the above.

Even persons well-schooled in the first 25 rules, might still underestimate their opportunities for demolishing a CT exam. I urge you to expand your repertoire by including testing tactics that make it appear they are assessing CT when they really are not. The final five rules are only for advanced test writers. In truth, they are the closely guarded secrets of the adept -- lost for centuries in unfound texts *Dune's* ancient Mentats. These five rules are so powerful that they can effectively undermine CT assessment even if some silly person violates all twenty-five of the earlier novice rules.

26. Target information recall about CT in a way that does not require students to actually use CT to come up with the right choice. This next example only looks like a CT item. In fact a person with a good memory could answer it correctly and a person with good CT who had not read the chapter could get it wrong. What more could a person want?

Q3: When a person argues that a claim must be correct simply because no one has brought up a good reason why it is wrong, that person is said to be committing the fallacy of

A= attacking the person.

B= false cause.

C= begging the question.

D= appeal to ignorance.

27. If an item actually demands some CT, complicate it by making students describe exactly what they are doing by using the proper technical vocabulary. If students cannot use the right words, how can we be sure they are thinking the right way? How can a person who never learned what "sound argument" means or how to distinguish induction from deduction, ever be expected to infer the contra-positive of a universal affirmative? What a rotten example Q4' is when compared to Q4. [I just had a dreadful thought. What would it mean for the quality of their CT if students did better on Q4' than Q4?]

Q4: "To judge the morality of an action one need only look at its consequences. Some actions have beneficial consequences; others do not. Killing an innocent person might be a great benefit to society. So, killing an innocent person can be morally correct." The passage in quotations is

A= not an argument.

B= an argument, the first sentence is its conclusion.

C= an argument, the second sentence is its conclusion.

D= an argument, the fourth sentence is its conclusion.

E= none of the above.

Q4' Consider the following passage: "(1) To judge the morality of an action one need only look at its consequences. (2) Some actions have beneficial consequences; others do not. (3) Killing an innocent person might be a great benefit to society. (4) So, killing an innocent person can be morally correct." Which sentence, if any, is presented as the main claim being supported by the others in the group?

A= None, B= (1), C= (2), D= (3), E= (4).

28. Totally ignore differences in cultures, gender-interests, domain-specific knowledge, familiarity with vocabulary, life-experiences, or any special information regarding the question content which might unfairly advantage or disadvantage a sub-group of students in their choice of answers. All that nonsense about biased questions is political claptrap anyway. That some students come to your exam having unfair advantages over other students is hardly the your responsibility. Use questions like this one:

Q5: When a stud hitter comes to the dish it would be fair for blue to

A= expand the strike zone.

B= squeeze the strike zone.

C= keep the strike zone the same.

D= Call time.

Hint 1 (not to be shared with students): The question is about baseball.

Hint 2 (Classified): In this context "fair" means treating people as they deserve to be treated, that is, differentiating between them only on the basis of relevant factors and only to an extent proportionate to the prevalence of those factors.

Hint 3 (Top Secret): "D" is wrong, and so is "A" (I think).

Hint 4 (Eyes Only): The right answer if the game is played in the USA is different than the answer if the game is played in Japan.

29. For pity sake, kill them with long tests, if you wish, but don't dare ask anything challenging. By all means avoid tough questions involving hypothetical reasoning, contrary-to-fact premises, or reasoning on the basis of explicit but questionable assumptions. It would be uncivilized of MC-testers to confront those who instinctively abhor such examinations with disconcerting evidence that MC-instruments might be more intellectually sophisticated than the weak-kneed examples paraded out whenever MC-testing needs a public whipping. Here's one type of question to avoid:

Q6: Consider the "sophonoglikipernan" relationship. It can be defined as follows: "Every human being now living has sophonoglikipernans. Nobody can be his or her own sophonoglikipernan. If someone is your sophonoglikipernan, then all of that person's sophonoglikipernans are your sophonoglikipernans too. If someone is your sophonoglikipernan, then you cannot be that person's sophonoglikipernan. Jacob and Kathy were the first humans to exist in the whole world." Which of the following must be true, if all of the above are true?

A= Either Jacob or Kathy has no sophonoglikipernans.

B= All humans are sophonoglikipernans to Jacob or Kathy.

C= Jacob and Kathy are sophonoglikipernans to each other.

D= Jacob and Kathy each are their own sophonoglikipernans.

E= There is a sophonoglikipernan who is no sophonoglikipernan to Jacob or Kathy.

30. Presume students think like the experts do. Never ask students how they figured out the correct answer was correct or the wrong ones were wrong. You don't want to know! They might be using bad CT to get the right answers or good CT but still getting wrong answers. If either of those things were happening, what would that say about the CT assessment strategy? I know many people make a big thing out of the novice/expert distinction when it comes to ways of perceiving issues and solving problems. So, let's ignore all that. Here's a case in point. As any logician can see, this next example is simply a valid deduction from two premises with a logically irrelevant sentence tossed in. Why then do more than half my students always get it wrong? What nonsense must they be thinking not to see it straight off?

Q7: Consider this group of statements: "If Adam appreciates anybody, he appreciates Barbara. There are people whom Barbara does not love, and Adam is one of them. But, everyone appreciates someone." Which of the following must be true, if all of the above are true?

A= Somebody appreciates everybody.

B= Adam appreciates Barbara.

C= Barbara appreciates Adam.

D= Barbara appreciates nobody.

E= None of the above.

There is much to learn yet about good assessment of CT. Nobody in CT

assessment wants the assessment strategy to drive the CT curriculum. Nobody wants the assessment tool to truncate the conceptualization of CT. Yet, I remain confident that when the dust has settled, plenty of opportunity will remain for those who would top off a perfectly sound concept of CT and a wonderfully thoughtful CT curriculum with a magnificently crude and ill-conceived CT examination. Toward this end these thirty rules are dedicated. If Murphy's Laws have taught us anything, it is that bad things happen. It's just that sometimes we have to work at them.

BIBLIOGRAPHY

Seriously, I would not presume to saddle any reputable scholar with the onus of being footnoted in support of any of the points in this essay. Much of what I covered is old ground anyway. Educational assessment and test construction in general are standard topics in most texts on educational research. On CT assessment in particular, Robert Ennis and Stephen Norris have published very helpful pieces. You will find many useful titles in my "A Critical Thinking Bibliography with Emphasis on Assessment," ERIC/Tests and Measurements Clearinghouse Document # TM 013578, and CT News Vol. 7, number 5, May/June 1989.