Thinking Skills and Creativity 25 (2017) 38–46

Contents lists available at ScienceDirect

Thinking Skills and Creativity

journal homepage: www.elsevier.com/locate/tsc

Predicting real-world outcomes: Critical thinking ability is a better predictor of life decisions than intelligence

Heather A. Butler[a,*,1], Christopher Pentoney[b], Mabelle P. Bong[c]

[a] Psychology Department, California State University Dominguez Hills, United States
[b] Division of Social and Behavioral Science, Claremont Graduate University, United States
[c] Psychology Department, Claremont McKenna College, United States

ARTICLE INFO

Keywords:
Critical thinking
Intelligence
IQ
Real-world outcomes
RWO
INSBAT
Halpern critical thinking assessment
HCTA

ABSTRACT

We all probably know someone who is very intelligent, but does blatantly stupid things. Despite evidence that intelligence predicts a variety of life outcomes, the relationship between intelligence and good thinking is less clear. This research explored whether critical thinking ability or intelligence was the better predictor of real life events. Community adults and college students (n = 244) completed a critical thinking assessment, an intelligence test, and an inventory of life events. Individuals with higher critical thinking scores and higher IQs reported fewer negative life events. Critical thinking more strongly predicted life events than intelligence and significantly added to the variance explained by IQ. There is ample evidence that critical thinking can be taught, so there is hope that teaching critical thinking skills might prevent the occurrence of negative life events. We advocate for critical thinking instruction as a way to create a better future for everyone.

Although the assessment of human intelligence has been hailed as psychology's greatest achievement, it also remains one of the most controversial (e.g., Nisbett et al., 2012). Critics of intelligence testing highlight the weaknesses inherent in attempting to measure such a complex construct, as well as the potential misuses of IQ scores. Most notably, scores on intelligence tests have been misused in the past to support racism (Vera, Feagin, & Gordon, 1995). The misuse of IQ scores, whether deliberate or inadvertent, is a serious issue, and we urge everyone to be vigilant for potential misuse of these measures. Aside from the clear misuse of intelligence testing, there remains serious concerns over the validity of the test scores and what conclusion can be drawn from them.

Critics argue that no single number can capture the complexity of human intelligence (Mayer, Caruso, Panter, & Salovey, 2012). In fact, one of the major controversies in this area of research is whether intelligence is a single construct or whether there are multiple intelligences. Other critics of intelligence testing have argued that environmental influences are critical determinants of intelligence (Nisbett et al., 2012). Like all psychometric tests, intelligence testing is often imprecise, which limits the extent to which we can rely on them when making critical academic and employment decisions. Each of these objections have some merit, but it is also true that intelligence, as assessed with IQ measures, has a utilitarian value.

Proponents of intelligence testing point to its concurrent and predictive validity for a wide range of diverse outcomes, including long-term job performance (Hunter & Schmidt, 1996), grades in school (Furnham & Monsen, 2009), academic achievement (Soars, Lemos, Primi, & Almeida, 2015), many types of criminal behaviors (negative relationship; Schwartz, Savolainen, Aaltonen, Merikukka, Paananem, & Gissler, 2015), including juvenile delinquency (Gordon, 1986); learning outcomes in college and work

* Corresponding author at: Department of Psychology, California State University, Dominguez Hills, 1000 E Victoria St Carson, CA 90747, United States.
E-mail address: hbutler@csudh.edu (H.A. Butler).
[1] Department of Psychology, California State University, Dominguez Hills, United States.

http://dx.doi.org/10.1016/j.tsc.2017.06.005
Received 1 July 2016; Received in revised form 8 June 2017; Accepted 21 June 2017
Available online 01 July 2017
1871-1871/

settings (Kuncel, Hezlett, & Ones, 2004; Kuncel, Ones, & Sackett, 2010), and for a highly gifted sample of adolescents, getting a doctorate and tenure in academic positions (Lubinski, Benbow, Webb, Bleske & Rechek, 2006). Despite evidence that intelligence has been associated with a variety of everyday life outcomes, the relationship between intelligence and good thinking is less clear.

If you think about the last time that you used the word "intelligent" to describe someone, you probably meant to say that the person knows a great deal of information and makes well-reasoned decisions. This everyday use of the term intelligence is somewhat at odds with how intelligence is actually measured. For the purpose of this paper, we are operationalizing intelligence as standardized assessments of intelligence that yield IQ scores. Consider, for example, the test structure for the Wechsler Adult Intelligence Scale (WAIS-IV). It includes vocabulary questions, visuospatial puzzles to solve, digits to recall in order both forward and backward, and a visual search for symbols (Wechsler, 1939). These subtests measure important components of intelligence, but individually or collectively, they do not capture the everyday meaning of intelligence, which includes making well-reasoned decisions, supporting conclusions with evidence, thinking in an unbiased manner, avoiding common well-documented biases in thinking such as not considering regression to the mean, weighting evidence that conforms to an existing belief more heavily than evidence that does not, and being misled by the way information is framed. A corollary to the problem that intelligence tests do not test the sort of skills that would seem to reflect good thinking is that there are many instances in which presumably intelligent people (high level politicians, physicians, lawyers, and even rocket scientists) do blatantly stupid things, such as allowing psychics to dictate financial decisions, rejecting solid scientific evidence based on a story from a friend, relying on small samples to generalize to a larger group, and confusing correlation with cause, to name a few (Roberts, 1988; Sternberg, 2002).

