The Halpern Critical Thinking Assessment and real-world outcomes: Cross-national applications☆

Heather A. Butler [a],[*], Christopher P. Dwyer [b], Michael J. Hogan [b], Amanda Franco [c], Silvia F. Rivas [d], Carlos Saiz [d], Leandro S. Almeida [c]

[a] School of Behavioral and Organizational Sciences, Claremont Graduate University, United States
[b] National University of Ireland, Galway, Ireland
[c] University of Minho, Portugal
[d] University of Salamanca, Spain

ARTICLE INFO

ABSTRACT

The Halpern Critical Thinking Assessment (HCTA) is a reliable measure of critical thinking that has been validated with numerous qualitatively different samples and measures of academic success (Halpern, 2010a). This paper presents several cross-national applications of the assessment, and recent work to expand the validation of the HCTA with real-world outcomes of critical thinking (e.g., contracting a sexual transmitted infection because you did not wear a condom). The real-world outcomes (RWO) inventory measures behavior in a wide range of domains, such as education, health, finance, and interpersonal relationships. Study 1 examined whether scores on the HCTA predicted real-world outcomes in three qualitatively different samples in the United States. Study 2 used the HCTA to assess the effectiveness of an online critical thinking course, and whether the HCTA predicted real-world outcomes in Ireland. Study 3 describes preliminary research involving the translation quality of the RWO (into Spanish and Portuguese) and explored differences in behaviors in the two countries. Overall, the HCTA is a useful tool for assessing critical thinking and predicts real-world outcomes of critical thinking. Cross-national implications and applications are discussed.

© 2012 Elsevier Ltd. All rights reserved.

1. Introduction

Over the last several decades educators, employers, and organizations around the world have expressed concern about student preparedness for a 21st century world (e.g., Association of American Colleges and Universities [AAC&U], 2010; Bureau of Labor Statistics, 2011; Galagan, 2010; Halpern, 2010b; Hunt, 1995). In response to these concerns worldwide educational reform in terms of the identification and assessment of student learning outcomes, has been incorporated into cross-national standards (European Higher Education Area, 2011). In the United States, the Commission of the Future of Higher Education (also known as the Spellings Commission; U.S. Department of Education, 2006) has placed increased emphasis on assessing learning outcomes, especially in critical thinking. Educational reform efforts have also been seen in the African Union, the

Euro-Mediterranean High Education and Scientific Research area, and the Latin America and Caribbean Higher Education Area (Redden, 2010). There is a large body of empirical evidence showing that critical thinking skills can be taught (for reviews, see Abrami et al., 2008; Chance, 1986; Halpern, 2003; Moseley et al., 2005; Nisbett, 1992), and that explicit critical thinking instruction is more beneficial than implicit instruction (Marin & Halpern, 2011). As more countries extol the benefits of critical thinking instruction, the need for an assessment of this key learning outcome becomes more critical, as does the need for a cross-national assessment.

Critical thinking has been defined many ways (e.g., Ennis, 2005; Halpern, 2003; Moseley et al., 2005; Sternberg, Roediger, & Halpern, 2007), but experts would generally agree that critical thinking involves an attempt to achieve a desired outcome by thinking rationally and in a goal-oriented fashion. Halpern (2003, 2010b) adds that while critical thinking involves a learned set of skills or cognitive strategies (e.g., problem solving strategies, calculating and correctly applying probabilities), critical thinking is also a disposition towards engaging in the thinking process. Recently, Stanovich (2009) argued that critical thinking is what intelligence tests fail to adequately measure. This idea echoes the general consensus among researchers that intelligence and critical thinking are separate constructs and was empirically tested in a series of studies that explored the relationship. Stanovich and West (2008) used SAT-scores as an estimate of cognitive ability and numerous well-known thinking biases (e.g., denominator neglect, conjunction effect, framing effects, anchoring effects, base-rate neglect, "less is more" effects, affect bias, omission bias, myside bias, sunk-cost effects, and certainty effects) as an estimate of critical thinking ability. Whereas some critical thinking biases were moderately related to cognitive ability, most critical thinking biases were not related to cognitive ability. Thus, critical thinking and intelligence are separable constructs, but share at least one common attribute – they are difficult to adequately assess.

