DRAFT Jan. 10, 2011

Measurement and Meaning of Critical Thinking

Nathan R. Kuncel, Ph.D.

Department of Psychology
University of Minnesota

75 East River Rd.
Minneapolis MN 55455
Kunce001@umn.edu

Draft Report for the National Research Council's 21st Century Skills Workshop, Janurary, 2011, Irvine, CA. 1-10-2011

Measurement and Meaning of Critical Thinking

Numerous reports and studies indicate that the United States workforce is under-skilled in multiple areas given the labor demands of the economy (Galagan, 2010) and this state has existed for some time (SCANS, 1999). Although some research reveals that progress has been made for some groups in skills like basic reading and quantitative literacy (Hauser, 2005) the overall outlook is pessimistic. On average, the basic skills of the United States workforce, such as reading and math, lag behind other industrialized nations. Improvements would clearly benefit the United States (OECD, 2010). One area that has been discussed as a skill for improvement is critical thinking (CT). This report presents a quantitative research synthesis on the correlates of CT measures as well as their predictive power. This evidence and a qualitative analysis of the primary study literature are integrated to evaluate the construct validity of CT.

I conclude that the broad definition of CT (an independent, domain general, construct that applies across all tasks or jobs) is generally not supported by the literature or theory. Instead a narrower definition is well supported where critical thinking is a finite set of very specific skills (e.g., gamblers fallacy, law of large numbers, correlation vs. causation). These skills are useful for effective decision making for many, but by no means all, decisions or tasks. Their utility is further curtailed by task specific knowledge demands. That is, a decision maker often has to have specific knowledge to make more than trivial progress with a problem or decision. Despite these limitations, proficiency with these narrower skills would be useful for the country and workforce. However, this is true of many skills. CT skills, narrowly defined, are but one set of skills among many that are valuable. Investment in explicit training of CT skills will necessarily involve trade offs with training others skills. Careful thought should be given to emphasizing CT skills over other economically useful skills (e.g., writing, applied mathematics, mechanical aptitude). This is particularly true given that no longitudinal research to my knowledge has demonstrated greater long term economic value for CT gains resulting from a CT training program over programs training basic work skills like reading and mathematics.

Because of the size of the critical thinking literature, this review is organized into 5 sections. The first discusses definitions of CT and reviews the kinds of evidence that should be considered when evaluating the construct. A critical discussion of the difference between correlations of CT measures with individual differences and CT training effects is laid out in the first section. This distinction is necessary to correctly evaluate and interpret the construct validity evidence. The second section examines classic convergent and discriminant validity evidence for critical thinking measures as well as predictive correlations with important outcomes. I examine the question, "Do CT measures behave the way they should and predict important things?" The third section briefly evaluates research on the development and training of critical thinking both through explicit instruction and from experiences. The fourth critiques the domain generality claim for the critical thinking construct. That is, it addressed the question "Is critical thinking a broad skills or class of skills that is applicable to all decisions or problem or is it best considered as a narrow construct?" On the basis of the evidence I make recommendations both for educational policy and necessary future research in the fifth section.

Critical Thinking: Definitions and Needed Research Evidence

There are numerous definitions of critical thinking. To illustrate their scope, several have been compiled in Table 1. Although different in many important ways, they generally focus on evaluating information and making decisions. Some are very broad (e.g., Halpern, 1998) to the point of arguably including all of problem solving, judgment, and cognition. Others are more specific (Bangert-Drowns & Bankert, 1990) and focus on a particular class of tasks.

Often a distinction is made in the CT literature between specific skills and what might be called dispositions or attitudes. This is ultimately a "can do" versus "will do" distinction. The former is the ability to correctly execute a CT skill, however defined. The later is the interest and willingness to execute the skill. A second distinction is often made for "meta-cognitive" skills. How these are defined vary, but it sometimes includes problem recognition, self-regulation, testing assumptions and conclusions.

The concepts of intelligence and expertise predate critical thinking and a critical scientific question is the extent to which CT is actually distinguishable from them. Definitions of both intelligence and expert performance are also presented Table xx for comparison. Simply on the basis of definitions there appears to be considerable overlap. The definitions suggest that even if the concepts are independent that one may contribute to or be the developmental outcome of another. For example, intelligence might facilitate the development of expertise. Alternatively, CT might require some degree of expertise.

Table 1

Definitions of Critical Thinking, Intelligence, and Expert Performance

Critical Thinking Definitions	Expert Performance Definitions	Intelligence Definitions (General)
"…cognitive skills or strategies that increase the probability of a desirable outcome--in the long run, critical thinkers will have more desirable outcomes than "noncritical" thinkers….Critical thinking is purposeful, reasoned, and goal-directed. It is the kind of thinking involved in solving problems, formulating inferences, calculating likelihoods, and making decisions." (Halpern, 1998, pp. 450-451)	"…we define expert performance as consistently superior performance on a specified set of representative tasks for the domain that can be administered to any subject." (Ericsson & Charnes, 1994, p. 731)	"Intelligence is the ability to undertake activities that are characterized by (1) difficulty, (2) complexity, (3) abstractness, (4) economy, (5) adaptiveness to a goal, (6) social value, and (7) the emergence of originals, and to maintain such activities under conditions that demand a concentration of energy and a resistance to emotional forces." (Stoddard, 1941, p. 255)
"Critical thinking is reflective and reasonable thinking that is focused on deciding what to believe or do." (Ennis, 1985, p. 45)	"…expertise is, by definition, the possession of a large body of knowledge and procedural skill" (Chi, Glaser, & Rees, 1982, p. 8)	"The capacity of an individual to understand the world about him and his resourcefulness to cope with its challenges." (Wechsler, 1975, p. 139)
"We understand critical thinking to be 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, 1990, p. 1)		a general, unified capacity for acquiring knowledge, reasoning, and solving problems that is demonstrated in different ways …" (Weinberg, 1989, p. 98)
"Critical thinking, the ability and willingness to test the validity of propositions" (Bangert-Drowns & Bankert, 1990, p. 3)		

Fortunately several operationalizations of CT exist that permit examination of key questions regarding the nature of CT.. Measures of critical thinking include the Watson-Glaser Critical Thinking Appraisal (WGCTA), Cornell Critical Thinking Test (CCTT), California Critical Thinking Skills Test (CCTST), and the California Critical Thinking Disposition Inventory (CCTDI). There are a fairly sizable number of studies examining relationships between each of these measures and both individual correlates and outcomes, with the majority of studies using the WGCTA, CCTST, and CCTDI. The next section quantitatively synthesizes this data to try to answer a number of key research questions that are typically used to examine construct validity. To set the stage for the interpretation of these data, the reminder of this section reviews important principles for evaluating construct validity evidence, providing a guide for the major issues addressed in this evaluation of CT.

