# A Serious Flaw in the Collegiate Learning Assessment [CLA] Test

**KEVIN POSSIN**

*Department of Philosophy*
*The Critical Thinking Lab*
*Winona State University*
*Winona, MN 55987*
*USA*
*kpossin@winona.edu*

**Abstract:** The Collegiate Learning Assessment Test (CLA) has become popular and highly recommended, praised for its reliability and validity. I argue that while the CLA may be a commendable test for measuring critical-thinking, problem-solving, and logical-reasoning skills, those who are scoring students' answers to the test's questions are rendering the CLA invalid.

**Résumé:** Le Collegiate Learning Assessment Test (CLA), loué pour sa fiabilité et sa validité, est devenu populaire et fortement recommandé. Je soutiens que, bien que le CLA puisse être un test louable pour mesurer la pensée critique, la résolution de problèmes et les compétences logiques, ceux qui corrigent les réponses des élèves aux questions du test rendent le CLA non valide.

**Keywords:** Collegiate Learning Assessment Test, CLA, critical-thinking assessment, critical-thinking skills, reliability, rhetoric, validity

## 1. Introduction

The Collegiate Learning Assessment (CLA) test has received a lot of great publicity. It was even featured in a marvelous *Doonsebury* cartoon. The Spellings Commission on the Future of Higher Education (2006) suggested using the CLA as a means of achieving better "accountability" in higher education—to ensure no undergrad left behind, so to speak. Arum and Roksa, in *Academically Adrift* (2011), used the CLA at 24 community colleges, colleges, and universities to measure students' improvement in critical thinking, reasoning, and writing skills. They discovered alarmingly meager average gains in those skills in the first two years of college—only .18 of a standard deviation—with 45% of the students exhibiting no statistically significant gains at all (p. 35).

These are just some of the hefty claims made using students' results on the CLA as evidence. So just how good is this test at measuring what its authors at the Council for Aid to Edu-

cation (CAE) claim it measures, viz., the higher-order skills of critical thinking, analytic reasoning, problem solving, and written communication?

## 2. Critical review of the CLA scoring

One cannot fault the CLA's focus here. The CAE is correct in saying that these skills are part of our "common learning," "critical in the knowledge economy," and "core to virtually all mission statements of colleges and universities" (Benjamin et al., 2009, p. 2). The Higher Education Research Institute (2009) found that professors agree, with 99% believing critical thinking is "very important" or "essential," and 87% believing similarly about effective writing.

How does the CLA measure the acquisition and enhancement of these crucial cognitive skills? By means of open-ended, real-world, performance-based tests. The CLA is claimed to assess these skills *holistically*, unlike multiple-choice assessment tests, which attempt to "define critical thinking as a discrete set of sub-skills that can be broken out separately, and then arranged along a series of dimensions." To which the CAE asks skeptically: "What are those constituent parts of critical thinking, and can problem solving be broken down into smaller, manageable pieces?" (Benjamin et al., 2009, p. 22).

There are three formats for the CLA: one Performance Task and two Analytic Tasks—Make-an-Argument and Critique-an-Argument (formerly called Break-an-Argument). All three Tasks are designed to measure how well students evaluate and analyze information and draw conclusions on the basis of that analysis. The CAE has posted the rubrics or criteria it uses in scoring student performance in each task type (CAE, 2011a). For example, students are scored on the basis of how well they assess the relevance and strength of evidence, recognize flawed arguments, recognize logical flaws (e.g., mistaking mere correlation for causation), construct cogent arguments, select the strongest evidence in support of conclusions, critically review alternative positions, and recognize that a problem is complex and lacking a clear answer (Benjamin et al., 2009, p. 42).

To see how successful the CLA is at its goal of measuring these critical-thinking skills, let's begin by examining the Performance Task, the most prominent of the three CLA Tasks—it is the one used by Arum and Roksa, in *Academically Adrift* (2011), and recommended by the Spellings Commission (2006).

