I sincerely thank Drs. Anastasi, Fasko, Hatcher, and Possin for their careful reviews of the Halpern Critical Thinking Assessment (HCTA). I appreciate their kind words and constructive criticisms. Not surprisingly, they raised questions that are central to understanding critical thinking and its instruction and measurement. I grouped my response to their comments around common themes and questions, which address many of the key issues in the assessment of critical thinking.

I. The Halpern Critical Thinking Assessment (HCTA):
Some Background Information
The HCTA provides a measure of how people think when they contemplate information that relates to real-world experiences. All questions are embedded in realistic scenarios in order to test how well respondents recognize the need for and apply appropriate critical thinking skills. The HCTA uses two response formats: constructed responses (short answer responses that assess recall)—how people respond to a situation in their own words—and recognition responses (multiple choice, ranking, or selecting good alternatives from a list), which is measures how well respondents can recognize a correct or optimal alternative among a list of possible answers. The HCTA is the only standardized measure of critical thinking designed for adults that uses both recognition and recall response formats with realistic scenarios. Scores are positively correlated with many measures that presumably reflect varying levels of critical thinking including grade point averages, level of education, and scores on standardized measures of achievement (multiple studies are listed in the HCTA manual). In addition, the HCTA is correlated with scores on a test of real world outcomes—what people say they do in real life. It has been subject to stringent psychometric analyses in many countries and in many languages (e.g., Ireland, Spain, China, Belgium). The HCTA was developed over several decades using data from thousands of test-takers and with input from a wide range of professionals.

II. Why Assess Critical Thinking?
The Power Balance Wristband improves energy, flexibility, and balance. Elite athletes including David Beckham and Shaquille O’Neal have endorsed it. O’Neal believes that it gives him a competitive edge on the court. “It’s no gimmick,” according to O’Neal (para. 1, Associated Press, 2011). Over 3.5 million bracelets were sold with sales in the United States alone over $35 million dollars. According to the company that makes these bracelets they work via an embedded hologram that uses “key frequencies of the human energy system” (para. 2, Pettit, 2011). But, when investigated by the Australian Consumer and Competition Commission, the company was forced to admit that “there is no credible scientific evidence that supports our claims” (para. 2, Caparell, 2011).

The list goes on—sneakers that tone your body (not because you wear them to exercise; the “secret” to toning success is in their design), bogus medical claims, financial scams, racist websites disguised as unbiased sources of information, and much more. These examples of the complexity of everyday life raise an important question: What are the critical thinking skills needed for informed citizenship at this time in history with information changing at an increasingly rapid rate? Employers responded to the question of what they want their employees to know and be able to do with a resounding unanimity: Their top choice is employees who can communicate effectively both orally and...
thinking that may sound correct, but is actually wrong. It is a kind of dishonesty. He raises the important question of whether “genuine” critical thinking involves being appropriately motivated. Ethics are an important part of every discussion about the use of critical-thinking skills, often revolving around a standard question about whether it is ever justified to deliberately misuse techniques such as either-or dichotomies or put downs to achieve a just goal. There are important ethical considerations, but they do not define critical thinking. Most definitions involve the effortful and consciously controlled processing that uses available information and sound cognitive strategies, and purposefully strives to overcome individual biases (e.g., Ennis, 1993; Moseley et al., 2005; Sternberg, Roediger, & Halpern, 1997).

I note here that the basic components of this definition of critical thinking are not specific to any discipline—this is not a “psychology definition.” For example, van Gelder (2013), a leading authority in critical thinking and a professor of philosophy offered this “pragmatic” definition: “First, the essence of critical thinking is correct or accurate judgement. That is, to think critically is to think in ways that are conducive to being ‘more right more often’ when making judgements.” This is essentially the same definition as the one that I have been using.