Stanovich and West (2014) is a staunch advocate of the idea that standardized intelligence tests "miss" important components of what he calls "reasoning" and "rationality." He says to think rationally means taking the appropriate action given one's goals and beliefs (instrumental rationality) and holding beliefs that are commensurate with available evidence (epistemic rationality; Stanovich & West, 2014, p. 81). In one study, Stanovich and West (2008) tested the hypothesis that more intelligent people would be more likely to avoid 'my-side bias' (the preference for conclusions and evidence that supports their world view) and more likely to prefer information that provided a balanced perspective on a controversial issue than less intelligent people. Contrary to expectations, they found that IQ measures did not predict thinking on these two types of reasoning tasks. More intelligent people were not more likely to avoid common cognitive biases or prefer balanced arguments than less intelligent people. Thus, intelligence does not appear to predict whether a person will use good reasoning or exercise good judgement.

The terms "critical thinking" and "wisdom" are labels given for the type of thinking that encompasses the skills that characterize good thinking. Grossman, Na, Varnum, Kitayama, and Nisbett (2013) conceptualize wisdom as a set of reasoning strategies including "pragmatic reasoning" (p. 944). They hypothesized that good reasoning should be associated with greater life satisfaction, greater longevity, and other positive outcomes. They cite many large scale studies that have failed to find a relationship between IQ and well-being and articulate the belief that this failure was due to the fact that "standardized intelligence tests do not do a good job of capturing people's ability to think about social relations or real-world decision-making" (p. 944). To remedy this problem, Grossman et al., devised a test of wise reasoning in which a random sample of adults from the United States was presented with social dilemmas (selected from Dear Abby newspaper columns). To assess the quality of the responses, the authors used six coding categories that represented wise reasoning: a) considering the perspective of others; b) recognizing the likelihood of change; c) considering multiple possible outcomes; d) searching for compromise; e) recognizing uncertainty; and, f) predicting conflict resolution. In addition, they used standardized measures of intelligence and personality. As predicted, they found that wise reasoning predicted well-being, a marginally predicted longevity (assessed with death records for the older participants five years later), and better social relations; whereas, cognitive ability as measured with intelligence tests did not. Although they used very different sorts of experimental stimuli and labels for the construct they were studying, both Grossman and his colleagues and Stanovich and West found that IQ tests do not assess good thinking.

In this paper, we use the term "critical thinking" as our label for good or clear thinking. Critical thinking involves thinking rationally in a goal-oriented fashion (e.g., Ennis, 1993; Halpern, 2003; Moseley et al., 2005; Sternberg, Roediger, & Halpern, 2007). It is a collection of skills and strategies that a thinker can use when the situation calls for them. It is also a disposition towards thinking careful and thoughtfully. For the last several decades, many psychologists have focused on the ways in which many people are consistently biased. For example, Ariely (2008) calls consistent predictable human biases "predictably irrational." Kahneman (2011) catalogues numerous tendencies to think in a seemingly irrational fashion as System 1 (fast) thinking. Slower and more rational thinking is System Lilienfeld, Lynn, Namy, and Woolf (2009) defined critical thinking as thinking intended to overcome cognitive biases. Although there are differences among these authors, all recognize that thinking is often biased, and that with consistent effort, at least some of these biases can be reduced or eliminated.

Sometimes critical thinking can be confused with pessimism and cynicism, but the word "critical" is to be used in a more positive way. To be a *critical* thinker means that you are an amiable skeptic who is willing to invest effort into your thinking process and that you are a flexible thinker. Halpern (2003) defined critical thinking as:

> 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. (p. 6)

The relationship between critical thinking and real-world decision-making has been established in an earlier study. Butler (2012) examined whether scores on a critical thinking assessment (Halpern Critical Thinking Assessment; HCTA) had any "real-world" validity beyond the correlations expected with academic achievement (e.g., grade point averages) and scholastic aptitude measures (e.g., SAT, ACT). The HCTA shows considerable evidence of reliability and validity when assessed with scores on academic