There are several assessments that measure critical thinking (interested readers are referred to Association of American College and Universities website for a short list of assessments; AAC&U, n.d.). Some of the many issues involved with the adequate assessment of critical thinking are the cross-cultural equivalences of these assessments and whether they can demonstrate acceptable psychometric properties. Furthermore, cognitive psychologists make a distinction between assessments that require recognition memory (e.g., multiple-choice, ranking) and recall memory (e.g., short-answer, essay). Recognition-based assessments are easier to score than recall-based items, but they are more susceptible to guessing. Many critical thinking assessments rely exclusively on recognition (e.g., the Measures of Academic Proficiency and Progress, the California Thinking Skills Test, the Cornell Critical Thinking Test), while others rely exclusively on recall (e.g., the ICAT Critical Thinking Essay Examination, the Cornell (also known as the Ennis-Weir) Critical Thinking Essay, and the Watson-Glaser II Critical Thinking Appraisal). The Halpern Critical Thinking Assessment is the only assessment that involves both recognition and recall.

1.1. The Halpern Critical Thinking Assessment (HCTA)

In the face of a need for adequate instruments to assess critical thinking, the potential raised by the HCTA's format – which considers both behavioral and motivational components of critical thinking, combines open and multiple-choice questions, and appeals to daily easy-to-relate-to situations (Ku, 2009) – gains even more relevance. The HCTA is a standardized instrument that consists of 25 everyday scenarios that respondents analyze and critique. Respondents first answer open-ended questions (recall-based) related to the scenario and then respond to forced-choice questions (recognition-based) related to the scenario. The scenarios involve thinking in a variety of domains including health, education, work, social policy, and others. For example, in one scenario respondents are told that a *newspaper reported several crimes committed by repeat-offenders who were released from prison early. One man who was interviewed demanded that the parole board members be fired.* First, respondents are asked to generate two questions that they would like to ask before making a decision about firing the parole board members. Second, respondents are shown a series of questions and rate the extent to which each question would help them make an informed decision. Five scenarios are presented for each subcategory of critical thinking, including: (a) verbal reasoning skills, (b) argument analysis skills, (c) skills in thinking as hypothesis testing, (d) using likelihood and uncertainty, and (e) decision making and problem solving skills.

Numerous studies have established the reliability and validity of the HCTA (for review see Halpern, 2010a). The validity of the HCTA has been evaluated using a variety of methodologies (e.g., correlational, pretest–posttest experimental designs), with respondents who vary widely in education level (e.g., high school students, community college students, state university students, private liberal arts students, graduate students, community adults) and nationality (e.g., China, Ireland, Portugal, Spain, United States, Viet Nam, etc.). The assessment has high face validity and can be easily communicated to a general audience because the 25 scenarios are familiar events (e.g., reading a newspaper article about crime, a friend talking about a new diet program). Numerous studies have evaluated the content validity evidence for the HCTA. These studies provide support for the five factor structure, a general critical thinking factor, and the separability of the recognition and recall facets of the assessment (Halpern, 2010a; Ku et al., 2006). Evidence of criterion validity has been established (Halpern, 2010a; Ku et al., 2006) by comparing scores on the HCTA to scores on standardized exams (e.g., SAT-Verbal $r = .58$, SAT-Math $r = .50$, GRE-Verbal $r = .12$, GRE-Quantitative $r = .20$), tests of reasoning (e.g., Arlin Test of Formal Reasoning $r = .32$), class grades (ranging from $r = .17–.41$), and scores on need for cognition scales ($r = .34$). Consistent with other assessments of critical thinking, most of the validity evidence for the HCTA is based on academic achievement (e.g., grades, standardized test scores) or measures of cognitive abilities (e.g., fluid intelligence, tests of reasoning).

1.2. Real-world outcomes of critical thinking

Surely, critical thinking skills need to predict more than course-related grades and cognitive test scores. There are many everyday situations that should be influenced by critical thinking. We might expect critical thinkers to make more informed decisions, so that they avoid certain negative life events that are indicative of poor decision-making. An inventory of negative life events, called the Decision Outcomes Inventory (DOI), was created by de Bruin, Parker, and Fischhoff (2007). It was originally used to assess the validity of the Adult Decision-Making Competence (A-DMC) index. The self-report inventory measures 34 life outcomes from many domains (e.g., interpersonal, business, financial, interpersonal). These life outcomes vary in severity from mildly negative (e.g., paying late fees for a movie rental) to severely negative (e.g., foreclosure on a home). The inventory consists of 28 item sets and six individual items. The item sets consist of a negative life event (e.g., *I threw out food because it went bad*) and a decision that precedes the negative life event that would have made the event possible (e.g., *I bought groceries for myself*). In addition to the item sets, there were six individual items for which no preceding question was necessary (e.g., *been in jail overnight for any reason*). See Table 1 for a list of the items. One benefit of such an inventory is that many life events (e.g., *throwing out groceries because they went bad before you could eat them*) are common experiences that are shared by people from many countries. Thus, this behavioral inventory lends itself to cross-national comparisons.