The starting question is whether or not various CT measures capture the same information about people. This is typically evaluated by looking at correlations between different CT measures. One limitation of this approach is that often people develop different measures for a reason and often that reason is a different opinion about the nature of the construct. This is a sign of an immature area of study.

The next question is the extent to which critical thinking measures measure the same thing as existing measures of intelligence or personality. This is typically evaluated by examining correlations between measures of other constructs. Similarly if the patterns of CT correlations with various measures (say personality traits) mirrors that of other conceptually similar constructs (e.g., intelligence) we will also be disinclined to believe that CT is a new concept. To be accepted as distinct, CT needs to have different patterns of relationships with other measures.

External correlates are examined next between CT measures and various important behavioral or outcome variables. CT is supposed to affect decision making and problem solving. It, therefore, should predict academic and work outcomes.

Having established basic relationships for CT measures, the third section of the report will examine the trainability and development of CT. Conceptualizations of CT include premises regarding the extent to which individuals can learn and apply CT to new situations. Examining the research base testing these premises is an important component of evaluating the CT construct.

Two caveats need to be mentioned about the correlational data and the training data. The first is that the presence of large correlations with other measure does not, necessarily, rule out the independence and importance of critical thinking. The most basic reason is that even two variables with large correlations (over .9) can still have meaningfully different relationships with other variables (McCornack, 1956). Having said this, large correlations with other variables argues against critical thinking. Although there are examples of variables correlating .7 and still behaving differently in meaningfully and practically important (Kuncel, 1999 for one example), within the field of individual differences psychology variables correlated near the limits of their reliability almost always behave the same. A second reason a positive correlation is not necessarily damning, is that two variables can be correlated due to development effects but not capture the same thing. For example, measures based on basic math skills (e.g., Algebra) are very strongly correlated with advanced math skills (Calculus). This does not mean that they are interchangeable in all settings. Tasks requiring calculus will tend to be better predicted by a Calculus assessment unless the two correlate near unity. This fact of learning leads to the third and more complex caveat. The third is the difference of observed correlation between individual differences "in situ" and the effect of training.

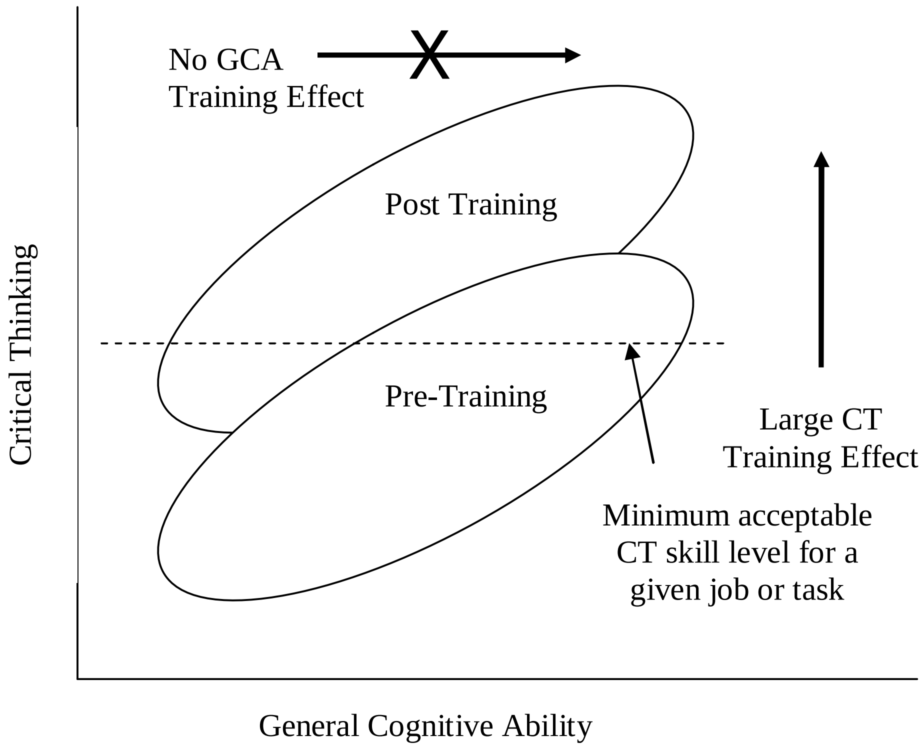
The later quantitative review of CT skills and their correlation with measures of general cognitive ability demonstrate moderate to large correlations. This effect is not necessarily a demonstration of the unimportance of CT or CT training. There is a potentially important difference between individual difference correlation and training effects. This is illustrated with a hypothetical example in Figure 1.

In the Figure, critical thinking scores and measures on a cognitive ability measures are depicted as strongly correlated. After training in CT, this correlation is preserved. In addition, CT scores show an mean increase (i.e., scores and skill go up). A score increase for the general cognitive ability measure is not observed suggesting that the training effect acts independently on CT. If this is the pattern typically observed, and CT is shown to be an important determinant of school and work performance or other life outcomes, then the results would make a good case for increased training of CT skills.

For example, we could assume there is a threshold for some minimum level of CT skills important for work and life success and mark this with a dashed line. Although skills levels cannot typically be marked with a cut off we could think of this threshold as a score that is associated with being able to effectively make a correct CT decision more often than not. In this example, although CT scores are predictable by general cognitive ability *both before and after the training*, training is shown to have an effect on the group's average performance and the percent of people above the threshold. The result of the training is that many of the "people" would move above the threshold for minimum skill competence.