On the right side of their (locked-down) computer screens, students are given access to a Document Library, consisting of

various sources of information, such as letters, research reports, newspaper clippings, diagrams, tables, and charts, which students are to use in preparing their answers to questions that appear on the left side of their screens along with a response box, into which the students have 90 minutes to key their answers.

Here is a portion of the current example, titled "Crime Reduction," provided by the CAE in their updated promotional materials:

> Pat Stone is running for reelection as Mayor of Jefferson, a city in the state of Columbia. Mayor Stone's opponent in this contest is Dr. Jamie Eager. Dr. Eager is a member of the Jefferson City Council. You are a consultant to Mayor Stone.
>
> Dr. Eager made the following three arguments during a recent TV interview: First, Mayor Stone's proposal for reducing crime by increasing the number of police officers is a bad idea. Dr. Eager said "it will only lead to more crime." Dr. Eager supported this argument with a chart that shows that counties with a relatively large number of police officers per resident tend to have more crime than those with fewer officers per resident….
>
> Mayor Stone has asked you to prepare a memo that analyzes the strengths and limitations of each of Dr. Eager's three main points, including any holes in those arguments. **Your memo also should contain your conclusions about each of Dr. Eager's three points, explain the reasons for your conclusions, and justify those conclusions by referring to the specific documents, data, and statements on which your conclusions are based.** (Benjamin et al., 2009, p. 47)

The first assigned question the student addresses is about Dr. Eager's claim that hiring more police "will only lead to more crime," based on Dr. Eager's chart. Students are scored on the basis of either (1) agreeing that more police *are* causing more crime, (2) suggesting that "more crime **might** necessitate more police," (3) saying that mere correlation does not imply causation or that the relation could go either way, or (4) offering a possible common cause. Only the first of these options is treated as incorrect; the other three possible answers are treated as correct, but *must* be stated in terms of *uncertainty* (Benjamin et al., 2009, p. 50).

On the one hand, I am very glad to see that the first answer (viz., agreeing with Dr. Eager) is being treated as just plain wrong. When I began researching the CLA (Possin, 2008), I was struck by how any answer was accepted so long as the writer offered some reason for it, no matter its justificatory power or

lack thereof—students were told to "Address the issue from any perspective—no answer is right." This invited sophistry instead of critical thinking; *rationalization* instead of justification. My fears were explicitly confirmed at that time by Marc Chun, CAE Director of Product Strategy, during a Web conference. These fears are not totally removed today, however, because of how the CAE says it develops its Performance Tasks: "[C]are is taken to ensure that sufficient information is provided to permit multiple reasonable solutions…to ensure that students could arrive at approximately three or four different conclusions based on a variety of evidence to back up each conclusion. Typically, some conclusions are designed to be supported better than others" (Benjamin et al., 2009, p. 40).

It's not clear, however, that students *are* being properly assessed by the graders for recognizing such differences in degrees to which conclusions are supported by data. According to the CAE, "'Might' is a key word here; the student should express uncertainty rather than a certainty in the explanation" (Benjamin et al., 2009, p. 50) of the correlation between the size of the police force and the frequency of crime. But "uncertainty" is not good enough: For students to say merely that Dr. Eager *might* be wrong, or that there *might* be some other relation explaining the correlation, is platitudinous—this is an *inductive* case, after all, so error is, *by definition, always logically possible*. One needs to offer a *more likely* or *plausible* alternative explanation for this correlation, e.g., that the increase in crime has caused the hiring of more police. Likewise, if the student offers a common-cause hypothesis, it must be plausible and not just some far-flung possibility.

Let's now examine the Analytic Task, Make-an-Argument. Students are given a prompt. The example currently provided in the CAE's promotional materials is, "Government funding would be better spent on preventing crime than in dealing with criminals after the fact" (Benjamin et al., 2009, p. 53). Students have 45 minutes to take any position on the topic and argue for it. The critical-thinking criteria used in scoring student responses are: "Clarifying a position and supporting it with evidence, considering alternative viewpoints or counter points to their argument, developing logical, persuasive arguments, [and exhibiting] depth and complexity of thinking about the issues raised in the prompt" (Benjamin et al., 2009, pp. 53-54).