assessments (Halpern, 2012), but what about measures of decisions and behaviors in everyday contexts? For example, do better thinkers have fewer negative life outcomes such as bankruptcies or sexually-transmitted diseases that result from failure to use a condom? De Bruin, Parker, and Fischoff (2007) created an inventory of 41 life outcomes that span a variety of domains such as business, interpersonal, and financial. As predicted, Butler found that higher scores on the critical thinking assessment were associated with fewer negative life events (r = −0.38). However, when she broke the data down into community sample, state university students, and community college students, there were interesting differences, with a strong correlation for the community adult sample (r = −59), a modest correlation for the state university sample (r = −0.29) and a smaller correlation for the community college students (r = −0.23), which was not statistically significant. The difference was due to the nature of the items listed in the De Bruin et al., inventory. They tended to be negative life events associated with adult life—college students, especially younger ones were less likely to have filed for bankruptcy or lost a job because of theft, possibly because they were too young to have purchased a house or held a job for an extended period of time. This assumption was supported when the HCTA was administered along with the same inventory of real-world outcomes to college students in the Netherlands (using a Danish language version of both instruments; De Bie, Pascal, & van der Meij, 2015). The Dutch language HCTA yielded good reliability and two separate factors—one for open-ended responses and one for forced choice responses, but scores on the HCTA were not significantly related to avoidance of negative life events for college students in the Netherlands. As the authors note, the use of college students as participants could have limited the range of negative life events (e.g., been kicked out of an apartment before the lease was up) and college students may be inappropriate for the inventory of negative life events that was used.

Building on the work of Butler (2012) and other researchers, we examined the independent and combined effects of intelligence as assessed with IQ scores and critical thinking as assessed with scores on a popular critical thinking assessment (HCTA). We expected to replicate earlier research showing that IQ can modestly predict some life events, but also expected to find that critical thinking is a better predictor of these events than IQ scores and that it would add significantly to predictive power when used in conjunction with IQ scores.

1. Method

1.1. Participants and procedures

Community adults and college students (n = 244[2]) in the United States took part in this research (32% college students, 68% community adults). A majority of the sample was female (69% female). Participants ranged in age from 19 to 28 years old (M = 21.06, SD = 0.74). The college students were recruited from their Introductory Psychology classes at a large public university in southern California and received research participation credit as compensation. The community adults were recruited from an online participant recruitment system (Amazon's Mechanical Turk) and were paid $12 for their participation. Studies of the comparability of samples recruited via on-line service such as Mechanical Turk have shown that results are similar to nationally representative samples with face-to-face administration on a variety of measures including personality assessment and political opinion polls (Clifford, Jewell, & Waggoner, 2015; DeSoto, 2016).

1.2. Materials and procedures

Participants completed an intelligence assessment, a critical thinking assessment, and an inventory of life events. All of the data were collected online. Participants were told to take breaks as needed between measures, but to complete all the measures within the same day. Participants were excluded from the analysis if they scored more than three standard deviations below the mean on the critical thinking (n = 2) or intelligence assessments (n = 4).

1.2.1. Intelligence assessment

The INSBAT was used to measure intelligence. It is a modular intelligence battery that is decision-oriented. The modules assigned for this study took 90–120 min to complete. The structure of the INSBAT is based on the hierarchical intelligence theory of Cattell-Horn-Carroll (Carroll, 1993; Horn, 1989; Horn & Noll, 1997). It can measure up to 14 facets of intelligence. For the purposes of our research we selected enough subsets to be able to compute general intelligence (g), which included fluid intelligence, crystalized intelligence, visual processing, quantitative reasoning, short-term memory, and long-term memory. Higher scores on the INSBAT indicate greater general intelligence (M = 100.26, SD = 11.92). The reliability of the INSBAT has been tested. Because test administrators are allowed to choose which subtests to administer, the reliability of the INSBAT varies from r = 0.70 to r = 0.95. The INSBAT is administrated and scored by the *Vienna Test System,* a proprietary program that is used in the administration and scoring of a wide range of psychometric assessments. The psychometric properties of the INSBAT including norms, reliability and validity can be found in Arendasy et al. (2012).

1.2.2. Critical thinking assessment

The Halpern Critical Thinking Assessment (HCTA) was used to measure critical thinking ability. The assessment takes 45–65 min

[2] The initial sample size was 329, but 85 participants were excluded either because they did not complete all three assessments or they scored three standard deviations below the mean on one or more of the assessments.

to complete. The HCTA measures five facets of critical thinking including: (a) verbal reasoning, (b) argument analysis, (c) hypothesis testing, (d) likelihood and uncertainty, and (e) problem solving. Participants were presented with 20 everyday scenarios (4 per facet of critical thinking) to analyze and critique. The assessment is unique in that it measures these facets through both constructed responses (recall memory) and forced-choice responses (recognition memory). Because of test confidentiality, we provide an example that is similar, but not identical, to a scenario. Respondents were told that a newspaper reported that children who attended preschool were better readers in first grade. Based on this information, respondents were asked if preschool should be made mandatory for all children. They were to explain their reasoning in one or two sentences. The ideal response would indicate that these are correlational data and although it is possible that preschool attendance increased reading ability, it is also likely that families who could afford preschool provided other academic advantages for their children. After writing a one or two sentence response, respondents were shown a series of related facts and asked to rate the extent to which each fact would help them make an informed decision. For example, suppose we knew that the children who attended to preschool were, on average, twice as likely to come from a family earning over $75,000 a year than the children who did not attend preschool. Hypothetical vignettes such as those used in the HCTA have been found to mimic real world behavior. Hainmueller, Hangartner, & Yamamoto (2014) reported that similar vignettes can be used to "elucidate how humans make multidimensional choices and evaluate objects" (p. 2395).

