
2.

What's Wrong with the California Critical Thinking Skills Test?

Critical Thinking Testing and Accountability

Leo Groarke

It is not hard to understand why **critical thinking** (“CT”) has been proposed as a goal of education. How could one deny that students should be taught to be proficient, judicious, and open- and fair-minded thinkers? The skills that this requires — most notably, the ability to evaluate the evidence for conflicting points of view — might plausibly be identified as the core ingredient in a good education. A commitment to CT seems particularly important to democracy, because democracies rely on their citizens’ ability to reach reasonable conclusions in the exercise of their democratic rights and influence.

Though the value of critical thinking thus seems unassailable, it is not obvious how critical thinking can and should be taught. Within universities (and, increasingly, at other levels of education), disciplines such as informal logic, rhetoric, pragma-dialectics, cognitive psychology, communication studies, and education theory have developed a variety of competing approaches to “stand-alone” and/or “subject-specific” critical thinking courses and curricula. The result has been hundreds of critical thinking texts, thousands of syllabi, and a growing cache of supplemental material which includes software, websites, bibliographies, lesson plans, data bases, and extensive collections of examples.

This is a positive development, but it raises many questions. Assuming that there are more and less successful ways to teach

critical thinking (and it would be peculiar to imagine otherwise), what are the key components of successful texts and courses? Which of the many competing approaches to critical thinking is to be preferred? Should different approaches be used in different circumstances? What evidence justifies the assumption that the skills (or dispositions) we try to teach in any critical thinking course are successfully learned? How do we know that they are transferred to other contexts? Can we prove that attempts to teach critical thinking create more engaged, reflective citizens? These “critical questions” have special force in a discipline which claims that it is dedicated to reflective criticism. This is a goal which implies that those of us who teach and study critical thinking have an obligation to critically evaluate the extent to which our courses — and the curricula, texts, and theories on which they are founded — really do turn students into better thinkers.

In practice, evaluations of attempts to teach critical thinking tend to be informal: those who teach and study critical thinking form opinions on the basis of their observations and experience. One should not minimize the experience underpinning these informal impressions, but conclusions founded on them are inherently problematic. Among other things, such conclusions are frequently contradictory: teachers committed to formal logic conclude that it aids their students; teachers who reject formal logic conclude that it is a pointless exercise; and so on. It is hard to see how contradictory conclusions of this sort can substitute for a systematic and critical approach to the assessment of critical thinking — especially as they do not, in any careful way, distinguish among the different factors that may contribute to the improvement of students’ ability to think critically (e.g., CT courses, other kinds of courses, and increased maturity).

Standardized critical thinking tests are sometimes suggested as a way to navigate these problems. According to this view, they provide a more consistent and objective way of measuring the results of critical thinking courses. In the context of attempts to defend critical thinking as an educational goal, they may seem particularly important. Van Gelder has even claimed that they cast doubt on the assumption that critical thinking courses improve

students' thinking. On the basis of a review of studies using such tests, he wrote that "currently it is difficult to make a convincing case that CT/IL [Critical Thinking/Informal Logic] courses make an appreciable difference to CT or informal reasoning skills" (van Gelder 2000). In discussing the studies he reviews, van Gelder goes even further, suggesting that "an important question, which is left unresolved by these studies, is whether CT courses harm their students. It appears possible that typical CT courses actually reduce CT performance" (ibid.).

Despite his general skepticism, van Gelder does not reject all approaches to critical thinking. In defending particular approaches, one might cite studies by van Gelder et al. (2004), Hatcher (2003), and Hitchcock (2003), who have demonstrated that their courses in critical thinking improve their students' performance on standardized critical thinking tests. If it can be shown that this improvement is not plausibly attributed to other causes (e.g., increased maturity, general education), one might take this as proof that these courses successfully improve students' critical thinking skills. By studying changes in performance that occur in other kinds of critical thinking courses, one might try to assess the relative value of different courses—an intriguing idea that Hatcher develops in his chapter.

In this way, standardized critical thinking tests appear to provide us with a way to systematically study and evaluate attempts to teach critical thinking. This approach might seem to provide a ready answer to demands for educational accountability — demands that we prove that our teaching methods successfully attain our education goals. But I shall argue that this approach to the evaluation of CT raises as many questions as it answers.

The problem with standardized CT tests can be put simply: an appeal to standardized tests can settle questions about the effectiveness of critical thinking courses only if such tests are dependable instruments which measure critical thinking abilities in a valid and unproblematic way. This assumption, frequently made by those who use such tests, is problematic. In such a context, it is easily argued that standardized tests do not answer the question "Do critical thinking courses actually improve critical thinking?"

so much as they replace it with the corollary “Do critical thinking tests actually measure critical thinking?”