1.3. Cross-national applications

The HCTA has been used to assess teaching and learning in many countries including China, Ireland, Portugal, the Netherlands, Spain, United States, and Viet Nam. Hau et al. (2006) conducted a cross-national comparison of the HCTA scores of students from China and the Unites States of America. More specifically, the study sought to establish the content validity of the assessment in each culture, which it did. The study found support for the five factor structure, and the separability of the forced-choice and open-ended response options. Ku et al. (2006) examined cross-national differences in the validity evidence with Chinese and American samples. Students grade point average (the mean of students' college grades; GPA), gender, and major were collected from both the Chinese and American students. For the Chinese students, GPA did not correlate with any of the HCTA subscales. For the students from the United States, the correlations between the HCTA and GPA were low or inconsistent. Thus, critical thinking scores were not related to GPA for these samples possibly because grades in college are not reflective of critical thinking ability. Scholastic Aptitude Test (SAT) scores were available only for the United States sample. Interestingly, SAT scores did correlate with HCTA. Halpern (2010a) suggests that this supports the general belief that GPA reflects diligence and persistence (e.g., turning in homework on time and rote memorization of course content), while SAT reflects aptitude.

Cross-national comparisons of the RWO inventory are especially critical. The inventory measures everyday behaviors in a variety of domains, which may have different linguistic and cultural significances. Most of the behaviors measured in the inventory would be considered indicative of poor decision-making regardless of culture (e.g., spending the night in jail), although the nuances of certain behaviors may need modification in order to be sensitive to cultural differences. For example, different cultures have different norms for what would be considered excessive credit card debt. Thus, a cultural and linguistic validation of this instrument would be necessary before it was used. The process of cross-national validation may involve back-translations to ensure linguistic equivalence, as well as item modification to ensure cultural validation.

The purpose of the present studies was to improve the assessment of critical thinking, expand the validity of the HCTA, and contribute to our understanding of cross-national differences in critical thinking assessment. Study 1 was conducted in the United States. The purpose of study one was to determine whether scores on the HCTA could predict real-world outcomes of critical thinking. It was predicted that those who scored higher on the HCTA would report fewer negative life events than those who scored lower on the assessment. Study 2 was conducted in Ireland. The purpose of study two was to assess the effectiveness of an online critical thinking course, and replicate the findings from Study 1 in Ireland. It was predicted that HCTA scores would be higher at posttest than pretest, and that the HCTA would predict real-world outcomes of critical thinking. Study 3 was conducted in Portugal and Spain. The purpose of study three was to evaluate the translation and psychometric properties of the RWO, and to determine whether there were any differences in the interpretation of the inventory or differences in the reported behaviors of students from Spain and Portugal. This study was a preliminary step that was necessary to ensure that this inventory could be used in future research that aims to explore the predictive ability of the HCTA in these countries.

2. Study 1: United States

The main purpose of study one was to explore whether scores on the HCTA would predict real-world outcomes of critical thinking. It was predicted that those who scored higher on the HCTA would report fewer negative life events than those who score lower on the assessment. This study also explored differences in HCTA scores in three different samples (non-selective community college students, more-selective state-university students, and community adults) and whether the relationship between HCTA scores and real-world outcomes differed based on these samples. Portions of this study are published elsewhere (Butler, in press).

Table 1
Real World Outcomes inventory.