The HCTA was administrated and scored by computer as part of the *Vienna Test System*. The grading system automatically calculates the quantitative (forced-choice) responses, and then guides the grader through the grading of the qualitative (constructed) responses with computerized grading prompts. For example, based on the example given previously, the computerized grading systems would present the grader with the participant's response and ask the grader whether the respondent's answer indicated any of the following: that the children in preschool may differ from those who did not attend preschool in systematic ways; policy makers would need random assignment to school conditions to determine cause, or there are many ways the two groups of children differ. The human grader indicates whether the respondent's answer clearly indicated this, less clearly indicated this, or did not indicate this at all. Respondents receive a greater number of points for accurate and comprehensive responses. Scores on the HCTA can range from 0 to 159 with higher scores indicating greater critical thinking. The HCTA has demonstrated strong reliability and validity (see the HCTA manual for relevant psychometric research; Halpern, 2012).

1.2.3. Real-world outcomes

The real-world outcomes (RWO) inventory was used to measure the proportion of negative life events experienced by the participants. The inventory takes 10–15 min to complete. The RWO was adapted from the Decision Outcomes Inventory (DOI) that was created by De Bruin et al. (2007). It was originally used to assess the validity of the Adult Decision-Making Competence (A-DMC) index. The self-report inventory measures negative life events from many domains (e.g., interpersonal, work, financial, health, education). These life events vary in severity from mildly negative (e.g., paying late fees for a movie rental) to severely negative (e.g., contracting a sexually transmitted disease because you did not wear a condom). The RWO is scored with the use of decision sets. The first item in the set asks participants whether they made a decision that would have made the life event possible (e.g., *I did my own laundry*). The second item in the set asks participants whether they experienced a negative life event (e.g., *I ruined a load of laundry*). The item sets help to control for differences in the proportion of life decisions the participants make for themselves. For instance, someone who makes very few decision for themselves (e.g., *students whose parents do their laundry*) would have an advantage, over those who make most life decisions for themselves, if only a basic checklist of negative life events was used. In addition to the item sets, there were some individual items for which no preceding question was necessary (e.g., *been in jail overnight for any reason*). New items were added to the original RWO, published by Butler (2012) that captured life events would be relevant to younger adults (e.g., social media posting, online dating, etc.). There were a total of 134 questions in the final version of the RWO, 83 item sets[3] and 8 individual items (see Table 1 for the revised scale).

The total RWO score is the proportion of negative life events experienced by the participant (the total number of negative life events divided by the total number of possible life decisions made by the participant). Scores range from 0 to 1 with higher scores indicating more negative life events or poorer real-world outcomes.

2. Results

The purpose of this research was to explore the relative contribution of critical thinking ability and intelligence to the prediction of real-world outcomes. The pattern of results was similar for college students and community adults, so the results were combined. Separate, zero-order, Pearson correlations were computed. Scores on the INSBAT and HCTA were moderately correlated ($r = 0.38$) indicating that although there is some overlap, they are distinct concepts. The relationship between scores on the critical thinking assessment and scores on the RWO inventory was moderate, $r = -0.330$. Those with higher critical thinking scores reported fewer negative life events. A similar relationship was found for intelligence test scores, $r = -0.264$. Those with higher intelligence scores reported fewer negative life events. Thus, both critical thinking and intelligence predicted negative real life events.

It was hypothesized that critical thinking ability would be a stronger predictor of real-world outcomes than intelligence. To test

[3] Several of the item sets involved the same preceding decision that would have made the negative life event possible. For instance, one decision that would have preceded a negative life event (e.g., enrolling in a college course) might have numerous negative life events associated with it (e.g., missed a class because you forgot to set your alarm, arrived to class and discovered that there was an exam that day that you had forgotten about). While this was taken into account when the score was computed, these preceding events were not repeatedly asked and were not included in the overall question count.

Table 1
Percentage of Real-World Outcomes Experienced.