The difficulties inherent in the second question reflect and exacerbate the many difficulties inherent in the first. It is difficult, for example, to be sure that attempts to teach critical thinking are successful because critical thinking (and higher-order thinking generally) is a complex activity that should, if the attempt to teach it is successful, be applicable to a broad array of different contexts (indeed, to all of life). The complexity and breadth that this complexity implies are, however, even more of a problem for testing than for teaching. How can one be sure that proficiency in such a complex and broadly applicable skills set can be measured by a standardized test which must be administered in artificial circumstances governed by so many practical constraints — the limited time available for testing, ease of marking, and so on?

These problems need further study. If they are considerable and serious, then standardized testing may not be the best way to evaluate critical thinking teaching. If critical thinking is, for example, too complex to be measured by standardized tests, then the informal assessments of critical thinking — assessments based on complex human judgments carried out over an extended period of time — may, for all their problems, be a more reasonable way to judge the efficacy of attempts to teach critical-thinking (see, for example, Case 1997).

The California Critical Thinking Skills Test

Within this broader context, my goal is to assess one specific test: the *California Critical Thinking Skills Test* (the “CCTST”). Currently available from the California company Insight Assessment, the CCTST is a popular test which is available in three forms (Form A, Form B, and Form 2000) and seven languages. It is the test used by van Gelder (2000, 2004), Hatcher (2003), and Hitchcock (2003) in their studies of critical thinking courses, and it has been used by educational institutions to monitor their students’ critical thinking skills.

Each of the CCTST forms consists of 34 multiple-choice questions designed to “target those core critical thinking skills regarded to be essential elements of a college education” (Facione et al. 2002, 1). Form 2000, which I discuss, retains 22 items from the original Form A, but adds 12 new items which “require one to apply reasoning skills to contexts more appropriate to the expectations of the new century” (ibid.). Despite its popularity (and even though it must be granted that the CCTST is an historically important attempt to formulate a test that measures critical thinking skills), I contend that the CCTST is a poor instrument for testing critical thinking skills.

In defending this conclusion I argue that:

- answers in the CCTST are mistaken or unreflective;
- one can reasonably defend conflicting answers to many CCTST questions;
- the instances of reasoning the CCTST uses as a basis for its questions are vague and artificial;
- the CCTST does not recognize many essential components of critical thinking;
- the CCTST is biased in favour of an outmoded conception of critical thinking; and
- there is little reason for believing that the unproblematic questions the CCTST does contain provide even a rough measure of CT skills.

If these contentions are correct, then the CCTST cannot be used to answer the important questions I have already raised about critical thinking as a subject. At best, it is irrelevant to these questions. At worst, its continued use serves only to confuse possible answers to them.

Issues of independence

Those who create and distribute standardized tests have an ethical obligation to ensure that their instruments accurately measure what they claim to measure (all the more so when tests are used as high-stakes tests). Because test makers and test distributors have a vested interest in positive evaluations, it is difficult for them to act as neutral judges of their own tests.¹ Such an obligation can best be met through independent scrutiny and assessment. Openness to impartial test evaluation is not a criterion for a valid test, but it is a condition that needs to be satisfied before users of a test can be confident of its validity

This is a condition which the distributors of the CCTST do not meet because they have refused to make its answers available for scrutiny.² Not all refusals of this sort are unreasonable. Distributors might reasonably protect their financial investment by placing limits on such reviews (by restricting access to established researchers, requiring non-disclosure agreements, etc.) but Insight Assessment has refused to make the CCTST answers available for scrutiny even under these restricted terms. Whatever motivates this refusal, it might easily be interpreted as an attempt to prevent a critical evaluation of the test. Given the nature of the CCTST, this lack of transparency cannot prevent its evaluation,³ but it still fails to embrace an openness to critical assessment which is one important precondition for an acceptable critical thinking test.⁴

These issues are exacerbated by the way in which the creators and distributors of the CCTST have attempted to confirm its validity. They have attempted to investigate (construct) validity by studying the CCTST performance of students who complete critical thinking courses. On the basis of their finding that such students register statistically significant gains in CCTST scores, they conclude that “the CCTST proved successful as a valid and reliable measure of CT skills” (Facione et al. 2002, 20; see also Facione 1990b).