Figure 1.

Hypothetical Example: Critical Thinking Training Effect Only after a Critical Thinking Training Program

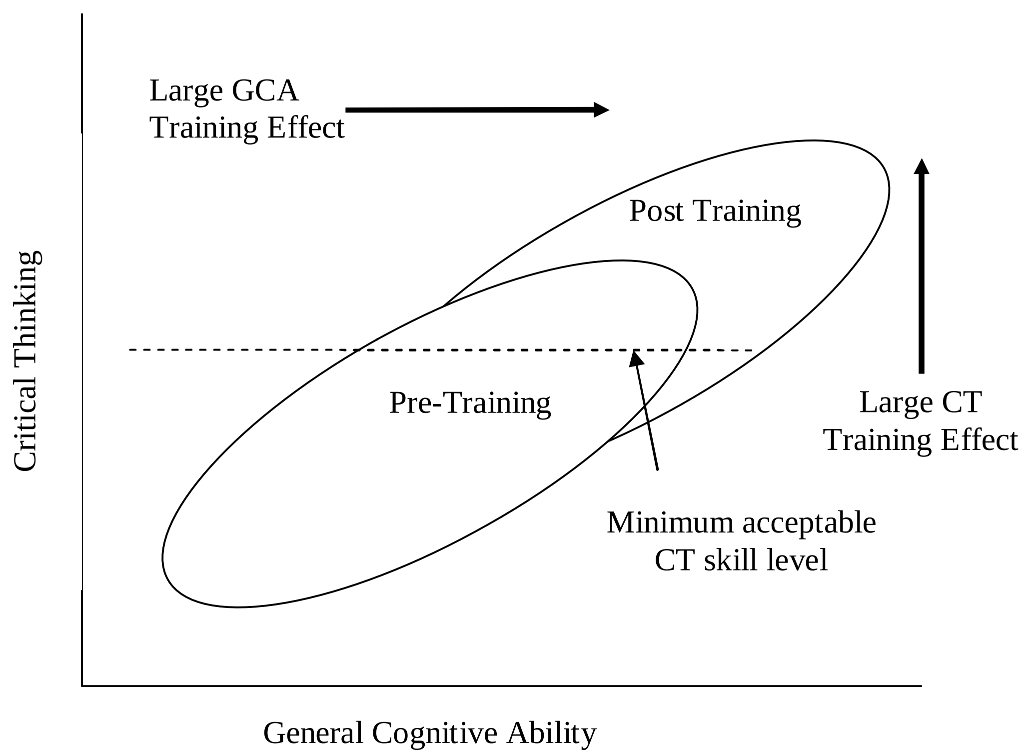


If this framework holds, it suggests that CT is associated with general mental ability but that it can be productively trained. Examples of this pattern of correlations and training effects exist in many areas, including both physical and knowledge domain. Many physical abilities are correlated but spending the time to training and practice specific skills results in skill improvement (catching, passing, running, strength). Another example would be nursing knowledge which is positively correlated with cognitive ability tests like SAT scores (e.g., Grossbach & Kuncel, 2010). Nursing students with high SAT scores tend to earn the highest grades and in nursing school and, subsequently tend to perform the best on nursing knowledge and licensure tests. This does not mean that nursing knowledge and SAT scores are the same and we certainly would not permit people to treat patients in a hospital simply based on strong verbal and math scores.

Figure 2 displays a different scenario where CT and cognitive ability both gain after training CT. This scenario would suggest that the two are one and the same. Such an extreme scenario is unlikely given that efforts to measure changes in CT scores often involve training that is test focused. Investigators deliver training that is specific to the test. This makes interpreting the results challenging because it is unclear whether gains have occurred on some underlying skill (CT) or merely knowledge relevant to the test and only the test.

Figure 2.

Hypothetical Example: Dual General Cognitive Ability and Critical Thinking Training Effect after a Critical Thinking Training Program



This leads us to what would be the acid test of the utility of CT training and even the CT construct itself. Gains on a test of critical thinking are interesting, but not compelling unless:

1. the gains predict important outcomes,
2. the gains cannot be explained by gains on other individual differences (e.g., intelligence),
3. the value of gains on critical thinking exceed the value of gains from other basic skill interventions, like being skilled at writing.

To the best of my knowledge, questions 1 and 3 have never been answered.

Correlates of Critical Thinking: A Quantitative Review

Convergent Validity Correlations

Evidence of convergent validity can be obtained from the few studies that correlate different critical thinking measures. For this report, 5 studies were located that reported correlations among different critical thinking measures. They correlated, on average, .41 with each other based on a total sample of 1,507 (see Table 2).

This relationship is on the low side. For example, the SAT-Critical Reading Tests and the SAT-Writing test, two tests that should be related to the fundamentally same construct, correlate .71 with each other (Kobrin, et al., 2008). However, authors have argued that different measures of critical thinking conceptualize the construct differently.

Table 2

Meta-analysis of correlations between different critical thinking skills measures

Critical Thinking Skills						
	N	k	r_{obs}	SD_{obs}	SD_{ρ}	80% cred.
Critical Thinking Skills	1,507	5	**.41**	.06	.03	.37

The question of convergent validity also can be approached by examining correlations between CT and superstitious and paranormal beliefs. Superstitious and paranormal beliefs have been cited as examples of poor critical thinking by scholars in that area (e.g, Halperin, 1998). As such measures of CT should be negatively associated with superstitious and paranormal beliefs. Interestingly, both cognitive ability measures and critical thinking measures are negatively associated with superstitious and paranormal beliefs and the relationships are both small and comparable in size. Table 3 presents results for correlations between both types of measures and scales developed to measure belief in superstitions and paranormal phenomena (e.g., ESP and UFO abduction). Meta-analytic estimates place the relationship for these beliefs with cognitive ability and critical thinking at r = -.13 and r = -.19, respectively.