Item	
1a)	Rented a movie
b)	Returned a movie you rented without having watched it at all
c)	Had to pay late fee because you returned it too late*
2a)	Bought new clothes or shoes
b)	Bought new clothes or shoes you never wore
3a)	Gone shopping for food or groceries
b)	Threw out food or groceries you had bought because they went bad
4a)	Done your own laundry
b)	Ruined your clothes because you did not follow the laundry instructions on the label
5a)	Been enrolled in any kind of school
b)	Missed a class because you slept through your alarm or forgot to set your alarm*
c)	Pulled an "all-nighter"*
d)	Forgotten to do a class assignment*
e)	Arrived to class only to realized that you had forgotten about an exam that day*
f)	Been suspended from school for at least one day for any reason
6a)	Had any kind of job
b)	Quit a job after a week
c)	Quit a job without giving at least 2 weeks notice*
7a)	Had a driver's license
b)	Had your driver's license taken away from you by the police
8a)	Driven a car
b)	Been accused of causing a car accident while driving
c)	Gotten a parking ticket⁺
d)	Gotten a speeding ticket⁺
e)	Gotten lost or gone the wrong way for more than 10 min while driving
f)	Locked your keys in the car
9a)	Bought any kind of car
b)	Had to spend at least $500 to fix a car you had owned for less than half a year
10a)	Taken a trip by airplane
b)	Missed a flight
11a)	Taken the train or the bus
b)	Taken the wrong train or bus
12a)	Had any form of ID (driver's license, passport, birth certificate)
b)	Had your ID replaced because you lost it
13a)	Lived in a rented apartment or other rental property
b)	Been kicked out of an apartment or rental property before the lease ran out
14a)	Carried a key to your home
b)	Had the key to your home replaced because you lost it
c)	Locked yourself out of your home
d)	Unintentionally left the door to your home unlocked*
15a)	Been responsible for electricity, cable, gas or water payments
b)	Had your electricity, cable, gas or water shut off because you did not pay on time
16a)	Been responsible for rent or mortgage payments
b)	Paid rent or mortgage payment at least 2 weeks too late
17)	Purchased lottery tickets in the last year*
18a)	Used checks
b)	Had a check bounce
19a)	Had a credit card
b)	Had more than $5000 in credit card debt
20a)	Been to a bar, restaurant, or hotel
b)	Been kicked out of a bar, restaurant, or hotel by someone who works there
21a)	Loaned more than $50 to someone
b)	Loaned more than $50 to someone and never got it back
22a)	Borrowed more than $50 from someone*
b)	Borrowed more than $50 from someone and never paid it back*
23a)	Had a romantic relationship that lasted for at least 1 year
b)	Cheated on your romantic partner of 1 year by having sex with someone else
24a)	Been married
b)	Been divorced
25a)	Had sex
b)	Had an unplanned pregnancy (or got someone pregnant, unplanned)
26a)	Had sex with a condom
b)	Had a condom break, tear, or slip off
27a)	Had an alcoholic drink
b)	Consumed so much alcohol you vomited
c)	Received a DUI for drunk driving
28a)	Been out in the sun
b)	Got blisters from sunburn
29a)	Purchased a product off the television (e.g., an infomercial)*
b)	Purchased a product off the television without reading information about that product's effectiveness*

Table 1 (*Continued*)

Item	
30a)	Read your horoscope*
b)	Found that your horoscope was accurate*
31)	Been in a public fight or screaming argument
32)	Forgotten a birthday of someone close to you and did not realize until the next day or later
33)	Broke a bone because you fell, slipped, or misstepped
34)	Been in a jail cell overnight for any reason
35)	Own a lucky object? (e.g., rabbit foot, etc.)*
36)	Paid to speak to a Psychic (i.e., in person or over-the-phone)*
37)	Purchased Airborne*
38)	Purchased Shape-ups or another shoe that tones with use*
39)	Purchased herbal remedies that enhance thinking or memory*
40)	Owned an object with healing properties (e.g., healing crystals, magnetic bracelets, mystical stones, etc.)*

Note: This inventory was originally developed by de Bruin et al. (2007). Three questions were removed, two questions were modified (denoted by +), and 20 questions were added (denoted by *) to make the inventory for relevant for college students.

2.1. Method

2.1.1. Participants and procedures

Respondents ($n = 131$) from the United States completed the HCTA and the RWO online, 35 were community college students, 46 were state university students, and 50 were community adults. The administration of both measures took approximately 90 min. The sample ranged in age from 18 to 71 ($M = 27.15$, $SD = 13.16$) and a majority of the sample was female (66.40%). The ethnic breakdown was 43.80% Caucasian, 21.90% Hispanic or Latino/a, 15.60% Asian, 7.80% African American, 5.50% multi-racial, and 5.40% reported another ethnicity or declined to state their ethnicity. Fifty-seven percent of the sample had a high school diploma, 14% had associate or technical degrees, 16% had baccalaureate degrees, and 12% had post-graduate educations. Additional details about this study can be found in Butler (in press).

2.1.2. Scoring of the HCTA

The HCTA is administered and scored by computer. It is part of the *Vienna Test System*. The program automatically calculates the quantitative (forced-choice, recognition) responses, and guides the grader through the grading of the qualitative (open-ended, recall) responses with computerized grading prompts. The grader answers grading prompts based on the respondents' answers. For example, based on the example given previously, graders would be asked whether the respondent's answer indicated the consideration of any of the following questions: the number of comparable parolees in similar cities who commit crimes compared to the number in the city, how parolee and repeat offender numbers compare with decisions made by a former parole board, procedures for monitoring early parolees after they are released, the number of criminals who are not granted early parole, and the criteria for early release. The 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 194.