All's well so far, with respect to this Task's goals and rubric. A problem appears, however, when we look closely at what the CAE presents as an exemplary "high quality response" and its "characteristics." Here is that student response (Benjamin et al., 2009, pp. 54-55), with my critique, paragraph by paragraph.

Government imposes order upon its citizens to pursue generally agreed-upon goals in society. An important function of American government, for example, is to protect the "life, liberty and the pursuit of happiness" of its citizens, a premise upon which the U.S. was founded more than two centuries ago. Guaranteeing this "inalienable right" through government action is easier said than done. In general, government does so by collecting taxes, enacting laws, and enforcing laws consistent with goals. Violating these laws, by definition, are crimes and the people who commit crimes are criminals. But the meaning of laws and the causes of crime are complicated. In all, there is no simple formula for investing taxpayer dollars and the statement oversimplifies the challenge of dealing with crime. While investing public dollars in crime prevention may have certain advantages, it is not necessarily "better spent" than "dealing with criminals after the fact."

If this last sentence is the student's statement of position, it's a very wimpy one: money spent on crime prevention *may* have its advantages but it *might* be better spent on incarcerating criminals. That's a platitude again, given that the position is an empirical claim. And the rest of the paragraph is just padding.

Laws are reflections of moral beliefs of society, that is, what we collectively believe to be right or wrong. These beliefs often change over time, and even by communities within broader society. Furthermore not all laws, or crimes, receive the same levels of enforcement. For example, while we might universally agree that certain violent acts (e.g., murder, rape, armed robbery) are indeed crimes that ought to be prevented at high dollar cost, we might not agree that others (e.g., underage drinking, jaywalking) deserve the same attention. And certain laws which may have been important at the time or in the jurisdiction where they were written, may no longer be relevant, although they may remain on the books. Given different interpretations, severity and changing nature of crime, it might be quite difficult (and costly) to create a program that effectively prevents crime in all its variety. Doing so would run the risk of addressing those crimes that either do not pose significant threat to "life, liberty and the pursuit of happiness" or, in the future, are no longer crimes at all. By contrast, dealing with criminals after the fact has the advantage of focusing resources on those who have indeed violated existing laws in society, in particular those laws society has chosen to enforce.

> This approach also allows society to reconsider laws for relevance in present-day society (i.e., through the courts) as violations occur, so that criminal behavior may be re-defined as concepts of morality may change.

This paragraph attacks a strawman: *No one* advocating a focus on crime prevention takes the position that money should be spent on preventing *all* crimes *equally, no matter their severity, with no consideration of their relevance*. This also sets up a false dichotomy: If one can't prevent *all* crimes, then one should focus instead on violations as they occur. Furthermore, the student's argument for preferring incarceration applies *equally* as well to the opponent's position, crime prevention—*both* approaches allow us to "reconsider laws for relevance in present-day society."

> Furthermore, preventing crime requires that we understand why crimes occur, so that we may know how to intervene. But crime is complex, stemming from many, many conditions pertaining to society and its members. These factors may divide along lines of the classic debate in biology over "nature vs. nurture" as determinants of behavior. Interpreting crime in this way, we might ask: Are criminals the result of the influence of their environment? Or are criminals born to commit crimes? If criminals are products of their environment, then crime prevention programs should address root causes of crime in society. But what are these root causes, and can they be disentangled from a combination of other factors? Are all people susceptible to the same causes, or does a crime prevention program need to accommodate all individual differences so that none will become criminals? Investing in a comprehensive crime prevention program that addresses all causes and all individuals would appear to be a costly proposition. It is difficult to imagine a program that could effectively do so, at any cost. Furthermore, addressing a root cause of crime would likely trigger a series of other causes that would need to be addressed. If, for example, robbery is related to high incidence of poverty and drug abuse, then crime prevention requires effective programs to address problems of poverty and substance abuse. But these, too, are complex problems related to issues of education, discrimination, mental health, and so forth. Where would the crime prevention program (and government investment) stop? By contrast, according to the "nature" argument, criminals are social deviants from birth. Addressing crime becomes a simple matter of identifying these individuals and removing them from society according to the crimes they commit, with-

out any need to address social or environmental concerns. So long as the number of criminals is few, the cost of separating these individuals from society (e.g., by sending them to prison) will also be relatively small, and government funding might be "better spent" on this approach.