Table 3

Meta-analysis of correlations between superstitious or paranormal beliefs and cognitive ability measures or critical thinking measures

	Superstitious/Paranormal Beliefs					
	N	k	r_{obs}	SD_{obs}	SD_{ρ}	Lower 90%
Cognitive Ability Measures	1,690	7	**-.13**	.18	.16	-.34
Critical Thinking	497	5	**-.19**	.09	.00	-.19

Discriminant Validity: Correlations with Cognitive Ability and Personality

A literature search yielded 19 correlations between CT skills and traditional measures of cognitive abilities (e.g., MAT, SAT). The average correlation of .48 was obtained from a sizable sample of data. This correlation suggests that critical thinking measures behave similarly to other cognitive ability measures and correlate about as well with them as they do with each other. For example, the SAT-Critical Reading and the SAT-Math correlate .50 with each other. If we consider that these two SAT tests are more reliable than the typical critical thinking test it becomes clear that CT measures are correlated about as well with each other (Table 2) as they are with other cognitive ability measures (Table 4). CT dispositions were more weakly associated with cognitive ability measures (r = .21) but were still clearly related to cognitive ability.

Table 4

Meta-Analysis of correlations between critical thinking measures and traditional cognitive ability measures

Traditional Cognitive Ability Measures						
	N	k	r_{obs}	SD_{obs}	SD_{ρ}	80% cred.
Critical Thinking Skills	6,461	19	**.48**	.14	.13	.31
Critical Thinking Dispositions	1,425	5	**.21**	.05	.00	.21

Three independent samples with correlations between CT skills and the personality characteristic, *Openness to Experience,* were analyzed. This trait measures…On average CT skills had modest but meaningful correlations with openness to experience (see Table 5, r = .24). A nearly identical value was obtained for dispositional measures of CT (r = .23). Other personality characteristics had relationship of similar magnitude (e.g., *Social Potency* from the CPI) but too few studies measured the same personality trait to permit research syntheses. Given the definitions of critical thinking, particularly the dispositional aspects, these correlations are expected. However, traditional cognitive ability measures behave similarly to critical thinking skills measures. Measures of general ability are also correlated with measures of Openness to Experience with meta-analytic estimates around .30 (Ackerman & Heggestad, 1997)

Table 5

Meta-analysis of correlations between critical thinking measures and personality measures of openness to experience

Openness To Experience						
	N	k	r_{obs}	SD_{obs}	SD_{ρ}	Lower 90%

Critical Thinking Skills	647	3	**.24**	.11	.08	.13
Critical Thinking Dispositions	582	3	**.23**	.12	.09	.11

In summary, the correlation between one measure of critical thinking skills and a different measure of critical thinking skills is lower than the correlation with traditional measures of cognitive ability. If nothing else this evidence suggests that any given measure of critical thinking is likely to have more in common with a traditional measure of verbal or quantitative skill than it does with an alternate CT measure. At least some of this may be attributed to, on average, lower measurement reliability for CT measures than many nationally administered measures like the SAT or ACT. But reliability does not fully account for the difference. Given the sizable overlap and poor discriminant validity, it is particularly important to determine if critical thinking is correlated differently than cognitive ability with important external variables like grades or job performance.

External Predictive Correlations

Demonstrating that a class of measures are strongly related to each other and have much smaller (ideally zero) relationships with existing measures is a good starting point for a young construct. The next step is demonstrating that the new construct is psychologically important. One method for doing this is demonstrating predictive power for important behaviors or outcomes. For example, there is unambiguous evidence that cognitive abilities and personality characteristics are associated with academic and work performance, occupational attainment, and even increased longevity (see Kuncel & Hezlett, 2010; Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2005 for two reviews among many). Given the language surrounding critical thinking, measures of CT should demonstrate important correlations with a range of outcomes

External Correlations: Grades and Grade Point Averages

Grades were the first variable examined. Grades earned in higher education should be influenced by critical thinking both within and outside of the classroom (e.g., managing time, goal commitments). The data from 12 independent samples with 2,876 subjects support a positive correlation between critical thinking and grades earned in higher education (principally GPAs, see Table 6). However, this correlation is similar to correlations obtained for measures like the SAT which was reported to have an average correlation with 1st year college GPA between .26 to .33 for the individual scales and .35 when the SAT scales are combined (Kobrin, et al., 2008). Dispositional measures fared similarly with an average correlation of .24.

Table 6

Meta-analysis of correlations between critical thinking measures and grades earned

	Grades					
	N	k	r_{obs}	SD_{obs}	SD_{ρ}	80% cred.
Critical Thinking Skills	2,876	12	**.27**	.10	.07	.18
Critical Thinking Dispositions	2,250	7	**.24**	.12	.10	.10

External Correlations: Job Performance

There are very limited data that quantify the relationship between CT measures and subsequent job performance. Three studies on the Watson Glaser were located (see Table 7). These yielded an average correlation of .32 with supervisory ratings of job performance (N = 293).

Simulated nursing performance through clinical decision-making assessments appear to have some correlation with critical thinking. Shin (1998) examined CT scores in a group of nursing students (N=234) and their relationship with a simulated clinical decision making task. The relationships was positive (r = .19) but small. Brooks and Shepard (1990) report similar (r = .25) results.

Although there are limited data, critical thinking appears to be associated with job performance. These correlations are similar to observed correlations for cognitive ability measures and job performance.

Table 7

Meta-analysis of correlations between critical thinking measures and job performance

Supervisory Rating of Job Performance						
	N	k	r_{obs}	SD_{obs}	SD_{ρ}	80% cred.
Watson-Glaser	293	3	**.32**	.04	.00	.32

Incremental Predictive Correlations over Existing Constructs

It is important to note that critical thinking measures could have large and important predictive relationships with a wide range of important personal and occupational outcomes and still fail as measures. Indeed my review and synthesis of the literature suggests that critical thinking measures are solid predictors of academic and work outcomes. However, if CT measures fail to add meaningfully to existing measures of cognitive abilities and personality then the evidence would not support their independent importance and the field would need to be concerned about there being a different, but redundant, name for a known set of individual differences. This would likely occur if they were strongly correlated with other measures and were not more predictive than the existing predictors. This is the very pattern of evidence that has been obtained from the correlational data.