III. What is Critical Thinking?
Virtually everyone will agree that we want to enhance critical thinking and identify the extent to which individuals can think critically. But, what exactly is critical thinking? Here is a definition that I have used for decades:

*Critical thinking is the use of those cognitive skills or strategies that increase the probability of a desirable outcome. It is used to describe thinking that is purposeful, reasoned, and goal directed—the kind of thinking involved in solving problems, formulating inferences, calculating likelihoods, and making decisions, when the thinker is using skills that are thoughtful and effective for the particular context and type of thinking task.*

Similar definitions have been used by multiple authors in the field of critical thinking, yet Possin criticizes this definition by asking what kinds of thinking skills, “Hopefully not those of sophistry” (Possin, 2013, p. 5). Sophistry is thinking that may sound correct, but is actually wrong. It is a kind of dishonesty.

IV. What are Critical Thinking Skills?
I have published many lists of skills that are grouped into various categories under the heading of critical thinking. (See Halpern, 2014, for lists with examples of each skill.) Five broad categories of critical-thinking skills that are tested with the HCTA are (1) verbal reasoning, (2) argument analysis, (3) hypothesis testing, (4) likelihood and uncertainty, and (5) decision making and problem solving. All assessments sample items from a larger pool of possible topics, and the HCTA is no exception. I thank Hatcher for
INQUIRY: CRITICAL THINKING ACROSS THE DISCIPLINES

compiling this list drawn from questions on the HCTA: “The twenty-five scenarios cover such knowledge and skills as: (a) argument recognition and evaluation, (b) problems with confusing correlation and causation, (c) the putative consequences of various practices, (d) problems with vague definitions, strategies for taking tests, (e) problems of inference from a small sample, (f) identifying unwarranted assumptions and hasty generalizations, (g) identifying instances of regression to the mean, (h) and questions requiring some understanding of probability” (Hatcher, 2013, p. 19).

V. Can critical thinking skills be taught (and learned) so that they transfer across domains of knowledge?

The answer to the question of whether one can improve in their ability to think critically is a simple, “Why not?” There is nothing in our knowledge of how people think and learn that would contradict the idea that we can become better thinkers if we take advantage of appropriate learning opportunities, which can occur in the home, on the job, at school, and most other places. University faculty teach mathematics, oral communication, and writing in the belief that these skills will transfer across content areas. Critical-thinking skills are similar to these other basic skills—they are needed in a wide variety of contexts and are generally considered “domain general,” which means that the skills learned in one academic domain can be applied in other academic domains (Angeli & Valanides, 2009).

Consider, for example, this question about the conclusions that can be made when people view a televised debate about a controversial topic and then log onto a website to indicate if they were “for” or “opposed to” the topic being debated. Suppose that among those residents of a particular state who voted in first hour following the broadcast, close to half voted for and half voted against the topic. Based on these data, the commentator for the show concluded that the people in his state were evenly divided on this topic and that both sides were equally effective in persuading listeners. Do you agree with the commentator’s conclusion?

This question is not about the controversial topic or this particular scenario, but what meaning can be drawn from a sample that may not be representative of the population of all people in this state. Test-takers are asked to provide two suggestions for finding a better way to gauge the way people think about the controversy and the effect of the debate on attitudes towards the controversial issue. Among many possible responses, test-takers could suggest that (a) the polls measure attitudes toward the controversial issue before and after the debate, and (b) the people who are studying this issue should get a sample that is more representative of the people in this state than those who will vote on their computer as soon as the show is over.

These scenarios are similar to “situational judgment tests” which are commonly used in selection (hiring), usually with written or video stimuli and multiple choice alternatives. There is a massive literature on their use. Jansen et al. (2013) highlight their use in valid personnel selection procedures. They explain that “Assessing the demands of situations refers to cognitive processes for deciphering (or reading) what is required to behave effectively in situations” (p. 327). Thus, one way to think of the HCTA is that it is an advanced variation on situational judgment assessments that have been shown to be effective in hiring decisions and that it is an advanced variation because the HCTA uses both constructed and multiple choice responses.