In 2010, the psychometric properties of the HCTA were tested with respondents in the United States ($n = 201$) who were either community adults, or students from a non-selective community college, a state-university, or a selective private liberal arts college. The mean total HCTA score for this standardized sample was 110.54 ($SD = 20.03$). Further analysis of the reliability of the assessment indicates that the HCTA has sufficient reliability (Cronbach $\alpha = .88$) and high inter-rater reliability ($r = .93$). The high inter-rater reliability for free recall responses is due to the unique scoring method (Halpern, 2010a). For the samples from study one, the overall mean score on the HCTA was 103.73 ($SD = 21.22$).

2.1.3. Scoring of the RWO

For the present studies, questions were added or removed from the DOI to make the inventory more age-appropriate for college students. The final inventory is reported here as the *real-world outcomes* (RWO) of critical thinking inventory.[1] Three questions from the original inventory were removed because they involved experiences that college students were unlikely to have experienced (e.g., foreclosure of a home). Two questions were modified and 20 new questions were added to the inventory to make it more applicable to college students (e.g., forgetting about an exam; see Table 1 for the full inventory). Respondents indicate whether or not they have experienced each life event within the past six months.[2] The total RWO score is the proportion of negative life events to possible life events. Scores on the RWO can range from 0 to 1 with higher scores indicating more negative life events or poorer real-world outcomes. The overall mean RWO score was .26 ($SD = .14$).

---

[1] An exploratory critical thinking disposition scale was included with the RWO inventory. The scale did not achieve sufficient internal consistency and was excluded from further analysis.

[2] In the original inventory respondents indicated whether they had experienced these events in the past 10 years, but this timeframe was modified in order to be more applicable to college students.

Table 2
The Halpern Critical Thinking Assessment (HCTA) predicted real-world outcomes (RWO) of critical thinking.

Variable	Real-world outcomes of critical thinking		
	Model 1 β	Model 2	
		β	95% CI
Constant			[.238, .314]
Community college students	−.112	−.103	[−.096, .032]
State university students	.054	.031	[−.044, .062]
HCTA score	−.416*	−.610*	[−.006, −.002]
HCTA by state university students		.184	[.000, .005]
HCTA by community college students		.168	[.000, .005]
R^2	.163		.185
F	8.254*		5.679*
ΔR^2			.022
ΔF			1.684

Note: N = 131. CI, confidence interval; HCTA, Halpern Critical Thinking Assessment score (centered). Community adults were used as the comparison group (constant).
* $p < .001$.

2.2. Results

A one-way ANOVA (sample type: community college students, state university students, and community adults) was conducted to examine differences in critical thinking. Scores on the HCTA varied based on sample type, $F (2, 128) = 8.54$, $p < .001$. Post hoc comparisons with Tukey HSD revealed that the community college students ($M = 92.31$, $SD = 17.50$) scored lower on the HCTA than both the state university students ($M = 105.15$, $SD = 21.49$), and the community adults ($M = 110.42$, $SD = 20.43$). The community adults and the state university students did not significantly differ from each other. Respondents with more years of education had higher scores on the HCTA than respondents with fewer years of education, Spearman's rho $(131) = .26$, $p = .003$. There were no other differences in HCTA scores based on other demographic information, such as gender, age, or annual household income (all $ps > .05$).

The primary interest of this study was whether HCTA scores would predict real-world outcomes of critical thinking. The study also explored the possibility that the predictive ability of the HCTA might vary based on the type of sample (community college students, state university students, and community adults). Both of these predictions were tested in a two-step regression analysis. The first step of the analysis tested the unique contribution of the main effects, HCTA scores and sample membership, in terms of predicting RWO scores. It was predicted that HCTA scores would predict RWO scores and this hypothesis was supported. It was also predicted that the samples would differ based on RWO, but this hypothesis was not supported. The two main effects (HCTA and sample membership) together explained 16% of the variance in scores, $R = −.40$, $F (3, 127) = 8.25$, $p = .001$. Those who scored higher on the HCTA reported fewer negative life events than those who scored lower on the HCTA ($\beta = −.42$, $t = −4.82$, $p = .001$). Sample membership did not predict the extent to which respondents experienced negative life events (RWO). Compared to the community adults, neither community college students ($\beta = −.12$, $t = −1.16$, ns) nor state university students ($\beta = .06$, $t = .70$, ns) differed in the number of negative life events they reported.