Passive and Explicit Efforts to Improve Critical Thinking

This section examines how environment and training affect critical thinking gain. A number of meta-analyses and reviews have been conducted to summarize this literature. Typically the instruction or training research examines pre-post gains on a critical thinking test after a course that is either focused exclusively on critical thinking or infuses critical thinking concepts into the course materials. Research examining passive effects either focuses on the overall gains in critical thinking scores after different amount of higher education (i.e., number of years) or after involvement in extra-curricular activities. On average, explicit instruction yields the largest gains while extra-curricular involvement and years in higher education tend to produce low to zero gains.

Critical Thinking Gains Associated with Student Involvement

Student involvement in activities is associated with nominal increases in critical thinking scores. Involvement was correlated .04 with gains in critical thinking on one of 4 standard measures (Gellin, 2003). The definition of student involvement was highly variable. A narrower examination of participation in clubs and organizations yielded a correlation of .11. In otherwords, a student who was high involved in clubs and organizations (Z=2) would be expected to have a 0.22 standard deviation increase in test scores. These results should not, however, be interpreted as causal as the research designs were based on observed correlations rather than experimental manipulations.

Critical Thinking Gains Associated with Training or Instruction

To be included in final draft

Critical Thinking Gains Associated with Years of College or Graduate Education

To be included in final draft

"This result is disconcerting for those who hold the view that growth in CT ability is a natural byproduct of either a college education or the maturation which occurs in young adulthood" (pp. 4-5, Facione,

Factors Associated with Critical Thinking Gains

Although the type of training seems to have some effect on critical thinking test score gains little research has examined the characteristics of students that are associated with larger gains. Previous research has demonstrated consistent and strong relationships between measures of general cognitive ability and subsequent learning in college and graduate school, laboratory skill acquisition settings, and in job training across a wide range of civilian and military occupations (Campbell & Kuncel, 2001). It would be surprising if that did not hold for critical thinking training. Most research on CT does not include correlates of skill gain, however, research on critical thinking suggests that CT gains are predicted by GCA as well. For example, Glaser (1942) reported a .33 correlation between an IQ measure and CT gains after instruction. This complicates the simplified examples presented in Figures 1 and 2 as those who are higher in GCA are not only likely to have stronger CT test scores initially, they are more likely to acquire even more of them with instruction.

Critical Thinking Gains Longitudinally Associated with Success Gains

A strong test of the value of teaching CT skills would be evidence that after training the gain in critical thinking skills is associated with improvements in school success, work performance, income, or other measures of life success (e.g., longevity).

Although studies may exist, to the best of my knowledge, there is no evidence testing this critical research question. As a result there are no comparative effectiveness studies comparing CT training with other work relevant skills.

Domain General Critical Thinking Skills

An ongoing debate in the critical thinking skills literature is a discussion of whether or not a person can think critically without having domain specific knowledge. Can we teach a general skill that is applied across all reasoning situations. For example, can a person think critically about arguments for different national economic policies without understanding macro-economics or even the current economic state of the country. At one extreme it seems clear that people cannot think critically about topics for which they have zero knowledge and their reasoning skills are intimately tied to the knowledge domain. The author has effectively zero basis for making judgments about how to conduct or even prioritize different experiments for CERN’s Large Hadron Collider. Few people in the world understand the topic of particle physics sufficiently to make more than trivial arguments or decisions. On the other hand, perhaps most people could try to make a good decision about which among a few medical treatments would best meet their needs (although even this is often beyond most people as typically presented, see Gigerenzer, Gaissmaier, Kurz-Milcke, Schwartz, & Woloshin, 2007).

I will argue that for the purposes of economic and societal improvement in the United States, critical thinking is best conceptualized not as a domain general skill but as a limited set of very specific skills that together can be useful for considering a relatively broad set of socially and personally important topics. However, this set of specific skills is not universally useful for all tasks, far from it. Critical thinking has become overly generalized to the point of being unwieldy. My position is not that some of the skills identified are unimportant but that they are far more specific and narrowly useful than is often suggested.

Because CT skills are not universally applicable to all problems, training them involves trade-offs and may be substantially less productive in some situations when training other skills is more important. In many settings the literatures on instruction systems design (e.g., Campbell & Kuncel, 2001) and the expertise literature (e.g., Ericsson & Charnes, 1994) may provide better insight into how to improve the performance of workers for the same dollar cost and time invested.

Because specific CT skills are not universally applicable to most, let alone all problems, training them involves trade-offs and may be counterproductive in some situations when training other skills is more important. In many settings the literatures on instruction systems design (e.g., Campbell & Kuncel, 2001) and the expertise literature (e.g., Ericsson & Charnes, 1994) may provide better insight into how to improve the performance of workers for the same dollar cost and time invested.

An examination of classic studies on critical thinking can illustrate how CT skills are not domain general. In the Nisbett et al (1987) paper they present research on a number of studies examining training effects for specific skills. It is key to note that although the paper is titled “Teaching Reasoning” Nisbett and colleagues conclude that “…transfer applies only insofar as there are common identical elements.” (p. 631). They observed that gains on one skill were not associated with gains on other skills. This indicates that these skills are more modular than is suggested by some authors and reasoning skills would be acquired piece by discrete piece.

One study in the paper demonstrating transfer across domains is presented in Figure 3. Here we have people trained in the law of large number within a domain and performance is examined later for a law of large numbers problem but in a different domain (sports versus ability testing). After learning about the law of large numbers people show an initial improvement and can apply it somewhat successfully to a new question form a different domain after a period of two weeks. But these people have not been taught universal reasoning skills. Home improvement can demonstrate how.