The underlying rationale for teaching courses in critical thinking is that critical thinking skills can be taught and learned, and when students recognize the need for these skills in appropriate contexts and apply them, they will be better thinkers. Thus, as Anastasi
noted, “It seems that many of the skills being measured by the HCTA can be easily improved by instruction. While this is certainly a primary reason for the importance of critical thinking assessment and a primary factor for why critical thinking assessment is so important to educators, this can lead to one potential concern for developing a valid assessment of critical thinking” (Anastasi, 2013, p 14). It is likely that most readers never had a course in critical thinking, yet have learned many of the skills. We expect that explicit instruction in critical thinking will lead to better thinking in a variety of settings; a large research literature supports this expectation (e.g., Helsdingen, van den Bosch, van Gog, & van Merrienboer, 2010; Marin & Halpern, 2011; Moseley, Baumfield, Elliott, Higgins, Miller, & Newton, 2005). The HCTA is not a test of knowledge of a content area; instead it is a test that requires the use of critical thinking skills.

V. Psychometric Properties of the HCTA
“Tests should be supported by evidence of reliability and validity for their intended purpose” (International Test Commission, 2013, p. 13). Although an overview of the constructs reliability and validity is presented in this section, an extensive and more technical discussion about various types of reliabilities and validity is beyond the scope of this paper. Interested readers can find more about these topics in many excellent books on test development.

A. What makes a critical thinking test valid?
The Standards for Educational and Psychological Testing define validity as “the degree to which evidence and theory support the interpretation of test scores entailed by proposed uses.” Validity is a matter of degree—it is not an all-or-none property of an assessment. The extent to which an assessment is valid depends on its intended use. The HCTA was written to assess five categories of critical thinking skills in adults. Although Possin states that “I agree with Halpern that these are indeed crucial categories of CT skills,” (Possin 2013 p. 6) he later questioned the validity of the HCTA because it would not be a good test of what was learned in a course he teaches on deductive reasoning. Possin is correct in his assessment—it is not a valid measure of what was learned in a course in deductive reasoning, nor is it intended to be. I corrected the typo he pointed out about increasing/decreasing running time and made the change he suggested about making inferences from small samples. I am working with the test publishers to make the HCTA as error-free, clear, and easy-to-use as possible. The latest version of the HCTA is shorter (20 questions instead of 25), and we achieved the reduction while also enhancing its psychometric properties. I hope to continue to learn from constructive feedback.

Hatcher would also like more questions on the HCTA that require deductive reasoning. It is interesting to note that Hatcher and Possin are professors of philosophy, and it is likely that philosophers weigh deductive reasoning more heavily as a component of critical thinking than psychologists or professors in other departments (e.g., education, communication, history, business, or economics). Hatcher recognized this likelihood when he said that: “it may be that people in psychology do not conceive of CT as so closely tied to deductive reasoning or evaluating arguments as those of us in philosophy”(Hatcher, 2013, p. 21). Deductive reasoning is one component of critical thinking. In the world outside of college classes, the ability to understand and make arguments, use likelihood and uncertainty, reason inductively, and recognize the way words influence our understanding are important components in thinking well. The HCTA is not a test of deductive reasoning, and if that is what a consumer is looking for, there are many such tests on the market.

The central question regarding validity is what makes a measure of critical thinking good? A good measure can predict
how people act in the real world. The HCTA is the only measure of critical thinking that has demonstrated real world validity—it can predict what adults do (more precisely, what they say they do) in their daily lives. There are very few assessments that actually have checked for relationships with real world use of critical thinking. Butler (2012) did exactly that and found that the HCTA correlates with people’s reports of the decisions and actions they have taken. Of course, Anastasi is correct in saying that there may be some biases in what people say (even anonymously on line) about their life choices. But that objection makes the results even stronger because it would make the likelihood of getting statistically meaningful relationships with the HCTA even less likely. Furthermore, the HCTA showed strong relationships with two factors extracted from the Assessment of Real World Outcomes: preference for “deeper” level processing and “effort preference.” This is a very strong test of the validity of the HCTA; one that is very rare in the assessment literature. Thus, the HCTA passes the most stringent tests of validity.