This paragraph is plagued by *two* false dichotomies: (1) that the focus on crime prevention must either "address all causes and all individuals" or it's not worth doing and (2) that criminality is *exclusively* either nature or nurture. The slippery-slope fallacy is also committed, when the student asks where crime-prevention programs would stop, mistakenly implying that if one cannot draw the perfect line, one is forced to apply those programs absurdly to "all causes and individuals." Finally, no argument at all is given for the claim that incarceration is less costly than crime prevention—in fact, evidence often indicates otherwise—it is merely asserted under the *assumption* that "the number of criminals is few." Under *that* assumption, however, crime *prevention* would be less costly too.

But my understanding is that the "nature vs. nurture" argument rages on, leading me to believe that neither determines an individual's behavior by itself. Sending individuals to prison, because they were born criminals, assumes that these people cannot become productive members of society. It denies these individuals their own "inalienable right," a reason many have come to America in the first place. Whether or not this is the case, keeping these individuals imprisoned assumes further that laws, and therefore the definition of crime, never changes. Unjust imprisonment in the name of dealing with criminals can never be government funding "better spent" in the United States.

This is another strawman fallacy: *No one* in favor of crime-prevention programs is arguing for incarcerating people prior to or beyond their committing a crime, such as in the movie *Minority Report*, in which clairvoyants nearly perfectly reliably predicted crimes so as to prevent them.

Neither investment in crime prevention nor investment in dealing with criminals by themselves can easily address the problem of crime in our society. Instead, some combination, along with investments in other societal improvements will be required to address problems of crime. More generally, how government funding should be spent to address the complex challenge of pro-

tecting citizen's rights to "life, liberty and the pursuit of happiness" is best determined by the continued interaction of lawmakers, law enforcement officials, the courts, and the citizenry, just as it has for more than 200 years.

The student *finally* takes a real position: viz., that a combination of prevention and incarceration, along with other "societal improvements," should be used. This is a substantive and reasonable position, which *should* have been stated in the *first* paragraph *and which is never argued for*. The student, however, immediately *reneges* on this position by stating that government spending on the issue of crime should simply be left to legislators, law enforcement, the courts, and the voters. Wait a second; the instructions said *you* should take a position and argue for it!

I'm sorry to say that after reading this student's response, I felt bullshitted (Frankfurt, 2005). And after reading papers for 26 years, I think I know it when I see it. It read so well, didn't it? But it was just rhetoric, not rational argumentation and criticism.

Lastly, let's examine the Analytic Task, Critique-an-Argument. The student is instructed to critically review the argument in the prompt. The updated example presented in the CAE's promotional materials is as follows.

> The number of marriages that end in divorce keeps growing. A large percentage of them are from June weddings. Because June weddings are so popular, couples end up being engaged for a longer time just so that they can get married in the summer months. The number of divorces gets bigger with each passing year, and the latest news is that more than 1 out of 3 marriages will end in divorce. So, if you want a marriage that lasts forever, it is best to do everything you can to prevent getting divorced. Therefore, it is good advice for young couples to have short engagements and choose a month other than June for a wedding. (Benjamin et al., 2009, pp. 58-9)

An interesting problem arises immediately with the rather elaborate instructions given to the student in this Task (to a lesser degree, this problem plagues the other two Tasks as well). The student is told to:

Discuss:
- Any flaws in the argument
- Any questionable assumptions
- Any missing information
- Any inconsistencies

…You will be judged on how well you do the following:

1. Explain any flaws in the points the author makes
2. Organize, develop, and express your ideas
3. Support your ideas with relevant reasons and/or examples
4. Control the elements of standard written English (Benjamin et al., 2009, p. 58)

By explicitly telling the students to exhibit these aspects of critical thinking, their *disposition* to exhibit those critical-thinking skills *on their own* is not being tested. I am sympathetic, however, with the dilemma the CAE faces here—one would hate to have students spin off on tangents and end up not having their critical-thinking skills tested at all; albeit a crucial aspect of critical-thinking skills is knowing *when* to apply them and not just how.