The second step of the analysis explored whether there was an interaction between HCTA scores and sample membership (i.e., whether the relationship between HCTA scores and RWO scores differed based on sample membership). Overall, this step in the analysis was significant, $R = −.43$, $F (5, 125) = 5.68$, $p = .001$, but the change in R^2 between the first and second step of the analysis was negligible (R^2 change = .02). Thus, it appears that the HCTA predicted real-world outcomes for each of the samples equally well (see Table 2).

3. Study 2: the Republic of Ireland

The main purpose of Study 2 was to assess the effectiveness of an online critical thinking course. A secondary goal of the study was to replicate the main finding of Study 1 (i.e., the HCTA predicted real-world outcomes of critical thinking) and extend the finding to an Irish population. It was predicted that the online critical thinking course would be effective and that gains in critical thinking would be seen through an increase in HCTA scores (pretest to posttest). It was also hypothesized that the HCTA would predict real-world outcomes in the Irish sample. Portions of this study are published elsewhere (Dwyer, Hogan, & Stewart, under review).

3.1. Method

Two groups of first year psychology students (an experimental group and a control group) from an Irish university participated in this study. The students ranged in age from 18 to 25, 26 were male and 48 were female. Whereas the experimental group ($n = 43$) engaged in a six-week e-learning critical thinking course (which was supported by the use of argument mapping; see Dwyer, Hogan, & Stewart, 2010, 2011), the control group ($n = 31$) received no critical thinking
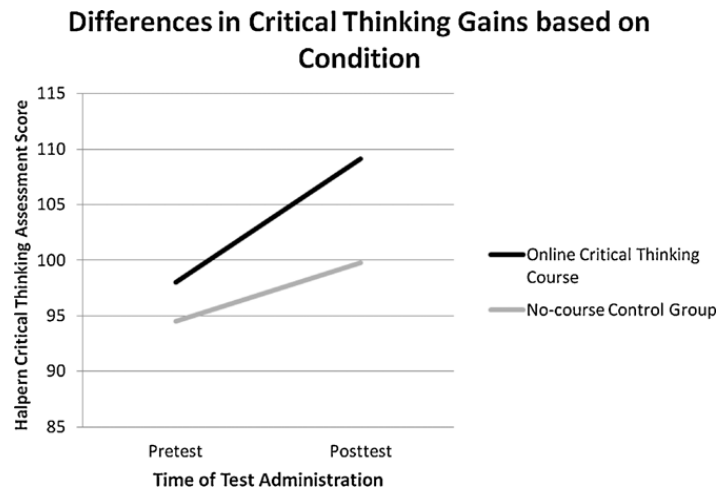
Fig. 1. Differences in Halpern Critical Thinking Assessment (HCTA) scores based of whether students participated in an online critical thinking course or a no-course control group.

intervention. The HCTA was administered to each group before and after the online course. The RWO was administered at pre-testing only.

Students in the experimental group participated in an online critical thinking course in which they viewed classes twice per week, completed two exercise sessions per week, and received detailed feedback for both exercises at the end of each week. Each class presented the educational material to students through argument maps (a visual representation of the structure of an argument). The exercises involved the manipulation of argument maps and completion of relevant critical thinking tasks using argument maps. Students who participated in the course used the *Rationale*™ argument mapping software (van Gelder, 2007) made available to them for purposes of completing their exercises, and they were encouraged to practice using the *Rationale*™ program outside of the course environment.

3.2. Results

The main purpose of this study was to ascertain the effectiveness of the online critical thinking course. The groups did not statistically differ in critical thinking scores at pretesting. A 2 (time: pretest, posttest) $\times$ 2 (condition: experimental group, control group) mixed ANOVA revealed two main effects. First, the HCTA scores for both groups improved from pretest to posttest, $F(1, 72) = 37.42$, $p = .001$, partial $\eta^2 = .34$. Second, the experimental group scored higher on the HCTA than the control group on overall critical thinking performance, $F(1, 72) = 6.72$, $p = .012$, partial $\eta^2 = .09$. Interestingly, the interaction between test time and condition was not statistically significant, perhaps due to the small sample size and lack of statistical power to detect an effect. However, planned follow-up comparison revealed that those who participated in the experimental group exhibited a significantly greater gain in critical thinking than those in the control group, $t(68) = -2.43$, $p = .018$, $d = .60$ (see Fig. 1; Dwyer et al., under review). A secondary goal of this research was to determine whether scores on the HCTA predicted real-world outcomes of critical thinking in an Irish sample. Critical thinking performance was significantly correlated with scores on the RWO, $r(70) = -.28$, $p = .019$. Thus, the main finding from study one, that the HCTA predicted real-world outcomes of critical thinking was replicated in the Irish sample.