Possin commented that some test items were ambiguous, but the other reviewers did not agree. Hatcher, for example, commented that “For the most part, the instructions are clear, and most of the scenarios portray situations that should be familiar to most students. Again, one thing I especially liked about the HCTA design was the long series of questions about a scenario, rather than the standard three or four choices present in the multiple choice questions from the CT tests that I have used” (Hatcher, 2013, p. 19).

B. Correlations with academic-related measures such as standardized examinations and grades.

Like most critical thinking assessments, the HCTA is validated against academic achievement measures. In all comparisons, across multiple studies in multiple countries with varied samples, correlations were significant and in the predicated direction. (Data and references are provided in the HCTA Test Manual.) Higher scores on standardized examinations, better grades, more selective admissions criteria, more advanced levels of education, and other measures all show positive relationships with scores on the HCTA. But, the correlations are, in general, not large. Anastasi ponders this relationship, and also asks if the HCTA predicts success in school for example, better than other measures (Anastasi, 2013, p. 16). An unstated assumption in this question is that success in school relies more heavily on the skills of critical thinking than other criteria such as memory for facts, interpersonal skills, and regular attendance. My response is that it should be a better measure of success in school than the standard admissions tests if instruction really reflects critical thinking as opposed to rote recall of information. In a recent study of a critical thinking intervention with 12th grade students in Hong Kong, Ku, Ho, Hau, and Lai (2013) found that when the intervention explicitly taught critical thinking skills, scores were improved in the constructed response portion of the HCTA, “which points to the interlocking relationship between assessment and instructional needs.” The skills that were explicitly taught transferred across contexts, showing that critical thinking skills can be learned, but standard assessment such as multiple choice questions may not reveal the learning.

The HCTA also correlates with measures of need for cognition, openness, and conscientiousness (e.g. Halpern, 2007; Hau et al., 2006; Ku et al., 2006; Ku et al., 2013; Marin & Halpern, 2011). These correlations are considerably lower than more construct-related measures (such as academic grades and standardized examination scores). In addition, there is good psychometric evidence for a general critical thinking factor and the separability of its recognition and free recall facets (Halpern, 2007; Hau et al., 2006; Ku et al., 2006). Verburgh (2013) reported that “the
use of constructed response questions was one of the main strengths of the HCTA when it comes to content validity” (p. 149). Fasko found the HCTA to be “reliable and valid instrument to evaluate critical thinking skills of adults.”

C. Reliability

Even though most experts agree that constructed responses are the best measure of what people actually think and do, most of the available critical thinking assessment use only multiple choice formats because of the time needed to grade constructed responses. In addition, another common problem with constructed responses is that it can be difficult to get good interrater reliabilities. The HCTA uses a computerized grading system that guides the grader with prompts that make it easy and relatively fast for anyone to grade constructed responses in a consistent manner. Inter-rater reliabilities are very high because the computerized grading system prompts the grader with questions about the constructed responses and then automatically computes a numerical grade for each question. In grading constructed responses, the grader decides the extent to which the response matches the content in the prompt. For example, a prompt might ask if the test-taker indicated that a larger sample size was needed. Each constructed response has multiple prompts.

Hatcher raised concern about whether the order in which topics were mentioned by the test-taker would affect the scores. (Hatcher, 2013, p. 21) The order in which topics are mentioned does NOT affect scoring, so this concern can be eliminated. Possin is concerned that there are other potentially correct answers that may not be included in the prompt. (Possin, 2013, p. 11) This is a legitimate concern. In response, I note that the prompts were extracted from the responses of well over 1000 test-takers. In addition, the grading of constructed responses always requires human judgment, so there is some latitude as to what to accept in a constructed response. As readers probably know, some standardized examinations now use computer-graded essays, which are designed to remove human judgment from the grading. The prompt system, as far as I know, is unique to the HCTA. It maintains human judgment in grading constructed responses while providing a framework that leads to reliable scoring.