Item	Percent who made decision	Percent who experienced outcome (given decision)
Rented a movie	78	
….Returned a movie you rented without having watched it at all		56
….Had to pay late fee because you returned it too late		49
Bought new clothes or shoes	96	
….Bought new clothes or shoes you never wore		51
….Bought new clothes or shoes you never wore when you were not in a good financial shape		34
Been enrolled in any kind of school	73	
….Missed a class because forgot to set your alarm		45
….Waited until the night before an exam to start studying		89
….Forgot to do a class assignment		75
….Arrived to class only to realized that you'd forgotten about an exam that day		33
….Been suspended from school for at least one day for any reason		10
….Arrived on the wrong day for an exam		10
….Overloaded with more classes than you could manage		49
….Been on academic probation		16
….Cheated on an exam		18
….Skipped an important class to do something fun (e.g., go to a concert, see friends, shop, etc.)		55
….When out with friends, instead of study for an exam		78
….Wrote a paper for a class that was more than 4 pages long		97
….Started writing the paper the night before it was due		74
Had any kind of job	77	
….Quit a job after a week		10
….Consistently went to work late		14
Had a driver's license	88	
….Had your driver's license taken away from you by the police		6
Driven a car	92	
….Caused a car accident		16
….Gotten a parking ticket		35
….Gotten a speeding ticket		13
….Locked your keys in the car		20
….Hit something with your car		39
….Ran a stop sign or traffic light		51
….Texted while driving		59
Taken a trip by airplane	50	
….Missed a flight		14
Taken the train or the bus	53	
….Taken the wrong train or bus		25
….Missed the train or bus		49
Had any form of ID (driver's license, passport, birth certificate)	97	
….Had your ID replaced because you lost it		15
Lived in a rented apartment or other rental property	48	
….Been kicked out of an apartment or rental property before the lease ran out		10
Carried a key to your home	92	
….Had the key to your home replaced because you lost it		16
….Locked yourself out of your home		37
Been responsible for electricity, cable, gas or water payments	50	
….Had your electricity, cable, gas or water shut off because you didn't pay on time		14
Been responsible for rent or mortgage payments	43	
….Paid a rent or mortgage payment at least 2 weeks too late		23
Purchased lottery tickets	46	
Used checks	60	
….Bounced a check		15
Had a credit card	56	
….Had more than $1000 in credit card debt		41
….Were charged a late fee because I was late or missed a payment		36
Had a credit or debit card	92	
….Had that card rejected for insufficient funds while trying to make a purchase		30
Been to a bar, restaurant, or hotel	94	
….Been kicked out of a bar, restaurant, or hotel by someone who works there		3
Had a romantic relationship that lasted for at least 6 months	66	
….Cheated on your romantic partner		18
Been married	14	
….Been divorced		21
Had sex	66	
….Had a one night stand		33
….Had unprotected sex		68
….been diagnosed with an STD		3

(continued on next page)

Table 1 *(continued)*

Item	Percent who made decision	Percent who experienced outcome (given decision)
….Had an unplanned pregnancy (or got someone pregnant, unplanned)		12
Had an alcoholic drink	78	
….Consumed so much alcohol you vomited		42
….Consumed so much alcohol you do not remember parts of the night (i.e., “blacked out”)		33
….Received a DUI for drunk driving		3
Been out in the sun for more than an hour	93	
….Forgot to wear sunscreen		86
Read your horoscope	65	
….Found that your horoscope was accurate		69
….Made a decision based on your horoscope		18
Been in a public fight or screaming argument	25	
Forgotten a birthday of someone close to you and did not realize until the next day or later.	52	
Broke a bone because you fell, slipped, or misstepped	8	
Been in a jail cell overnight for any reason	3	
Own a lucky object? (e.g., rabbit foot, etc)	18	
Purchased Toning Shoes	4	
Purchased herbal remedies that enhance thinking or memory.	9	
Owned an object with healing properties (e.g., healing crystals, magnetic bracelets, mystical stones, etc)	7	
Sent a text message	97	
….Sent text to the wrong person		63
….Constantly texting and not paying attention to surroundings		51
Watched television	98	
….Spent so much time watching television that it negatively affected work/school/relationships		35
Eat at a dining hall or buffet	82	
….Ate too much or ate unhealthy food too often		83
Skyped with family	36	
….Family member saw something you wish they had not seen during a Skype call		6
Lived with roommate(s)	42	
….Argued with roommate(s) over who should do household chores		57
Got new piercing(s) or tattoo(s)	32	
….Piercing(s) got infected		41
Started a new diet	50	
….Started a new diet without researching its effectiveness or consulting a doctor		66
….Selected a new diet plan based on my friend’s advice		44
Met someone new online (e.g., in a chatroom, Facebook, etc.)	42	
….Shared personal information with someone new that I met online (e.g., address, phone number, where you work or go to school, etc.)		71
….Proposed or accepted a date from someone new that I met online.		38
Went on a first date	40	
….Met my date somewhere I was unfamiliar with		35
….Met my date in an area with few other people		50
Spent time with a new group of people	72	
….Did something dangerous to impress the new group of people		19
Was at least 18 years old at the time of the last presidential election	90	
….Did not vote		40
Adopted a pet	25	
….Gave this pet away or abandon it because I couldn’t take care of it		20
Gambled for money	30	
….Lost more than $100 gambling		30
….Borrowed money to gamble		10
Smoked cigarettes	24	
Took out a student loan	31	
….Owe more than $13,000 in student loans		52
Thought about getting a flu shot	49	
….Decided not to get a flu shot because it will make me sick		57
Posted to a social networking website (e.g., Facebook, Twitter)	87	
….Regretted something I posted		44
….Got in trouble for something I posted (e.g., got fired from a job, caused a fight, hurt someone)		14
….Posted during class		48
….Posted during work		45
Sent an email	96	
….Sent an email with spelling errors to someone important		27
….Sent an email to the wrong person		16
Had a mobile/cellular phone	95	
….Lost a mobile/cellular phone		20
….Had to pay extra because you went over your call/text/data limit		22
Received chain letter or email	63	
….Forwarded chain letter or email		22