Rather than providing another detailed analysis of CAE's exemplary "high quality" student response, I will just say that this time the response was quite good at pointing out that we need information about comparative *proportions* (and not just *numbers*) in order to have any evidence of a correlation of divorces to June weddings, and that that correlation would not be identical to a causal relation. But the student repeatedly phrased his or her objections by merely saying that the claims in the prompt *might* be wrong, and that there *might* be alternative explanations for this possible correlation and for why couples postpone engagements. Again, merely pointing out that an empirical claim *might* be wrong is platitudinous; and to offer a *merely possible* alternative explanation is likewise. One needs to offer a *plausible* alternative explanation in order to raise a legitimate criticism; because, after all, the conclusion of any inductive argument by definition might be false, even given the truth of its premises.

During a recent conference call with Jeffrey Steedle and Marc Chun, from the CAE, the criticisms I have just presented were characterized as "ad hoc." I was puzzled by this charge, because this would mean that my criticisms were without independent evidence. But the independent textual evidence I was using was the student responses provided by the CAE in its promotional materials. If anything was an ad hoc rescue, it was *their* charge of my being ad hoc. Perhaps what they meant was that I was making a hasty generalization from a small sample. But I was not using a dangerously small *random* sample; I was using *their* examples that were offered as being *representative* of "high quality" student responses. I can't help but conclude that someone doesn't know what 'ad hoc' means. And this quite naturally brings me to my conclusion.

## 3. The fatal flaw

After this examination of the CLA test, I think I have discovered a serious, perhaps fatal, weakness among its many strengths. Its *goals* are to be commended—measuring students' higher-order skills of critical thinking, analytic reasoning, problem solving, and written communication—these *are essential* to higher learning. And the *rubrics* (Benjamin et al., 2009, pp. 41-3) and the *CLA Scoring Criteria* (CAE, 2011a) used to assess the students' application and enhancement of these higher-order skills are spot on. So, the graders *seem to be looking* for the right things in the students' responses. But, according to my findings, they just aren't correctly finding them very well. Remember, the critical-thinking criteria the graders are to use in scoring the students' responses in Make-an-Argument are: "Clarifying a position and supporting it with evidence, considering alternative viewpoints or counter points to their argument, developing logical, persuasive arguments, [and exhibiting] depth and complexity of thinking about the issues raised in the prompt" (Benjamin et al., 2009, pp. 53-4). While obviously *believing* that these criteria are met, the graders are *in fact* falling for numerous informal fallacies, platitudes, and evasions. They are being persuaded by arguments and criticisms that are simply *not cogent*. How can this be happening?!

I think it's because *the graders cannot see the trees for the forest*. They are trained to take only a *holistic* view of critical thinking, ignoring the *component* skills of critical thinking that are often the focus of the multiple-choice assessment tests so disparaged by the CAE. I noted earlier how skeptical the CAE is that component generic critical-thinking skills can be clarified enough for study, instruction, and testing. But many of us in philosophy departments all over the world do it everyday, as we teach courses in critical thinking and informal logic. We teach students how to identify and dissect arguments, taxonomize arguments as inductive or deductive so as to apply the appropriate cogency conditions for their assessment, and identify and avoid popular formal and informal fallacies that result from not meeting those cogency conditions (Possin, 2002a). *Then* we *also* instruct students on how to synthesize and apply all those component critical-thinking skills in the holistic tasks of discovering and arguing for the most rational position on an issue while critically reviewing competing positions and their arguments (Possin, 2002b). *One cannot successfully do the holistic latter without learning the component former*; just as one cannot build a brick stairway without using component bricks.