Instead of resolving questions about the CCTST (and the questions about critical thinking courses that motivate its use),

such conclusions constitute a classic begging of the question: the evidence that the CCTST is valid assuming the validity of critical thinking courses, the proof that critical thinking courses are valid assuming the validity of the CCTST. This is a circle which would have to be broken (or at least explained) before one could reasonably claim that the CCTST studies provide independent evidence for the conclusion that the CCTST is valid. Without further argument, the correlation between improved CCTST performance and the successful completion of critical thinking courses may be just as plausibly attributed to similar biases that they may share. This is a hypothesis which is not easily dismissed given that the CCTST and the courses in question have been created by individuals who share a particular approach to critical thinking (one that places, for example, great emphasis on the aspects of critical thinking that correspond to introductory formal logic).

Problem Questions, Problem Answers

Though these issues of independence are cause for concern, and though they raise serious questions about the evidence given for the validity of the CCTST, they do not themselves show that the CCTST is unreliable. In arguing that the CCTST is indeed unreliable, I want to begin with a catalogue of problems inherent in the test questions. In elaborating these problems, I will argue that the questions and answers the CCTST contains are often unreflective, sometimes mistaken (usually because they are imprecise), and founded on attempts to mimic ordinary reasoning, attempts that are artificial and ambiguous when they are presented outside the context of a more detailed description of the circumstances in which they are supposed to arise. In all the cases that follow, I argue that a critical thinker may reasonably favour a response to a question that is neither available on the CCTST nor favoured in its expected answers.

Question 1

The first question in the CCTST expects a critical thinker to conclude that the Sparklers will probably beat the Mustangs (but may lose) in a soccer match, on the basis of the knowledge that the Sparklers beat the Wildflowers and the Wildflowers beat the Mustangs (test answer B).

For a variety of reasons, this is a prediction a critical thinker should reject. First, one should recognize that the results of games are difficult to predict, especially in circumstances in which the Sparklers may have beaten the Wildflowers 3-2 in a penalty kick shootout, while the Wildflowers beat the Mustangs 1-0 on a single penalty kick. In a circumstance such as this, the teams are too closely matched to allow one to predict the outcome of their game. And all the more so given that teams in the “recreational” league in question have (according to the scenario described in the CCTST) been explicitly designed “to be evenly matched.” As anyone familiar with such leagues is bound to know, the matches they sponsor are by their nature characterized by inconsistent play and dramatic changes in individual teams, as different players show up (or not), depending on other family obligations.

Faced with the scenario the CCTST proposes, a critical thinker in a real-life situation should not draw a conclusion; instead he or she should refuse to predict the outcome of the upcoming game. Critical thinking in such a circumstance requires that one recognize that the situation is too uncertain to allow any reasonable prediction about who will certainly or even probably win the game.

Question 5

The CCTST asks us to recognize that “Ezernians tell lies” “means the same thing” as “If anyone is Ezernian, then that person is a liar.” This equivalence treats “Ezernians tell lies” as a universal statement equivalent to “All Ezernians tell lies.” This equivalence is sometimes assumed in formal logic, but it misrepresents ordinary language in which statements of the form “Xs are Ys” function as general rather than universal claims. In ordinary

language, this means that “Ezernians tell lies” claims a general truth which, unlike a universal statement, is compatible with exceptions. One might compare “The French are fond of red wine and cheese,” which is not mistaken if a few French persons do not hold these preferences (or “Lions eat meat,” which is not disproved if a vegetarian raises a lion on soy-based alternatives). In the CCTST, a generalization that is recognized as admitting of exceptions is included in question 7.

Question 8

We are asked to draw the conclusion that “Whatever else, Nero was certainly insane” on the basis of four premises:

1. Nero was emperor of Rome in the first century AD.
2. Every Roman emperor drank wine and did so using exclusively pewter pitchers and goblets.
3. Whoever uses pewter, even once, has lead poisoning.
4. Lead poisoning always manifests itself through insanity.

This is a peculiar inference on a test which purports to measure critical thinking skills because a critical thinker faced with premises such as these should not be drawing a conclusion, but should instead be asking how the premises can be justified. How could one ever know that every Roman emperor drank wine using exclusively pewter pitchers, that using pewter only once produces lead poisoning, and that such poisoning always manifests itself through insanity?

Even if we ignore the epistemological issues the above question raises, the CCTST inference cannot be justified. It is apparently founded on the notion that the conclusion of a deductive inference is always certain. This is a common misconception: deductive inferences produce conclusions which are only as certain as their premises (which in this case are notably *uncertain*).⁵ In the case in question, someone who accepts the proposed premises must certainly (on pain of contradiction) accept that Nero was insane,

but he or she needs not accept that it is *certain* that Nero was insane. If the premises are marginally acceptable but not certain (as they appear to be) then the conclusion is acceptable but uncertain.