this hypothesis a hierarchical multiple regression was conducted with scores on the RWO inventory as the dependent variable. Scores on the intelligence test (INSBAT) were entered into the model in the first step and scores on the critical thinking assessment (HCTA) were entered into the second step of the model. The first step in the model was statistically significant, $R = 0.264$, $F(1, 243) = 18.318$, $p = 0.001$. Those who scored higher on the intelligence test reported experiencing fewer negative life events. Critical thinking scores were entered into the second step of the hierarchical regression and this step of the model was also statistically significant, $R = 0.403$, $F(2, 242) = 23.624$, $p = 0.001$. Both predictors were statistically significant. Those with higher scores on the critical thinking assessment reported fewer negative life events ($\beta = -0.323$, $t = -5.194$, $p = 0.001$), and those with higher scores on the intelligence test reported fewer negative life events ($\beta = -0.158$, $t = -2.534$, $p = 0.012$). Both predictors predicted the occurrence of negative life events, although critical thinking ability was a stronger predictor than intelligence. Furthermore, critical thinking uniquely contributed to the prediction (R^2 change = 0.093, F change = 26.980, $p = 0.001$). See Table 2.

3. Discussion

Over one hundred years of research on intelligence testing has shown that scores on standardized tests of intelligence predict a wide range of outcomes, but even the strongest advocates of intelligence testing agree that IQ scores (and their near cousins such as the SATs) leave a large portion of the variance unexplained when predicting real-life behaviors (Hunt, 2012). We began this study by (a) asking if critical thinking ability would be a better predictor of real-life behaviors than IQ scores, and (b) whether the combination of these two measures would be a better predictor than either one alone. Our results support both of these hypotheses—critical thinking ability had a greater association with real life decisions, and it added significantly to explained variance, beyond what was accounted for by intelligence alone. These results leave us cautiously optimistic about the assessment of critical thinking ability and the validity of critical thinking scores to predict what people actually do (or say they do) in real life.

The biggest problem in determining if any measure has "real-world" validity is the operationalization of "real-world." In this study, we used a behavioral inventory that was originally developed by De Bruin, Parker, and Fischhoff (2007). Scores on this inventory show the proportion of decisions that had negative outcomes (e.g., failure to pay a bill on time). Although this inventory accounts for life events that vary based on domain (e.g., medical, financial, interpersonal) and severity of the outcome, the inventory only measured a limited number of human experiences. We encourage other researchers to contribute to this list of life events, so that the scale can more adequately capture the full range of life outcomes.

As noted in previous studies, another weakness of the inventory is that it was designed for community adults who would have had ample time and opportunity to make a range of positive and negative decisions. Participants in this study were primarily adults, but overall the sample was weighted toward young adults. In prior research, the relationship between critical thinking and real-world outcomes was not as strong for young adults as it was for older adults. Life outcomes were added to the original inventory with the intention of capturing more experiences of young adults, but it is likely that the additions were not sufficient. Thus, the predictive ability of critical thinking might be even stronger with a list of possible outcomes that are more closely aligned with the decisions made by young adults. Additionally, the findings are limited by the measures used, in that, each construct was assessed with a single measure. It is possible that the results might differ if another critical thinking assessment, intelligence test, or behavioral inventory was used.

Although the methodology of this study (a correlational design) does not allow us to make casual statements (e.g., improving our critical thinking skills will prevent us from making bad decisions) or rule out alternative explanations of these results (e.g., critical thinkers are more selective about reporting certain life outcomes), we are optimistic that improving critical thinking and intelligence will have a positive impact on our everyday lives. Jones (2016) makes a persuasive case for the benefits of raising the IQ of all people. He envisions a world in which all people have the opportunity to become more intelligent. Such a world would have less corruption, with citizens that are more patient, cooperative, and better-informed, with concomitant economic benefits. He supports the idea that

Table 2
Scores on the Critical Thinking Assessment (HCTA) and Intelligence assessment (INSBAT) predicted Real-world Outcomes (RWO).

	Real-world outcomes of critical thinking			
		Model 2		
Variable	Model 1 β	β		95% CI
(Constant)				[0.797, 1.155]
IQ	.264***	−0.158*		[−0.004, 0.000]
HCTA		−0.323***		[−0.005, −0.002]
R^2	.066***		.156***	
F	18.318***		23.624***	
ΔR^2			0.093	
ΔF			26.98	

Note. $N = 244$. CI = confidence interval. HCTA = Halpern Critical Thinking Assessment score. IQ = INSBAT intelligence score.
* $p < 0.05$.
** $p < 0.01$.
*** $p < 0.001$.

intelligence can be enhanced with reference to the Flynn Effect, which has shown that IQ scores have increased an average of approximately one standard deviation (15 points) per generation since 1930 (Flynn, 2007). In a similar vein, Rindermann, Kodila-Tedika, and Christainsen (2015) make the case that countries with higher levels of intelligence have good governance. Other researchers have questioned whether the rise in IQ scores is attributable to increasing levels of intelligence or whether it is due to better test-taking skills and the increase in number of years of education across generations (e.g., Deary, 2001). Regardless of the cause for the rise in IQ scores, researchers agree that positive outcomes are associated with it.