Figure 3

Figure from Nisbett et al. (1987) displaying correct use of the law of large numbers across domains in a control group experimental group and after a delay

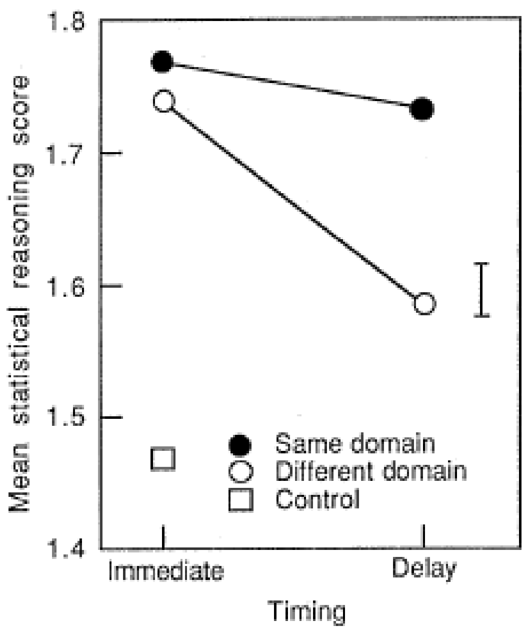


Fig. 1. Mean statistical reasoning score as a function of training and problem domain. Vertical bar represents one standard error of the mean ($N = 231$).

Critical Thinking in Construction and Home Improvement

A practical illustration can help clarify. The author was recently involved in two projects (Figure 4a, b). One was installing a walk in closet organizer system and the second was decking, lighting, and installing a pull down staircase into a garage attic. Many decisions needed to be made including the right materials to use, how to get materials into the spaces, and what are the risks, limitations, and costs of different approaches. More long range thinking was needed including determining how the spaces be used (affecting how they are laid out and constructed) or what configurations would be useful now given the family in question and appealing to a possible unknown buyer in the future.

Figure 4a,b

Closet organization and garage attic installation

No reasonable case can be made that either project required considering the law of large numbers. The gambler fallacy, sampling bias, or issues around affirming the consequent were also not used. Performance on these projects would not be likely to improve after training on the law of large numbers. However, people do work like this every day. So skill at the law of large numbers allows one to be effective solving problems involving the law of large numbers and not much else. Training and apprenticing as a carpenter and taking shop classes would help with effective decision making for these tasks.

Many Critical Thinking Skills are Often Field Specific

The Nisbett et al (1987) study also illustrates the field specificity of specific reasoning and thinking skills. They conduct some interesting research examining what they label as statistical and methodological reasoning in graduate students in law, medicine, psychology, and chemistry (see Figure 5). Both in cross-sectional and longitudinal research psychology students gain the most in these skills during graduate school while chemistry graduate students have little to zero gain.

Figure 5

Gains in statistical and methodological reasoning

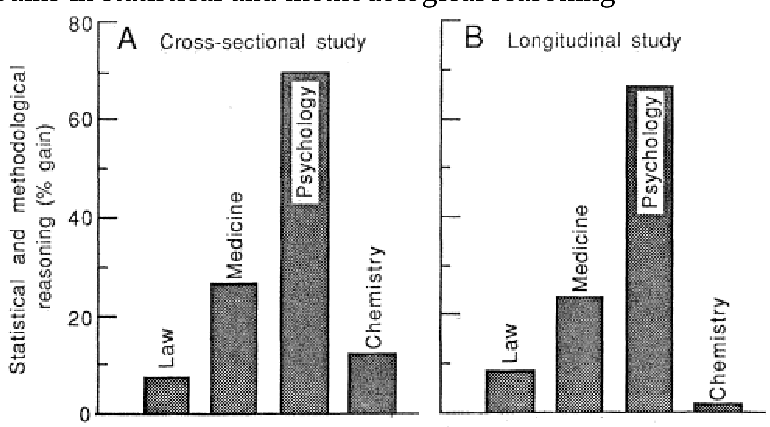


Fig. 2. Percentage of change in statistical and methodological reasoning score after 2 years of study as a function of graduate discipline. (**A**) The cross-sectional study examined first-year students and simultaneously enrolled third-year students. Sample sizes for first-year students were law, 213; medicine, 127; psychology, 25; and chemistry, 31. The sample sizes for third-year students were 50, 48, 33, and 26, respectively. (**B**) The longitudinal study examined the same students at the beginning of their first year and at the beginning of their third year. Sample sizes were law, 77; medicine, 87; psychology, 24; and chemistry, 18.

Do we really believe that PhD level chemists don't learn how to reason? After all, on average students applying for doctoral work in chemistry score appreciably higher than psychology doctoral students on tests of quantitative reasoning. Careful examination of the Nisbett et al (1987) study reveals the problem. The study tested students on statistical and methodological reasoning including "methodological reasoning dealing with different types of confounded variable problems, for example, self-selection problems (26), sample bias problems…" (p. 630). What is a self-selection effect in chemistry? How does sample bias play when a mole of Hydrogen-2 deuterium is a mole of deuterium is a mole of deuterium? These are not critical thinking skills of importance to the profession of chemistry and we might argue that their time is better spent becoming expert chemists. I present an initial list of enormously important critical thinking skills for chemists, engineers, and physical scientists in Table 8. The Laws of Thermodynamics are necessary for thinking at all effectively about an enormous range of problems from applied work in industry to theoretical problems. Psychology students do not learn these critical thinking skills and we can reasonably conclude psychology graduate students don't gain in their skill applying these natural laws to a range of problems during graduate training.

Table 8

Some Critical Thinking Skills for Chemists, Engineers, and Physical Scientists

Zeroth Law of Thermodynamics: Thermodynamic equilibrium and temperature
First Law of Thermodynamics: Work, heat, and energy
Second Law of Thermodynamics: Entropy

Note that this does not mean that the skills examined by Nisbett and colleagues are unimportant. They are very important for thinking critically about research or situations that involve self-selection effects or sampling bias. It might also be desirable as citizens and consumers of the news media for chemists to be more skilled at these questions. But we are now talking about tradeoffs between very specific skills rather than a discussion of the improving Critical Thinking (capital "C", capital "T") for chemists.