4. Study 3: Spain and Portugal

This study explored the psychometric properties of the RWO in Spain and Portugal. This preliminary research was necessary to ensure the validity of future research that aims to explore the predictive ability of the HCTA in these countries. A "cultural" translation of the RWO to Portuguese and Spanish was conducted. Since this inventory is centered on specific situations that have to be dealt with and solved every day, the translation process had to respect the cultural equivalence of such situations and responses to them. Moreover, the psychometric properties of this inventory were analyzed and empirical data that allowed us to make some considerations concerning the quality of decisions made on a daily basis were collected.

4.1. Method

Two university professors and a professional translator converted RWO to Portuguese and Spanish, which was followed by its back-translation to English. Despite their geographic proximity and some shared history, Portugal and Spain are two very distinctive countries, each with their own idiosyncratic identity and culture. Throughout the translation phase, a number of items were particularly challenging to translate, whether because of the idiosyncrasy of a given word or expression (e.g., *Pulled an "all-nighter"*, or even, *I prefer relying on "my gut" when I make decisions*), or because of linguistic and/or cultural

specificities of each country that could prevent the items from sounding authentic and not merely a translation. Another set of items had to undergo adaptations due to manifest cultural differences. For example, *Quit a job without giving at least 2 weeks' notice*, was problematic because in Portugal and Spain it is customary to give at least a month's notice when leaving a job. Additionally, *I believe in some conspiracy theories* was a problematic item because the term "conspiracy theory" is not broadly popularized in the two countries. Finally, items involving certain sums of money (e.g., *Had more than $5000 in credit card debt*) had to be altered, considering not only the exchange rate, but also economic and financial differences differentiating the USA from Portugal and Spain.

Once the translations were complete we compared the Portuguese and Spanish versions of RWO for equivalence, while respecting intrinsic linguistic and cultural specificities. The *Think Aloud* method was used with a sample of 14 students enrolled in the 3rd year of a psychology degree course. We insured that RWO items were clear, comprehensible and relevant to these students before proceeding.

A sample of 475 students from three different universities – two in Portugal, called Portugal University 1 ($n = 227$) and Portugal University 2 ($n = 51$), and one in Spain, called the Spain University ($n = 197$), participated in the study. These students were either in their first ($n = 378$), second ($n = 51$) or third (and final) year ($n = 45$) of a college degree. They were enrolled in Psychology ($n = 216$), Humanities ($n = 137$) or Engineering ($n = 122$) courses. The sample ranged in age from 18 to 50 years old ($M = 20.7$, $SD = 5.19$); 73.5% were female and 26.5% were male (see Table 3).

4.2. Results

The RWO was computed by summing the total number of negative life events. The scores could range from 0 to 53. The Cronbach's alpha coefficient was calculated to study differences in the internal consistency of the RWO based on country, gender, and course. Overall, the RWO evidenced a fit internal consistency index with most internal consistencies at or above .70. A 2 (gender: male, female) × 3 (university sample: Spain University, Portugal University 1, Portugal University 2) ANOVA was conducted to examine differences in RWO scores. There were no differences in RWO scores based on gender, $F(1, 411) = 2.60$, $p = .108$, partial $\eta^2 = 0.003$. There was a main effect of the university, $F(2, 411) = 7.25$, $p = .001$, partial $\eta^2 = 0.033$. Planned contrasts with Sidak correction indicated that students from the Spain University obtained a significantly higher RWO score ($M = 8.81$, $SD = 3.95$) than students from Portugal University 1 ($M = 7.22$, $SD = 4.80$; $t[411] = -3.63$, $p = .001$, $d = 0.362$) and from Portugal University 2 ($M = 8.11$, $SD = 5.10$; $t[411] = -1.06$, $p = .27$, $d = 0.153$). This suggests that daily decision making of students from the Spanish university may be poorer than the decision making of the students from the two Portuguese universities, and this could be for many different reasons. For example, differences between the three universities might reflect differences in the admissions processes at these institutions. In Portugal, access to higher education is limited by a politic of *numerus clausus*, which restricts the number of students admitted at each university every year. The *numerus clausus* politic is not implemented in Spain, and as a result a higher number of students are admitted into college every year. For example, in Portugal University 1 about 60 students are admitted every year into a psychology degree, whereas in the Spain University this number is 350. Another possible factor contributing to the difference between the universities is that the students who participated from each university had different discipline backgrounds. This was a preliminary investigation of the psychometric qualities of the RWO and a necessary first-step towards conducting a direct assessment of the predictive ability of the HCTA in Portuguese and Spanish populations. It provides the basis for future cross-national research in non-English speaking countries.