Question 12

We are asked to draw a conclusion on the basis of data gleaned from research on preschools and the extent to which they help prepare students for kindergarten. The intended conclusion is founded on the way that students who attended preschool and those who did not attend preschool perform on a standardized test of kindergarten readiness. Those students who attended preschool scored 50-60 points, whereas those who did not attend preschool scored an average of 32 points. The CCTST concludes that “attending preschool is correlated with kindergarten readiness” (test answer E) but one could reasonably argue that more testing is needed before a plausible hypothesis can be formed (test answer B). In this regard it is significant that the students who did not attend preschool “were all from low-income households” and that the students who attended may, for all we know, be from high-income households (a distinct possibility if the preschools were located in affluent neighbourhoods). In such circumstances, it may be life in a high-income household, not preschool attendance, which is correlated with kindergarten readiness. To find out, one would have to investigate how students in preschools in low-income areas perform on the test in question.

Question 17

“Little Christopher” presses his nose against the window, wishes for the sun to come up, watches it rise, and concludes that he can make the sun come up whenever he wishes. The CCTST asks one to explain this as poor reasoning because it is an instance of the fallacy *post hoc ergo propter hoc* (test answer A). This is the answer one expects from a logic student (or professor), but it is a mistake to think that it must, therefore, be the “best” way to explain what is wrong with the reasoning. If one wants to explain

to little Christopher's friend, Jamie, why the reasoning is wrong, one will do better to point out that the world goes around the sun with or without Christopher's wishing it (test answer B). One can even imagine contexts in which one could plausibly argue that Christopher's reasoning is good because he is "only a child" (test answer C): one might argue that, despite his erroneous conclusion, it is significant that someone as young as Christopher has recognized that causal conclusions should, in some crucial way, be founded on an observed correlation between a cause and an effect.

Question 19

We are told that there are "two popular arguments in favour of the death penalty" The problems with one of the arguments are explained and the test-taker is asked to evaluate the reasoning. But one might easily object that this is difficult to do without knowing more about one's goals in arguing. If one's goal is the argument that is the most philosophically defensible (the traditional goal of logic), one might lean in one direction. If one's goal is to convince an audience (the traditional goal of rhetoric) then one might lean in another direction. If one imagines oneself at a philosophy conference where one is trying to establish the morality of the death penalty, one might reasonably object to question 19's focus on popular arguments in favour of the death penalty (test answer A). In such a context, one might argue that the popularity of an argument is irrelevant.

One might evaluate the argument in a different way if it is propounded by a politician in the context of an upcoming referendum on the death penalty—a circumstance where popular opinion (even if misguided) is an appropriate focus of attention. In these new circumstances, one might argue that the reasoning is poor because only one of two popular arguments has been addressed (test answer B). In yet another context—say, a conversation with a group of social scientists (who typically reject the deterrence argument)—it might not matter that a popular argument based on deterrence is mentioned but not addressed. In

this context, one might argue that the argument is a good argument (test answer C).

Question 23

We are provided with a list of height relations (L is shorter than X, Y than L, M than L, M than Y) and asked what information “must” [the test’s emphasis] be added to require that Y is shorter than J. Of the answers given the only possibility is C (“J is taller than L”) but J could be shorter than L and still taller than Y — if, to take one example, L is 5’, X is 6’, Y is 4’, M is 3’ and J is 4.5’. Thus, it is not true that the information “J is taller than L” must be added to imply that Y is shorter than J. There are many possibilities one could add (for example, “Z is taller than Y and shorter than J”). In this particular case, it appears that the CCTST question is misstated. It should ask: “Which of the following would imply that Y is shorter than J?” This question would require the intended answer. Though precision is one of the hallmarks of critical thinking, the CCTST mistakenly treats it as equivalent to the question “What information must be added to make this true?”

Question 24

A paragraph of reasoning begins with the sentence “A standard deck of 52 playing cards contains exactly four kings, four queens, and four jacks” and ends with the sentence “So, from what we know now, we can conclude that among the 52 playing cards in a standard deck, there are precisely four each of jacks, queens, and kings.” According to the CCTST, the reasoning is “poor” because “It proves nothing, as in ‘The sky is blue because it’s blue’ (test answer A).

But the claim that reasoning in the paragraph has the form “The sky is blue because it’s blue” is contentious. The latter is an inference of the form “A, therefore A.” The reasoning in Question 24 has the form “A, B, C, D, E, therefore A.” These are importantly different inferences. In one, the conclusion repeats the premise; in the other, the conclusion is deduced from a list which

contains it. It is difficult to think of plausible inferences of the form $A \vdash A$ (I don't doubt that there are some), but it is not difficult to think of examples of the form $A, B, C, D, E \vdash A$.