There is considerable agreement that critical thinking skills can be enhanced, especially with specific instructional strategies designed for that purpose. The large research literature on this topic has been reviewed by Halpern (2012) and in a more recent meta-analysis by Huber and Kuncel (2016). The meta-analysis was conducted in response to a popular book that was very critical of higher education. Arum and Roksa (2010), the authors of *Academically Adrift: Limited Learning on College Campuses*, concluded that over one-third of college students showed no improvement in critical thinking skills during their time in college. Arum and Roksa's study has been criticized for being pessimistic and for the way critical thinking was assessed. More specifically, the lack of incentives students were given to demonstrate their critical thinking skills. It is difficult to measure critical thinking and to get students to take an assessment that has no impact on their grade seriously. The results of Huber and Kunkel's meta-analysis indicated that both critical thinking skills and the disposition to use those skills improves substantially over a normal college experience. This is consistent with a growing body of literature from all parts of the world and at all grade-levels that indicates critical thinking skills can be enhanced. Interestingly, no specific program or courses in the meta-analysis were identified as having increased critical thinking skills. This is perhaps not surprising given that good thinkers need to employ an inventory of critical thinking skills that cut across disciplines, and also need disciplinary knowledge to use them effectively.

We are making a strong plea for more instruction in and attention to critical thinking skills. We can imagine a world where many more people think critically. Around the world, people are called upon to vote on a wide range of critical issues. The irrational (uncritical) voter is a threat to all of us, as are irrational politicians, business executives, and scientists. We believe that we can create a better future by enhancing critical thinking skills of citizens around the world. This optimism is tempered with the reality that so far, we only have data showing that individuals make fewer negative decisions in their personal lives when they are better thinkers, and can only imagine the impact of a world-wide increase in better thinking. We have nothing to lose by trying.

References

Arendasy, M., Hornke, L. F., Sommer, M., Wagner-Menghin, M., Gittler, G., Häusler, J., ... Wenzl, M. (2012). *Intelligenz-Struktur-Batterie (Intelligence structure battery) – INSBAT. modling*. Austria: Schuhfried Publishers.

Ariely, D. (2008). *Predictably irrational: The hidden forces that shape our decisions*. New York, NY: Harper Collins Publishers.

Arum, R., & Roksa, J. (2010). *Academically adrift*. Chicago, IL: The University of Chicago Press.

Butler, H. A. (2012). Halpern critical thinking assessment predicts real-World outcomes of critical thinking. *Applied Cognitive Psychology*, *26*, 721–729.

Carroll, J. B. (1993). Educational psychology in the 21st century. *Educational Psychologist*, *28*, 89–95. http://dx.doi.org/10.1207/s15326985ep2802_1.

Clifford, S., Jewell, R. M., & Waggoner, P. D. (2015). Are samples drawn from Mechanical Turk valid for research on political ideology? *Research and Politics*, *1–9*. http://dx.doi.org/10.1177/2053168015622072.

De Bie, H., Wilhelm, P., & Van der Meij, H. (2015). The halpern critical thinking assessment: Toward a Dutch appraisal of critical thinking. *Thinking Skills and Creativity*, *17*, 33–44.

De Bruin, W. B., Parker, B., & Parker, A. M. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology*, *92*, 938–956.

DeSoto, A. (2016). *Under the hood of mechanical turk*. Observer. Retrieved from http://www.psychologicalscience.org/index.php/publications/observer/2016/march-16/under-the-hood-of-mechanical-turk.html.

Deary, I. J. (2001). *Intelligence: A very short introduction*. New York, NY: Oxford University Press.

Ennis, R. H. (1993). Critical thinking assessment. *Theory into Practice*, *32*(3), 179–186.

Flynn, J. R. (2007). *What is intelligence?* Cambridge, U.K: Cambridge University Press.

Furnham, A., & Monsen, J. (2009). Personality traits and intelligence predicts academic school grades. *Learning and Individual Differences*, *19*, 28–33.

Gordon, R. A. (1986). Scientific justification and the race-IQ-delinquency model. In T. F. Hartnagel, & R. A. Silverman (Eds.), *Critique and explanation* (pp. 91–131). New Brunswick: Transaction Books.

Grossman, I., Varnum, M. E. W., Na, J., Kitayama, S., & Nisbett, R. E. (2013). A route to well-being: Intelligence versus wise reasoning. *Journal of Experimental Psychology: General*, *142*, 944–953.