Who, then, is authoring the answer keys and scoring the students' responses to the CLA's Tasks? I asked Jeffery Steedle, Measurement Scientist, at the CAE. Here is our email correspondence:

> [KP] I was wondering about the educational backgrounds and areas of specialization of the authors of the Performance Tasks and the authors of their paradigmatic answers that are used for comparatively scoring what counts as meeting the criteria listed on the scoring rubrics. I know the graders come from a diversity of areas and they are trained to ensure better reliability. But I'm particularly interested in the areas of expertise and education of those drafting the answers that the graders are trained to be looking for in the students' test responses.

> [JS] We don't typically write "paradigmatic" answers. When we train scorers, we provide actual student responses as examples at each scale point (1 through 6 on multiple scales). We also provide a document that we call "response features," which catalogues common (valid) ideas that students may discuss in their responses. This is initially created by the person who developed the task, but it is commonly updated in light of what we see in student responses. It is not an exhaustive list, and scorers are given the leeway to award credit for other valid points that students make in their arguments.

> The developers are people who are trained to develop tasks according to our task specifications. For the most part, they have been a mix of measurement professionals affiliated with CAE and experienced CLA scorers. The majority of scorers have backgrounds in the liberal arts (predominantly English literature and composition) or education. They're roughly split between having master's degrees and PhDs. One requirement is prior experience evaluating student writing at the college level.

So the task and response authors, as well as the scorers, come from a diversity of disciplines, such as measurement, English, and education—*not applied logic*. Since the CAE prides itself on assessing what 99% of professors believe to be essential critical-thinking skills, it was only natural that it had the RAND Corporation do a reliability and procedural validity study of the CLA using a very diverse panel of 41 faculty, from the social sciences (9), English (8), the physical sciences (7), philosophy (4), math (4), history (2), the arts (2), business law (1), and other disciplines (2) (Hardison & Vilamovska, 2009, p. 20). But, as Richard Paul (1995) discovered, university professors

"have little understanding of critical thinking nor how to teach for it, but also wrongly and confidently think they do." Here is a small fraction of Paul's findings:

> Though the overwhelming majority [of the faculty] (89%) claimed critical thinking to be a primary objective of their instruction, only a small minority (19%) could give a clear explanation of what critical thinking is. Furthermore, according to their answers, only 9% of the respondents were clearly teaching for critical thinking on a typical day in class…. When asked how they conceptualized truth, a surprising 41% of those who responded to the question said that knowledge, truth, and sound judgment are fundamentally a matter of personal preference or subjective taste…. [O]nly a very small minority could clearly explain the meanings of basic terms in critical thinking. For example, only 8% could clearly differentiate between an assumption and an inference, and only 4% could differentiate between an inference and an implication.

My personal experience confirms Paul's findings; for example, in my decades of university committee work, I have yet to come away from a single meeting without jotting down at least one new fallacy committed by the attending professors, to share with my Critical Thinking class and add to the bulging set of exercises in my *Critical Thinking Software* (Possin, 2002a).

So, doing a validity study to see if one's staff of scorers is accurately measuring critical-thinking skills by correlating its results with the judgments of a diverse set of professors is, to paraphrase Wittgenstein (1953, §265), like going out and buying *several* copies of the tabloids to assure oneself that their story about the UFO landing is true.

Why does the CAE treat the meager requirements of having a graduate degree and "prior experience evaluating student writing at the college level" as jointly sufficient for qualifying as a CLA scorer? Because they believe that the expertise involved in acquiring, applying, and assessing general critical-thinking skills simply arises from learning any single "highly situated and context-bound" discipline, and that this view is "supported by research." According to the CAE, "through practice with a particular subject area, learned knowledge becomes sufficiently generalized to enable it to transfer to the realm of enhanced reasoning, problem-solving, and decision-making skill that can be demonstrated across content domains" (CAE, 2011b). I only *wish* the acquisition of such generic and transferable critical-thinking skills were that easy!