The latter, for example, is the form of inference I use when checking a grocery list to deduce what should be put in the shopping cart. In other situations, such an inference might be appropriate when teaching deductive reasoning or when dealing with children, or in other cases where one needs, in painstaking ways, to make things clear; or when the passage in question is one part of a long argument in which it is particularly important to recognize that there are four of each face card in a standard deck of cards (i.e., a circumstance in which it makes good sense to repeatedly reinforce an audience's commitment to this proposition).

It is true that the argument in question is circular, but it cannot be dismissed on these grounds. The same can be said of all good deductive arguments — which might be approved, not rejected, because they are (as test answer B explains) inferences in which “the reasoning is an accurate restatement of the facts.”

Question 33

In a situation in which an assistant fails to send an important package, we are asked to judge a friend's argument that there are (setting aside union issues) sufficient reasons for firing him: “He has lied. He is disorganized and loses important things. He did not even check with you about sending the package late once he found it.” One could argue that the reasoning is “good, because the assistant has performed in exactly these substandard ways” (test answer D). It is plausible to suppose that someone should be fired if he has acted in these ways.

However, one can imagine contexts in which it is more plausible to conclude that the friend's reasoning is “poor, because the friend does not know the circumstances of work in your office” (test answer A). Imagine a situation where the assistant who has misbehaved has a long record of superior performance and his unhelpful behaviour can be attributed to difficult circumstances

that require some compassion (e.g., his father has died, his teenager is in trouble with drugs, etc.).

Someone who reflects on the vicissitudes of human conflict may reasonably argue that one can never understand a situation of this sort until one has heard “both sides of the story.” But this suggests that the right answer to Question 33 is B: that the friend’s reasoning is poor because he or she has not given the assistant a chance to explain himself.

Faced with Question 33, how can the critical thinker choose between answers A and D and possibly B? On the one hand, one might reasonably suppose that all the essential information has been given in the test question, and that one should not imagine further complicating circumstances (a supposition that favours answer D). On the other hand, one might reasonably hypothesize that the CCTST is designed to test one’s care in reasoning, and that in this instance it is testing one’s ability to recognize that complicating circumstances have not been explicitly ruled out.

Question 34

The same kinds of problems are evident in Question 34, which refers again to the misbehaving assistant. In this case, we are asked to imagine that our daughter elaborates the argument that “If you fire your assistant you will get in trouble with the union; but if you do not, you will get in trouble with your boss! No matter what, you will get in trouble eventually.” This is reasonably judged to be a good dilemma argument “because right now there seem to be no other options” (test answer C). It is, however, possible to make a case for rejecting one of the conditionals in the dilemma, i.e., the claim that “if you fire your assistant you will get in trouble with the union.” This is not explicitly stated in the CCTST’s original description of the situation. One might say on these grounds that the reasoning is poor “because you cannot be sure what the union will do” (test answer B).

Without more information, it is difficult to choose between answers B and C. On the one hand, this is the kind of contract violation that is likely to precipitate a union grievance. On the

other hand, violations of a union contract may not result in grievances (because the individual affected does not wish to pursue a grievance, because the union leadership decides not to pursue it, and so on). There is no way to tell what should be expected in this particular case.

Why Such Problematic Questions?

Putting aside the problems with specific questions, the CCTST might be criticized for its commitment to artificial examples of reasoning that are, at best, distantly related to the kinds of reasoning or critical thinking required in real-life contexts. Within the CCTST, this artificiality is reflected in premises that are fanciful (“*Whenever* it is snowing, streets and sidewalks are wet and slippery,” “*All* college students graduate sooner or later,” etc.); in arguments presented out of context; and in inferences that are embedded in scenarios which are described in a manner that does not provide the details necessary to properly assess them.

In such contexts, the CCTST asks us to judge arguments and explanations without knowing to whom they are addressed, what circumstances prompted them, and the argumentative details of the situation in which they are advanced. In these and many other cases (consider Questions 3, 6, and 16), one may wonder whether the examples that form the basis of CCTST questions can reasonably be used to test one’s ability to think critically in the “complex and many layered” situations that demand real-life reasoning. Why should we believe that an ability to answer the CCTST’s artificial questions shows that someone can think critically about politics, his or her favourite television show, advertising on the Internet, a business proposition, ethics, and so on? What compelling evidence shows that this is so?