The use of grading prompts is a unique way of enhancing interrater reliabilities for constructed responses. In her doctoral dissertation, Verburgh (2013) examined the psychometric properties of the HCTA using a Dutch language translation and independent raters. She reported interrater reliabilities that are considerably higher than those reported in the current edition of the test manual (a new manual is in preparation). Verburgh found interrater reliabilities ranging from .84 to .93. These are high values that attest to the value of scoring prompts. Anastasi agrees, “In summary, educators and employers are searching for instruments that adequately assess an individual’s ability to think critically. The HCTA appears to provide them with a useful option that can provide a valid, reliable measurement of such skills” (Anastasi, 2013, p. 17).

VI. The Logic of Alternative Forms

The enterprise of teaching critical thinking—or any other domain general skill—is based on the assumption that knowledge acquired in one situation or setting will transfer to another situation or setting. The underlying idea when teaching critical thinking is to teach students a set of domain general skills that can generalize across various disciplines. The specific skills that are taught constitute the backbone of scientific/rationale reasoning. For example, several authors convincingly argued that argumentation is at the core of both scientific and critical thinking (e.g., van Gelder, Bissett, & Cumming 2004). Other critical-thinking skills such as understanding that the finding that two constructs are correlated does not necessarily imply that one
causes the other are also relevant to scientific domains and play a central role in educated rationale decision making in daily live.

In assessing critical thinking skills our task is to tap into domain-general skills. HCTA covers a broader range of skills at the heart of the critical-thinking construct. In teaching critical thinking we are generally interested in far-transfer. The main task we are facing in measuring critical thinking or any other construct is that test scores in general do not reflect only the trait in which we are interested. More precisely, test scores can be contaminated by various sources of construct-irrelevant variance (for an overview see Arendasy and Sommer, 2013), who outlined several possible sources of construct-irrelevant variance, including settings in which the trait of interest is measured at multiple time-points (e.g. pre-test before taking the course and post-test after completing the course). For example, item memorization and memorization of test design principles can lead to an increase in test scores that does not correspond to a true increase in the latent trait measured, which in this case is critical thinking. The psychometric literature and meta-analyses indicate that “retest effects” are omnipresent in cognitive ability and skill assessment. This line of research indicated that retest score gains that do not correspond to a true increase in the latent skill measured are most pronounced when using identical retest forms (administering the same item set repeatedly). In a recent study Arendasy and Sommer (2013) demonstrated that the detrimental effect of retesting can be reduced by using alternate test forms constructed by means of item cloning. In general, item clones share the same problem structure but differ in terms of surface features assumed to have no or little effect on the item parameters (for an overview see Arendasy & Sommer, 2012). The alternative forms of the HCTA were built using the principles of item cloning.

In the last decades there has been a move to use automatic computer-generated alternative items to construct valid and reliable tests. This approach has become increasingly popular because it also helps to resolve the problem posed by compromised items. In testing, items are “exposed.” Test takers tell others about specific items, and if there is an attempt to track progress over time, test takers will recognize items they encountered earlier if they are given the same items repeatedly. Imagine if the SATs or the licensing boards for any profession or any other serious assessment used the same questions every year. Thus, the use of a “pool of items” became a standard procedure in educational and occupational testing to circumvent possible detrimental effects of construct-irrelevant sources of variance that may compromise test score validity in applied practical settings.

Since one use of the HCTA is to evaluate the effect of critical-thinking instruction in educational settings, we are faced with the same problems as classic high-stakes assessments. For all these reasons, we chose to use alternate forms to overcome, or at least reduce, unwanted effects on test scores. All items and scales are subject to multiple stringent statistical controls to examine the extent to which scores obtained on alternate test forms can be directly and unambiguously compared to each other. This is done by means of both item response theory and structural equation modeling based methods of evaluating measurement invariance across alternate test forms. The main benefit of this approach is that if measurement invariance holds, individual differences within and between test forms can be unambiguously attributed to the same latent skill measured (for an overview see Mislevy et al., 2013). This finding implies that reliable score differences between test forms reflect true differences in the latent skill measured and cannot be attributed to differences in the characteristics of the individual alternate test forms. Given the theoretical arguments and their empirical support, I was surprised by Possin’s criticism of the practice of alternative
forms as it is a central tenet of standardized testing to circumvent the detrimental effects that call the validity of test scores and test score comparisons across time-points of measurement into question. (Possin, 2013, p. 4) In fact, items for the HCTA are continuously being developed and tested to ensure that the HCTA and its subscales can be used in applied settings in the form of multiple, psychometrically and conceptually equivalent test forms to rule out the possibility that mean score gains are attributable to any factors other than a true gain in critical thinking skills. This is an ongoing task. I thank Possin for pointing out a typographical error in one of the questions (which I had been assured was corrected previously) and I thank him for his suggestion about making a slight change in one item so that the item became more isomorphic to its matching item used in the other test form.