What "research" led the CAE to believe that this is how critical-thinking skills are so mysteriously acquired? They make reference to (Shavelson & Huang, 2003) and (Klein et al., 2005), which contain *no* empirical research on critical-thinking skills and their acquisition, but which do refer the reader to (Brandford et al., 2000). However, there we are told that (my emphases):

- Knowledge that is overly contextualized can *reduce* transfer….(p. 53)
- One way to deal with the lack of flexibility [of context-bound learning] is to ask the learner to solve a specific case and provide them with an additional similar case; the goal is to *help* them abstract general principles….(p. 62)
- Transfer is also enhanced by *instruction that helps* students represent problems at higher levels of abstraction. (p. 63)
- Transfer can be improved by *helping* students become more aware of themselves as learners who actively monitor their learning strategies and resources [i.e, by metacognition]…. (p. 67)

The research, then, indicates that *students need a great deal of help* practicing component critical-thinking skills across *many* contexts in order for those skills to become generic enough to be transferable and applicable. This is exactly the kind of instruction and practice students receive in a dedicated critical thinking or informal logic course, in which component critical-thinking skills are studied in multiple contexts and then ultimately applied holistically to multiple topics. Critical-thinking skills are *not* statistically significantly enhanced by content-*specific* courses, e.g., introduction to philosophy, or by content-*independent* courses, e.g., symbolic logic (Possin, 2008). Marcus Gillespie (2012) recently demonstrated this at Sam Houston State University, using the *Critical Thinking Test* [CAT]: The general education course, *Foundations of Science*, a critical-thinking course dedicated to the *explicit* study of inductive and scientific reasoning in the context of *various* subject-matter case studies, enhanced the critical-thinking skills of the students more than students on average achieve otherwise after 4 years of university coursework. Other content-*specific* science courses, such as introductory chemistry, biology, and physics courses, which Gillespie used as control groups, demonstrated no statistically significant gains. So much, then, for leaving the task of magically enhancing critical-thinking skills to "immersion" and "critical thinking across the curriculum."

Let me note one exception that is *not* really an exception: Only by adding a *separate, dedicated, generic* critical-thinking curriculum to his general psychology course, was Tom Solon (2006) able to demonstrate impressive gains in his students' critical-thinking skills as measured by the *Cornell Critical Thinking Test Level Z* (Ennis & Millman, 1985).

My point is that to enhance students' critical-thinking skills, they should be *deliberately* and *explicitly* studying critical thinking with the assistance of those with real expertise in those skills. And just possessing a graduate degree is very poor evidence of having acquired that expertise in critical-thinking skills. Hence, what the CAE needs to do is to make sure its response authors and graders *truly are experts* in *both* the wide array of *component* critical-thinking skills and their compilation and holistic application to the projects of making rational decisions, solving problems, and writing position papers and critical reviews using *cogent* arguments and criticisms instead of fallacious ones. Personnel at the CAE may be excellent "measurement scientists," ensuring the *reliability* of the CLA; but they appear to be missing the mark on its *validity*—measuring *rhetorical* skills instead of actual critical-thinking skills.

One last issue that I want at least to mention here is that all of the CLA Tasks are currently computer graded; graders continue to confirm the computer-assigned scores but on only 10% of the student responses. The CAE tries to reassure us that "CLA computer-assisted scoring is as—and, in some cases, more than—accurate as two human scorers [*sic*]," with the correlation of scores being .80-.88 between graders and being .84–.93 between computer-scoring and grader (Elliot, 2011, pp. 3-4). This is some evidence that the CAE's scoring system is *reliable*; but consistency is a fickle virtue when one is consistently *wrong*. If the accuracy of the CLA graders is in doubt, and the computer-assisted grading system is strongly correlated with the graders' scoring, then the accuracy of the computer-assigned grading is in doubt too. And still having the scorers rechecking 10% of the student responses just brings us back to that passage from Wittgenstein again. [For a more detailed critical review of the CLA's computer-assisted scoring, please see (Ennis, 2012).]

## 4. Conclusion

I have identified a serious, if not fatal, ailment on the part of the CLA. I have argued for my diagnosis and have offered a prescription, recommending a shift in the way the CLA is scored, so that student responses are judged more on the basis of com-

ponent critical-thinking skills and less on the basis of rhetorical skills.