Anyone familiar with the development of critical thinking and its related disciplines will recognize that the artificiality that tends to characterize the examples in the CCTST reflects the artificiality that characterized early attempts to teach logic in a manner suited for general students (notably, in early editions of Copi: see, for

example, Copi 1961). In both cases, the attempt to teach reasoning skills is characterized by constructed rather than actual examples of reasoning; focuses on answers that only reflect aspects of thinking explained in terms of the limited resources available in propositional and syllogistic logic; and emphasizes many of the simplest kinds of inference making to the exclusion of many more complex aspects of ordinary reasoning (e.g., questions of premise acceptability and more complex kinds of inference).

In the wake of developments in informal (and even formal) logic, critical thinking, and related disciplines, this approach to critical thinking reflects an outmoded conception of critical thinking which has been roundly criticized (for an overview of some of the standard criticisms, see the articles by Johnson and Blair in Johnson 1996). There is therefore little reason to believe that test questions reflecting the CCTST's limited conception of reasoning can validly measure critical thinking abilities as they are understood when critical thinking is proposed as a goal of education — a goal that implies the ability to think critically in the midst of the complexities and nuances that characterize reasoning in real-life contexts.⁶

What's Missing From the CCTST?

The issues raised by the artificial examples in the CCTST suggest that it fails to test one's ability to deal with many of the complexities that characterize critical thinking in real-life situations. It is reasonable, then, to ask whether key critical thinking competencies are missing from the CCTST. In attempting to answer this question, something must be said about the definitions of critical thinking, because it is one's definition of critical thinking that determines what competencies and complexities critical thinking must encompass.

The CCTST is based on the definition of critical thinking proposed in the **American Philosophical Association's 1990 "Delphi" report** (*Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and*

Instruction). It identifies six core critical thinking skills: interpretation, analysis, evaluation, inference, explanation, and self-regulation; and defines critical thinking as the “purposeful, self-regulatory judgment which results in interpretation, analysis, evaluation and inference, as well as explanation of the evidential, conceptual, methodological, criteriological or contextual considerations upon which that judgment is based” (Facione 1990a, 2). The Delphi Report associates each of the six core skills identified in this definition with a specific set of sub-skills.⁷

A detailed discussion of the Delphi definition — or the general issues raised by any attempt to define critical thinking — lies beyond the scope of this chapter.⁸ In place of such a discussion, it will suffice to note that the different definitions of critical thinking that have been proposed recognize it as an ability (or set of abilities, or set of abilities and dispositions) applicable to a broad array of real-life contexts. When those of us who champion critical thinking say that we want students to be critical thinkers, we mean that we want them to be individuals who critically evaluate the claims, beliefs, arguments, attempts at persuasion, etc., that surround them in the many different facets of their lives: when they argue in class; when they watch television; when they read magazines, newspapers, and books; when they participate in formal and informal conversations; when they graduate and pursue professional careers; and so on.

This aspect of critical thinking raises an obvious question about the CCTST: Does its interpretation of the Delphi definition⁹ encompass the essential skills and competencies that characterize critical thinking in a broad array of real-life contexts? In answering this question, one might usefully compare the understanding of ordinary reasoning implicit in the CCTST and that evident in current research coming out of disciplines usually associated with critical thinking (which are often referred to as the interdisciplinary amalgam of disciplines and sub-disciplines called “argumentation theory”). In the last twenty years such argumentation theories have made great progress in the attempt to establish and extend a more sophisticated understanding of informal argument, discussion, dialogue, and debate. It is

significant that they are, in marked contrast with the CCTST, characterized by both a much clearer focus on real, rather than concocted, examples of critical thinking, and a much more sensitive account of the nuances and complexities of real-life reasoning.

Though space does not allow a detailed account of the understanding of critical thinking that has emerged in argumentation theory (for an overview, see Johnson 1996; Groarke 2002; van Eemeren 2002), I will note that its scope encompasses, among others, the following elements:

- the principles of argumentative communication that inform critical inquiry;
- the different expectations that govern dialectical exchange in different kinds of circumstances (see, e.g., Van Eemeren 2002);
- techniques of persuasion, bias, and the relationship between argument, audience, and ethos (see, e.g., Tindale 1999, 2004);
- an in-depth understanding of fallacies and argument schemes which play a central role in ordinary reasoning (see, e.g., Walton 1992, 1998; Hansen and Pinto 1995);
- the dialectical obligations that attend arguments in real-life contexts (see Johnson 2000, and in this volume); and
- the nature of visual argument and persuasion that surround us on television, in advertising, and on the Internet (see Groarke 1996; Blair 2003).