VII. Weighting of dimensions of critical thinking

Anastasi supports the idea of multiple dimensions of critical thinking and the differential weighting of the dimensions, and asks about the theoretical reasons that underlie the dimensional weighting. (Anastasi, 2013, p. 15) The criteria for weighting are explained in the HCTA Test Manual. Verbal reasoning, for example, was given a lower weighting than the other dimension because connotations of words are often difficult to convey in multiple language translations, and they can be a particular problem when test-takers are not taking the HCTA in their best language, which is increasingly common given the rise in international students and immigration patterns world-wide. It is not possible to use only words that have unambiguous meanings across multiple languages as Anastasi suggests. All of the components of critical thinking are not equally representative of the construct of critical thinking. The HCTA was developed over several decades and reflects the judgment of multiple professors and other stake-holders in the development of critical-thinking skills.

Anastasi also asks about the decision to weight multiple choice and constructed questions equally. (Anastasi, 2013, p. 15) There are many ways to design an assessment. In the HCTA, every scenario is assessed with both multiple choice and constructed response questions. The correspondence (or its lack) between these two types of responses for exactly the same question is known as “method variance.” The cognitive processes involved in recognizing a correct response and in generating a correct response are not the same. Often someone can recognize a correct response when it is presented, but cannot recall the correct response when there are no alternatives to select among. Determining the correct response in recognition tests is objective and easy: even with exceptionally high interrater reliabilities for the constructed response questions, there is always human judgment involved on the part of the grader. Thus, each type of response has positive and negative aspects. The decision to weigh these two types of response equally was based on an understanding of the strengths and weaknesses of these response formats. Other tests of critical thinking are based only on recognition responses or recall responses, not both, and thus miss a critical component of critical thinking. Anastasi raised a question about the use of recognition-only items as an assessment option. (Anastasi, 2013, p. 15) He correctly pointed out that such a measure misses the additional component of constructed response. Some assessment administrators prefer to use only recognition response formats because there is no time spent in grading (recognition responses are graded by computer) and thus the cost is reduced. Thus, it is an option for those who want to use it.

VIII. Nuts and Bolts: Ease of Use and Pragmatic Concerns

Fasko used the CD version of the HCTA and described its installation on his computer as easy (Fasko, 2013, p. 26). We plan to use Fasko’s comments about the instructions
that accompany the CD version of the HCTA as we write instructions for a new test launch planned for early in 2014 that will use an interface that user-friendly. As a test author, I have no control over the cost. I note here that the cost of the HCTA is considerably less than that of other standardized assessments. Consider, for example, the ETS Major Field Assessments, which consist of multiple choice questions designed to test knowledge of many typical college majors such as psychology or chemistry. These assessments cost $24 to $25 (depending on volume), and these are multiple-choice only tests. Test users (i.e., consumers) do not have the option of viewing or grading the test questions, which is an option with the HCTA.

The test publisher for the HCTA, Schuhfried, Inc. compiled a comparison cost chart for other standardized tests of critical thinking. In a personal correspondence, Mayr (who heads international sales) reported that the Watson-Glaser, a competitive test, costs between $22 and $69 for use on-line. The test manual has to be purchased separately. By comparison, the HCTA costs between $22 and $24 for the most expensive option in which all constructed response questions are graded. The manual is included in this price. If users prefer to grade the constructed response questions themselves, they can use the grading prompts and costs are approximately $12. There are also options to use the HCTA on individual computers. The cost for this option is as low as $4 for high volume users. Thus, there are many ways to administer the HCTA, with the on-line version and all constructed responses scored being the most expensive and the desktop version being very inexpensive.