Hainmueller, J., Hangartner, D., & Yamamoto, T. (2014). Validating vignette and conjoint survey experiments against real-world behavior. *Pnas*, *112*, 2395–2400.

Halpern, D. F. (2003). *Thought and knowledge: An introduction to critical thinking* (4th ed.). Mahwah, NJ: Erlbaum.

Halpern, D. F. (2012). *Halpern critical thinking assessment (HCTA) scoring guide*. Modling. Austria: Schuhfried Publishers.

Horn, J. L., & Noll, J. (1997). Human cognitive capabilities: Gf-Gc theory. In D. P. Flanagan, J. L. Genshaft, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (pp. 53–91). New York: Guilford.

Horn, L. R. (1989). *A natural history of negation*. Chicago, IL: Chicago University Press.

Huber, C. R., & Kuncel, N. R. (2016). Does college teach critical thinking? A meta-analysis. *Review of Educational Research*, *86*, 431–468. http://dx.doi.org/10.3102/0034654315605917.

Hunt, E. (2012). What makes nations intelligent? *Perspectives on Psychological Science*, *7*, 284–306.

Hunter, J. E., & Schmidt, F. L. (1996). Intelligence and job performance: Economic and social implications. *Psychology, Public Policy, and Law*, *2*, 447–472.

Jones, G. (2016). *Hive Mind: How your nation's IQ matters so much more than your own*. Redwood City, CA: Stanford University Press.

Kahneman, D. (2011). *Thinking, fast and slow*. New York, NY: Farrar, Straus and Giroux.

Kuncel, N. R., Hezlett, S. A., & Ones, D. S. (2004). Academic performance, career potential, creativity, and job performance: Can one construct predict them all? *Journal of Personality and Social Psychology*, *86*, 148–161.

Kuncel, N. R., Ones, D. S., & Sackett, P. R. (2010). Individual differences as predictors of work, educational, and broad life outcomes. *Personality and Individual Differences*, *49*, 331–336.

Lilienfeld, S. O., Lynn, S. J., Namy, L., & Woolf, N. (2009). *Psychology: From inquiry to understanding*. Boston, MA: Allyn & Bacon.

Lubinski, D., Benbow, C. P., Webb, R. M., & Bleske-Rechek, A. (2006). Tracking exceptional human capital over two decades. *Psychological Science*, *17*, 194–199. http://dx.doi.org/10.1111/j.1467-9280.2006.01685.x.

Mayer, J. D., Caruso, D. R., Panter, A. T., & Salovey, P. (2012). The growing significance of hot intelligences. *American Psychologist*, *67*(6), 502–503. http://dx.doi.org/10.1037/a0029456.

Moseley, D., Baumfield, V., Elliott, J., Gregson, M., Higgins, S., Miller, J., & Newton, D. P. (2005). *Frameworks for thinking: A handbook for teaching and learning*. Cambridge, MA: Cambridge University Press.

Nisbett, R. E., Aronson, J., Blair, C., Dickens, W., Flynn, J., Halpern, D. F., & Turkheimer, E. (2012). Intelligence: New findings and theoretical developments. *American Psychologist*, *67*, 130–159. http://dx.doi.org/10.1037/a0026699.

Rindermann, H., Kodila-Tedika, O., & Christainsen, G. (2015). Cognitive capital, good governance and the wealth of nations. *Intelligence*, *51*, 98–108.

Roberts, S. V. (1988). *White house confirms Reagans follow astrology, up to a point*. New York Times. Retrieved from http://www.nytimes.com/1988/05/04/us/white-house-confirms-reagans-follow-astrology-up-to-a-point.html.

Schwartz, J. A., Savolainen, J., Aaltonen, M., Merikukka, M., Paananen, R., & Gissler, M. (2015). Intelligence and criminal behavior in a total birth cohort: An examination of functional form, dimensions of intelligence, and the nature of offending. *Intelligence*, *51*, 109–118.

Soars, D. L., Lemos, G. C., Primi, R., & Almeida, L. S. (2015). *Learning and Individual Differences*, *41*, 73–78.

Stanovich, K. E., & West, R. F. (2008). On the failure of cognitive ability to predict myside and one-sided thinking biases. *Thinking & Reasoning*, *14*, 129–167. http://dx.doi.org/10.1080/13546780701679764.

Stanovich, K. E., & West, R. F. (2014). What intelligence tests miss. *The Psychologist*, *27*(2), 80–83.

Sternberg, R. J., Roediger, H. L., & Halpern, D. F. (2007). *Critical thinking in psychology*. New York, NY: Cambridge University Press.

Sternberg, R. J. (2002). *Why smart people can be so stupid*. New Haven, CT: Yale University Press.

Vera, H., Feagin, J. R., & Gordon, A. (1995). Superior intellect?: Sincere fictions of the white self. *The Journal of Negro Education*, *64*, 295–306.

Wechsler, D. (1939). *The measurement of adult intelligence*. Baltimore, MD: Williams & Wilkins.