These aspects of reasoning, which have been shown to play a crucial role in reasoning in real-life contexts, are conspicuously absent from the CCTST, which has no questions that would allow us to measure a thinker's ability to evaluate real-life problems appropriately or to make sound decisions about what to believe or do. Even if there were no problems with the questions and answers assumed in the CCTST, the failure to recognize and test for such

abilities would make it difficult to accept that this particular test can function as a reliable measure of critical thinking skills.

Some Concluding Comments

The ruminations in this chapter leave little room for confidence in the CCTST's ability to reliably measure critical thinking skills. The test is problematic in many ways. Most notably, it contains many contentious answers, relies on artificial examples which are removed from the real-life contexts where critical thinking must take place, fails to recognize key aspects of ordinary reasoning that play a role in critical thinking, and focuses on rudimentary reasoning skills which represent a very limited conception of critical thinking. For these reasons, it is difficult to defend the use of the CCTST as a way to test critical thinking abilities and, more broadly, to teach these skills.

It would be premature to conclude that reliable tests of critical thinking are impossible. The problems with the CCTST highlight the many nuances and complexities of ordinary reasoning that make the design of a good test difficult. That said, other tests (like the Ennis-Weir) and other approaches to testing (like Fisher and Scriven's multiple rating items) must be judged on their own merits. More significantly, perhaps, we should not prejudge attempts to create better tests because it is possible that they will provide valuable instruments that will allow us to study and understand attempts to teach critical thinking. More study and discussion will have to determine the extent to which testing can adequately measure the complex and difficult aspects of critical thinking (e.g., what the Delphi Report calls "self-regulation"). Such work will be worthwhile even if it reaches negative conclusions because it will still clarify the nature of critical thinking teaching and assessment.

In the meantime, it should be said that the value of standardized critical thinking tests is easily exaggerated. My attempt to answer the questions about critical thinking stated at the beginning of this chapter has not shown that standardized tests provide a better

measure of critical thinking abilities (and the efficacy of critical thinking courses) than the informal assessments that have characterized the field. We need to remain open-minded, but we should also be wary of the kind of standardized testing Giancarlo-Gittens warns about in Chapter One of this book — tests often used in high-stakes situations that have major ramifications for students and teachers, and for the development of critical thinking as a field.

References

- American Philosophical Association. 1990. *Critical thinking: A statement of expert consensus for the purposes of educational assessment and instruction*. ("The Delphi Report"). ERIC Doc. No. ED 315-423.
- Blair, J. 2003. The rhetoric of visual arguments. In *Defining visual rhetorics*, ed. C. Hill and M. Helmers, 41-62. Mahwah, NJ: Lawrence Erlbaum Associates.
- Case, R. 1997. Principles of authentic assessment. In *The Canadian anthology of social studies: Issues and strategies for teachers*, ed. R. Case and P. Clark, 389-400. Vancouver: Pacific Educational Press.
- Copi, I. 1961. *Introduction to logic*. New York: Macmillan.
- Dumke, G. 1980. *Chancellor's executive order 338*. Long Beach, CA: California State University.
- Ennis, R., and E. Weir. 1985. *The Ennis-Weir critical thinking essay test*. Pacific Grove, CA: Midwest Publications.
- Facione, P., N. Facione, S. Blohm, and C. Giancarlo. 2002. *The California critical thinking skills test*. Millbrae, CA: Academic Press/Insight Assessment.
- Facione, P. 1990a. *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction. Executive summary "The Delphi Report."* Millbrae, CA: Academic Press.
- _____. 1990b. *The California critical thinking skills test: College level technical report #1-Experimental validation and content validity*. Millbrae, CA: Academic Press. ERIC Doc No. ED 327-549.
- Fisher, A., and M. Scriven. 1997. *Critical thinking: Its definition and assessment*. Norwich, UK: Centre For Research In Critical Thinking, University of East Anglia.
- Groarke, L. 2021. Informal logic. *Stanford Encyclopedia of Philosophy*. Online. Available at <http://plato.stanford.edu/entries/logic-informal>.