Possin also questions that accessibility of the HCTA because the on-line version works on Internet Explorer, and he wants all Mac users to be able to access the HCTA. (Possin, 2013, p. 6) There are easy fixes for this problem. There are many Windows emulators that work on the MAC or students can take the HCTA on-line at a computer lab that has Windows based machines. Approximately 90% of the personal computing market uses windows, so it is unlikely to be a real problem. Possin also noted that when taking the HCTA he could not go back and review his responses to earlier scenarios. (Possin, 2013, p. 6) This is a deliberately programmed feature of the HCTA. Test-takers first respond to a scenario with a constructed response. They then progress to a second screen where the scenario is repeated, but this time the response format is forced choice. Given that recognizing a correct response is often easier than generating one, we did not want test-takers to be able to edit a constructed response based on the alternatives they are given for the forced choice responses.

IX. Looking Forward: Additional Research

The HCTA allows new types of research on the teaching and assessment of critical thinking. A new stream-lined version of the HCTA interface will be available in March, 2014. A new test manual will accompany the rollout of the new interface, with norms based on over 1000 people and reports of additional research. The new manual will provide clear evidence of the equivalence of the alternative forms based on the larger standardization sample.

Anastasi raised the question of the relationship between critical thinking and intelligence (Anastasi, 2013, p. 14). A colleague (Butler, manuscript in preparation) just completed this study. I was delighted to learn that scores on a standardized intelligence test predicted real world outcomes, and adding scores from the HCTA significantly improved the predictions. Thus, critical thinking is not the same as intelligence and it contributed independently of intelligence to what people do (say they do) in real life situations. Stanovich (2009, p. 3) wrote “IQ tests are good measures of how well a person can hold beliefs in short-term memory and manipulate those beliefs, but they do not assess at all whether a person has the tendency to form beliefs rationally when presented
with evidence.” What we really want for our politicians, lawyers, doctors, citizens who vote, employees at all levels, and everyone else is to gauge their ability to think critically, which is largely absent from intelligence tests. In a series of experiments, Stanovich (2009, p. 39) has found that “Rational thinking can be surprisingly dissociated from intelligence.” This is exactly what Butler found in her most recent research.

The HCTA is constantly evolving, so I appreciate the helpful comments from all of the reviewers. I sincerely appreciate their kind words and generally positive assessment of the HCTA (with some reservations by Possin). In her doctoral dissertation on the assessment of critical thinking, Verburgh (2013) reported that students found the HCTA to be “interesting, fascinating and motivating” (p. 145). Perhaps that is the best review of all.

References


Author Information
Diane F. Halpern is the Dean of Social Sciences at the Minerva Schools at KGI and McElwee Family Professor of Psychology at Claremont McKenna College. She is a past president of the American Psychological Association and the Society for Teaching of Psychology. Diane has published hundreds of articles and over 20 books including, Thought and Knowledge: An Introduction to Critical Thinking (5th ed., 2014); Sex Differences in Cognitive Abilities (4th ed.), and Women at the Top: Powerful Leaders Tell Us How to Combine Work and Family (co-authored with Fanny Cheung). Her other recent books include Psychological Science (4th ed. with Michael Gazzaniga and Todd Heatherton) and the edited book, Undergraduate Education in Psychology: A Blueprint for the Future of the Discipline.

Diane’s most recent projects are the development of Operation ARA, a computerized game that teaches critical thinking and scientific reasoning (with Keith Millis at Northern Illinois University and Art Graesser at University of Memphis) and the Halpern Critical Thinking Assessment (Schuhfried Publishers) that uses multiple response formats, which allow test takers to demonstrate their ability to think about everyday topics using both constructed response and recognition formats.