I am not optimistic that the patient will heed my advice, however, because doing so would come at the high price of rendering the CAE's large database of past scores obsolete. But I think it is worth the price—the CLA is a commendable assessment tool; it just needs to be used correctly.

## References

Arum, R. & Roksa, J. 2011. *Academically Adrift*. Chicago, IL: University of Chicago Press.

Benjamin, R., Chun, M., Hardison, C., Hong, E., Jackson, C., Kugelmass, H., Nemeth, A., & Shavelson, R. 2009. Returning to learning in an age of assessment. Retrieved January 2012, from http://www.collegiatelearningassessment.org

Bransford, J., Brown, A., & Cocking, R. 2000. *How People Learn: Brain, Mind, Experience, and School.* Washington, DC: National Academy Press.

Center for Assessment & Improvement of Learning. *Critical Thinking Assessment Test* [CAT].
http://www.tntech.edu/cat/home/

Council for Aid to Education [CAE]. 2011a. CLA scoring criteria. Retrieved January 2012, from
http://www.collegiatelearningassessment.org

Council for Aid to Education [CAE]. 2011b. CLA framework/construct definition. Email attachment from Jeffery Steedle, CAE.

Elliot, S. 2011. Computer-assisted scoring of performance tasks for the CLA and CWRA. Retrieved January 2012, from http://www.collegiatelearningassessment.org/files/Computer AssistedScoringofCLA.pdf

Ennis, R. 2012. Grading the critical thinking aspects of the CLA Test. Central Division Meeting of The Association for Informal Logic and Critical Thinking.

Ennis, R. & Millman, J. 1985. *The Cornell Critical Thinking Test Level Z*. Pacific Grove, CA: Midwest Publications.

Frankfurt, H. 2005. *On Bullshit*. Princeton, NJ: Princeton University Press.

Gillespie, M. 2012. The *Critical Thinking Assessment Test* [CAT]: Assessing CT in science courses. Central Division

Meeting of The Association for Informal Logic and Critical Thinking.

Hardison, C. M. & Vilamovska, A. 2009. The *Collegiate Learning Assessment*: Setting standards for performance at a college or university. Retrieved January 2012, from http://www.rand.org/pubs/technical_reports/TR663.html

Higher Education Research Institute [HERI]. 2009. *The American College Teacher: National Norms for 2007-2008* . Los Angeles, CA: HERI, University of California Los Angeles.

Klein, S., Chun, M., Hamilton, G., Kuh, G., & Shavelson, R. 2005. An approach to measuring cognitive outcomes across higher education. *Research in Higher Education,* 46 (3), 251-75.

Paul, R. 1995. Study of 38 public universities and 28 private universities to determine faculty emphasis on critical thinking in instruction: Executive summary. Retrieved January 2012, from
http://www.criticalthinking.org/pages/center-for-critical-thinking/401

Possin, K. 2002a. *Critical Thinking: A Computer-Assisted Introduction to Logic and Critical Thinking*. Winona, MN: The Critical Thinking Lab.

Possin, K. 2002b. *Self-Defense: A Student Guide to Writing Position Papers*. Winona, MN: The Critical Thinking Lab.

Possin, K. 2008. A field guide to critical-thinking assessment. *Teaching Philosophy*, 31 (3), 201-28.

Shavelson, R. & Huang, L. 2003. Responding responsibly to the frenzy to assess learning in higher education. *Change* 35 (1), 10-19.

Solon, T. 2006. Generic critical thinking infusion and course content learning in introductory psychology. Central Division Meeting of Association for Informal Logic and Critical Thinking.

Spellings Commission on the Future of Higher Education. 2006. *A Test of Leadership: Charting the Future of the US Higher Education*. Retrieved January 2012, from
http://www2.ed.gov/about/bdscomm/list/hiedfuture/reports/pre-pub-report.pdf

Wittgenstein, L. 1953. *Philosophical Investigations*. New York, NY: Macmillan Publishing.