Making Trivial Progress with Specific Critical Thinking Skills: Three Examples

Three examples can help illustrate the limitations of the critical thinking concept as a domain general set of skills that are independent of specific knowledge. An example of more everyday critical thinking might be discussions of national health care, a health claim from Dr. Oz, or Glenn Beck's argument that the Mercury dime was evidence of Fascism creeping into US society as far back as 1916.

While discussing health care in the US, Giuliani, a presidential candidate, claimed US health care was superior based on evidence for prostate cancer death 5 years after diagnosis (Gigerenzer et al., 2007). There was a 44% 5 year survival rate in the United Kingdom compared to 82% in the United States. On the surface this suggests that outcomes are better, almost a factor of 2, in the US with its health care system. The actual data don't support this conclusion. The problem is that diagnosis occurs far earlier in the US due to the PSA screening test while diagnosis was based on physical symptoms in the UK which typically occur much later. The actual death rates, making a more direct comparison, are effectively identical across the two countries with the US spending more money on early screening. These discussions occurred constantly during the Correct evaluation of the evidence require a fair amount of specific knowledge of medical research and some modest mathematic skill.

Dr. Oz is a popular television and print physician and health advice giver. When asked in the Sept, 2009 *Readers Digest*, “What are the two most important things people should do to keep themselves young?” His first things was, “Walk. When you can’t walk a quarter mile in five minutes, your chance of dying within three years goes up dramatically.” Oz, perhaps knowingly, confuses correlation and causation to provide a dramatic, but poor, piece of evidence. The evidence ignores that the same data also supports both reverse causal explanations and unobserved causal variables. For example, when people are very sick and likely to die soon, they often cannot walk a quarter mile in five minutes or factors like obesity contribute to both to premature death and low walking capacity. Fortunately, walking is good exercise but this is known from superior experimental research. Here evaluating the evidence does requires a very specific skill (alternative explanations for correlational data). However two things need to be noted. First, making good progress with evaluating the claim and generating alternative explanations requires some general knowledge of medicine. Second, this skill would not be useful for evaluating the following claim.

Beck is a popular pundit and entertainer in the United States. His claim about the dime is based on it including the fasces (a bound bundle of sticks with an ax) as decoration on the reverse side (see Figure 5). The fasces became a symbol of Fascism in the 20th century. Thinking about this claim requires a fair amount of specific knowledge.

First is that cultures have reused and recycled symbols and names many times. How many Nike sportswear owners know that Nike was a Greek goddess?

Figure 5

Reverse side of a mercury dime

Similarly ancient Hindu's and numerous other cultures are not time traveling Nazi's despite using swastika as a decorative symbol. In the case of the fasces, it is also an ancient symbol dating back at least as far as the Roman Republic of antiquity. Fascism is often dated to the publication of The Manifesto of the Fasci of Combat in 1919, three years after the creation of the Mercury dime. Aside from global skepticism this everyday problem requires specific knowledge.

Expertise as an Alternative Perspective

Performance failures in highly skilled professions are sometimes attributed to innate ability or a lack of global critical thinking (see Ericsson, 2004 for a review in medicine). New nurses or veterinarians are criticized for low critical thinking skills and failure to made, for example, accurate diagnoses. An alternative perspective is that these individuals have not developed high levels of expertise. The prescriptions in this case would not be global critical skills training but more practice and instruction within the discipline. Making a correct diagnosis for an infrequently occurring disease or condition requires many opportunities to make such a diagnosis and receive feedback on the accuracy of one's decision. Accuracy and practice follow lawful patterns in physicians based on the frequency with which they encounter diseases and specialists become especially accurate at diseases within their area of specialty (Ericsson, 2004).

The problem is not having a global skill but lacking practice to develop a specific skills. Even for meta-cognitive skills involving management and self-monitoring of task performance, some theories argue against the notion of teaching general meta-cognitive skills arguing that these broader skills only developer ***after*** people developing specific expertise within a number of fields (for a review see Schraw & Moshman, 1995). In other words, people really don't become critical thinkers they become critical thinkers for a domain. Ericsson and Charnes (1994) argued "There is no reason to believe that changes in the structure of human performance and skill are restricted to the traditional domains of expertise. Similar changes should be expected in many everyday activities, such as thinking, comprehension, and problem solving, studied in general psychology." (p. 745).

Unacknowledged Trade-Offs

Whether the position presented here is accepted or not, it is still the case that there is only so much time for instruction, practice, and opportunities to apply new knowledge across multiple situations. Both explicit training programs for critical thinking and the attempt to incorporate CT instruction into the content of courses must ultimately be regarded as a trade-off between training CT skills and training something else. In some settings, like professional training, a needs assessment should be conducted to address the actual nature of the training need. If we believe that nurses lack "critical thinking" it is important to determine if the training need is actually the content of a CT course or is the need to have more knowledge and practice identifying declining symptoms in an ICU, applied hands on experience with simulated patients, or additional specific instruction in how to interpret medical research. Often descriptions of poor critical thinking for highly trained workers sounds like a lack of expertise rather than issues with affirming the consequent.

Conclusion

The evidence collected here suggests that critical thinking measures behave very similarly to measures of cognitive ability. Scores on these measures can be improved by direct instruction but show small gains with college education or involvement. No evidence demonstrates long term improvement for critical life outcomes due to gains on critical thinking measures. As a result there is also no evidence comparing the effectiveness of CT training to other job or life skill training program. Finally, theh suggestion that good critical thinking is a knowledge independent domain general skill is not supported by the evidence. Instead they are a class of relatively independent skills that, like all skills, are relevant for specific tasks. The decision to train these skills should be weighed against other training options including reading, writing, mathematics, civics, and general scientific knowledge. A report card for critical thinking is presented in Table 9. Its overall status would be greatly improved if compelling evidence existing supporting some of the most critical research questions.

The concerns raised here about critical thinking, broadly defined, should not be interpreted as an argument against efforts to improve the quality of instruction or efforts to increase transfer of training, where appropriate. It is certainly the case that different training methods will yield people who differ in their ability to function effectively and display behaviors more or less in line with what is called good critical thinking. For example, rote memorization of anatomical components is less likely to provide a good understanding of the body than training that attempts to help chunk the knowledge. Similarly, many of the specific skills identified in the CT literature are useful, for specific tasks.