- _____. 1999. Deductivism within pragma-dialectics. *Argumentation* 13: 1-16.
- _____. 1996. Logic, art and argument. *Informal Logic* 18(2 and 3): 116-31.
- Hansen, H., and R. Pinto. 1995. *Fallacies: Classical and contemporary readings*. University Park: Pennsylvania State University Press.
- Hatcher, D. 2003. On assessing and comparing critical thinking programs: A response to Hitchcock. Paper presented at *Informal Logic @ 25*, University of Windsor, May 14-17. Online. Available at <http://www.humanities.mcmaster.ca/~hitchckd/response.htm>.
- Hitchcock, D. 2003. The effectiveness of computer-assisted instruction in critical thinking. In *Informal Logic at 25: Proceedings of the Windsor Conference*, ed. J. Blair, D. Farr, H. Hansen, R. Johnson, and C. Tindale. CD-ROM. Windsor, ON: Ontario Society for the Study of Argument.
- Johnson, R. 2000. *Manifest rationality: A pragmatic theory of argument*. Mahwah, NJ: Lawrence Erlbaum Associates.
- _____. 1996. The rise of informal logic: Essays on argumentation, critical thinking, reasoning and politics. *Studies in Critical Thinking and Informal Logic* No. 2. Newport News, VA: Vale Press.
- Paul, R., and L. Elder. 2001. *Critical thinking: Tools for taking charge of your learning and your life*. Upper Saddle River, NJ: Prentice-Hall.
- Tindale, C. 2004. *Rhetorical argumentation*. Thousand Oaks, CA: Sage Publications.
- _____. 1999. *Acts of arguing: A rhetorical model of argument*. Albany: SUNY Press.
- van Eemeren, F., ed. 2002. *Advances in pragma-dialectics*. Amsterdam and Newport News, VA: Vale Press.
- van Gelder, T. 2000. *The efficacy of undergraduate critical thinking courses: A survey in progress*. Online. Available at <http://www.philosophy.unimelb.edu.au/reason/efficacy.html>.

- van Gelder, T., M. Bissett, and G. Cumming. 2004. Cultivating expertise in informal reasoning. Special issue on informal reasoning. *Canadian Journal of Experimental Psychology*.
- Walton, D. 1998. *Appeal to popular opinion*. University Park: Pennsylvania State University Press.
- _____. 1992. *Slippery slope arguments*. Oxford: Clarendon Press.

Notes

1. As Paul and Elder (2003) recognize, vested interests of this sort are one of the major obstacles to critical thinking, and manifest themselves in a natural tendency to "think of the world in terms of how it can serve us" (214).
2. I personally discussed this issue with Insight Assessment (the test distributors) on two occasions, asking them for the official answers. I purchased the test packet and explained that I would only use the answers to assess the test, but they would not release the official answer key.
3. Many of my criticisms (for example, that the CCTST is founded on questions which are vague, founded on mistaken assumptions, and susceptible to different interpretations) hold no matter what answers one proposes. That said, most of the questions on the CCTST have obviously intended answers that will be evident to anyone who knows the field. In order to deal with a few cases about which I was unsure I consulted with researchers who worked on the original test.
4. One might question whether, as a matter of standard practice, the critical thinking community should use any test which is not made available for independent assessment.
5. The intended answer illustrates a fallacy of misplaced modality which often characterizes assessments of deductive arguments. It mistakenly assumes that the conclusion of a deductive argument is certain. It should instead be said that a deductive argument, with premises P and conclusion C , has the form $P \vdash C$, and establishes (only) that the conclusion is as certain as the premises. (Groarke 1999)
6. The more one is sensitive to the different aspects of real-life reasoning (context, audience, premise acceptability, etc.), the more the questions on the CCTST must strike one as puzzling, peculiar, and open to

different interpretations. Especially in view of its time constraints, one will score better on the CCTST if one ignores the nuances of good reasoning (as formal logicians sometimes do) and answers questions without the reflection they invite.

7. Interpretation, for example, is defined as the ability "to comprehend and express the meaning or significance of a wide variety of experiences, situations, data, events, judgments, conventions, beliefs, rules, procedures or criteria" and said to include as sub-skills "categorization," "decoding significance," and "clarifying meaning" (Facione 1990a, 6-7).
8. For an overview of these issues, see Fisher and Scriven (1997); see also the discussion in the Introduction and Chapter Three by Ralph Johnson in this book.
9. The CCTST interpretation is only one of many possibilities and one that might be criticized in many ways. One aspect of the Delphi definition, for example, is its commitment to "self-regulation." Putting aside the question of whether it is a disposition rather than a skill, self-regulation encompasses a willingness to critically examine and re-examine one's beliefs. There is no doubt that regulation of this sort is a cornerstone of critical thinking, but it is difficult to see how it can be tested in a test like the CCTST. In circumstances in which we wish to establish the extent to which someone is committed to an open-minded examination of their beliefs, we need to observe their willingness to engage criticisms of these beliefs, their response to countervailing evidence, and so on. These are skills and dispositions that are not tested by the CCTST, which functions as a more general test of reasoning skills. The difference between reasoning skills and self-regulation is evident in individuals who have sophisticated reasoning skills but are dogmatic about their beliefs.