Table 9

Summary of Critical Thinking Literature Review

Key Consideration	Grade	Notes
Convergent Validity	**P**-	Different interpretations of CT may limit inter-correlations. Correlations lower in magnitude than correlations with cognitive measures.
Discriminant Validity	**P**-	Large correlations with intelligence and personality that are consistent in pattern with existing measures.
Incremental Predictive Power	-	Correlations with existing intelligence measures and equal or lower predictive power indicate a lack of incremental power.
Training Gains	+	Consistent gains for specific measures, poor gains for education and experiences during college.
Training Gains Yield Success Gains	No Data	No evidence of long term positive gains or positive longitudinal outcomes.
Training Gains Yield Global CT Gains	No Data	Modest transfer of training effects for same skill in a different context.
Evidence of Knowledge Independent Critical Thinking Skills	+ or **P**-	Depends on how CT is defined.
Evidence for Universal Applicability	-	Beyond link to GCA data and theory do not support universal applicability.

References

Ackerman, P. L., & Heggestad, E. D. (1997). Intelligencce, personality, and interests: Evidence for overlapping traits. *Psychological Bulletin, 121*, 219-245.

Brooks, K. L., & Shepard, J. M. (1990). The use of the Watson Glaser Critical Thinking Appraisal to predict success in a baccularate nursing program. Journal of Nursing Education, 26, 278-281.

Campbell, J. P., & Kuncel, N. R. (2001). Individual and team training. In N. Anderson, D. S. Ones, H. Sinangil, & C. Viswesvaran, (Eds.), *Handbook of Industrial, Work, and Organizational Psychology* (pp. 287-. London, UK: Sage Publications.

Chi, M. T. H., Glaser, R., & Rees, E. (1982). Expertise in problem solving. In R. J. Sternberg (Ed.), *Advances in the Psychology of Human Intelligence*. (Vol. 1). Hillsdale, NJ: Erlbaum.

Ejiogu, K. C., Rose, M., Yang, Z., and Trent, J. (2007). Incremental Validity of Numerical Reasoning over Critical Thinking. Poster session presented at the 115th annual convention of the American Psychological Association, San Francisco, CA, August 17-20.

Ennis, R. H. A logical basis for measuring critical thinking skills. Educational Leadership. XX , 44-48

Ericsson, K. A. (2004). Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. *Academic Medicine, 79*, 70-81

Facione, P. A. (1990). The California Critical Thinking Skills Tests – College Level. Technical report #2. Factors predictive of CT skills. Millbrae, CA: California Academic Press. Retrived from ERIC documents # 015819

Galagan, P. (2010). *Bridging the Skills Gap*. Alexandra, VA: American Society Training and Development

Gellin, A. (2003). The effect of undergraduate student involvement on critical thinking: A meta-analysis of the literature 1991-2000. *Journal of College Student Development, 44*, 746-762.

Generality of thinking skills is inversely related to their usefulness. Gives the examples of not believing everything one hears.

Gigerenzer, G., Gaissmaier, W., Kurz-Milcke, E., Schwartz, L. M., & Woloshin, S. (2007). Helping doctors and patients make sense of health statistics. *Psychological Science in the Public Interest, 8*, 53-96.

Glaser Phi kappa phi journal

Glaser, R. (1984). Education and knowledge: The role of knowledge. *American Psychologist, 39*, 93-104.

Gordon, M. (1980). Predictive strategies in diagnostic tests. Nursing research, 29, 39-45.

Gunning, C. 1981 diss

Hauser, R. M. (2005). *Measuring Literacy: Performance Levels for Adults.* Washington DC: National Academies Press.

Holzemer, W. L., & McLaughlin, F. E. (1988). Concurrent validity of clinical situation. Western Journal of Nursing Research, 10, 78-83.

Ketefian 1981

Kobrin, J. L., Patterson, B. F., Shaw, E. J., Mattern, K. D., & Barbuti, S. M. (2008). The validity of the SAT for predicting first year college grade average. Research Report 2008-5. New York: The College Board.

McMillian, J. H. (1987). Enhancing college students' critical thinking: A review of studies. *Research in Higher Education, 26*, 6-29

McPeck, J. (1992). Critical thinking and subject specificity: A reply to Ennis. *Educational Researcher, 19*, 10-12.

McPeck, J. (1992). Thoughts on subject specificity. In S. P. Norris (Ed.). The Generalizability of Critical Thinking (pp. 198-205). New York: Teachers College Press.

Nisbett, R. E., Fong, G. T., Lehman, D. R., Cheng, P. W. (1987). Teaching reasoning. *Science, 238*, 625-631.

OECD (2010). PISA 2009 Results: Executive Summary.

Rubinstein, M. F., & Firstenberg, I. R. (1987). Tools for thinking. New Directions for Teaching and Learning, 1987, 23

Sa, W. C., West, R. F., & Stanovich, K. E. (1999) The domain specificity and generality of belief bias: Searching for a generalizable critical thinking skill. Journal of Educational Psychology, 91, 497-510.

SCANS (1999). *Skills and Tasks for Jobs*. Washington DC: US Department of Labor.

Schraw,G.&Moshman, D. (1995).Metacognitive theories. *Educational Psychological Review* 7, 351–371.

Shin, K. R. (1998). Critical thinking ability and clinical decision making skills among senior nursing students in associate and bacculate nursing programmes in Korea. *Journal of Advanced Nursing, 27*, 414-418.

Stice, J. E. (XX) Teaching Problem Sovlving.

Tanner, C. A. 1977 dissertation

Tiessen 1987

Watson, G., & Glaser, E. M. (2008). *Watson-Glaser Critical Thinking Appraisal*. Pearson.

Weinberg, R. A. (1989). Intelligence and IQ. *American Psychologist, 44*, 98-104.

Woods, D. R. How might I teach problem solving? In J. E. Sum (ed.). Developing Critical Thinking and Problem Solving Abilities. New Directions for Teaching and Learning, no. 30. San Francisco: Josrey-Bass. Summer 198