5. Discussion

The primary objective of these studies was to expand the validity of the Halpern Critical Thinking Assessment (HCTA) cross-nationally and to determine whether HCTA scores predicted real-world outcomes of critical thinking. In general, the results were consistent with these predictions. In the United States (Study 1), the HCTA predicted scores on the real-world outcomes (RWO) of critical thinking inventory. Furthermore, the HCTA predicted RWO scores equally well for community college students, state university students, and community adults. In the Republic of Ireland (Study 2), gains in critical thinking (HCTA scores) were seen for an experimental group of students who participated in an online critical thinking course, compared to the gains seen by the control group. Additionally, scores on the HCTA predicted RWO scores in this Irish sample. In Spain and Portugal (Study 3), the psychometric qualities of a RWO translation were assessed in a preliminary step towards the validation of the HCTA in these countries.

These studies contribute to the development, assessment, and validation of critical thinking assessments. First, they describe various applications of a critical thinking assessment in different countries. Second, they explore new perspectives in the validation of critical thinking assessments, namely validation with real-world outcomes. In two countries, the relationship between critical thinking scores and real-world outcomes of critical thinking was established. Given the plethora of evidence that critical thinking can be improved through instruction (see Chance, 1986; Halpern, 2003; Moseley et al., 2005; Nisbett, 1992), this is an important contribution.

It is important to note several limitations of this research. First, and foremost, the RWO inventory is a self-report measure. The implications of this research would be much stronger if actual behaviors had been observed. It is possible that critical thinkers are simply better at knowing which behaviors to report and which behaviors not to report. A second limitation of any research questions that were answered with a correlational design is that we cannot determine causation based on

Table 3
Characterization of the sample in Study 3.

| Country | University | N | Gender | | Age | | Degree course | | |
			Female	Male	M (SD)	Min–Max	Psychology	Humanities	Engineering
Portugal	University 1	n = 227 (47.8%)	n = 173 (76.2%)	n = 54 (23.8%)	21.9 (6.52)	18–50	n = 19 (8.4%)	n = 137 (60.4%)	n = 71 (31.3%)
	University 2	n = 51 (10.7%)	n = 6 (11.8%)	n = 45 (88.2%)	20.2 (3.27)	18–31	–	–	n = 51 (100%)
Spain	University	n = 197 (41.5%)	n = 170 (86.3%)	n = 27 (13.7%)	19.4 (3.15)	18–43	n = 197 (100%)	–	–

this method of inquiry. Thus, we cannot determine whether critical thinking ability causes people to make poor decisions, vice versa, or whether another factor is influencing both critical thinking and poor decision-making. Future research could explore the causal link between critical thinking and real-world outcomes of critical thinking, with special emphasis on the role of education and behavioral outcomes. Third, only one direct cross-national comparison was made with these studies (Study 3) and this preliminary research did not include a direct standardized measure of critical thinking. Future research should include a standardized assessment of critical thinking and direct cross-national comparisons.

The Halpern Critical Thinking Assessment is a reliable and valid measure of critical thinking. This set of studies expanded the cross-national applications of the assessment, and added a new perspective to traditional validation measures, such as academic aptitude and performance. Furthermore, this research might provide higher education with a new outlet to demonstrate the utility of a critical thinking education.

References

Abrami, P. C., Bernard, R. M., Borokhovski, E., Wade, A., Surkes, M. A., Tamim, R., et al. (2008). Instructional interventions affecting critical thinking skills and dispositions: A stage 1 meta-analysis. *Review of Educational Research*, *78*, 1102–1134.

Association of American Colleges & Universities. (2010). *Raising the bar: Employers' views on college learning in the wake of the economic downturn*. Retrieved from the AAC&U website: http://www.aacu.org/leap

Bureau of Labor Statistics. (2010–2011). *Occupational outlook handbook, 2010–2011 edition*. Retrieved from http://www.bls.gov/oco/ocos021.htm

Butler, H. A. Halpern critical thinking assessment predicts real-world outcomes of critical thinking. *Applied Cognitive Psychology*, in press.

Chance, P. (1986). *Thinking in the classroom: A survey of programs*. New York, NY: Teachers College, Columbia University.

de Bruin, W. B., Parker, A. M., & Fischhoff, B. (2007). Individual differences in adult decision-making competence. *Journal of Personality and Social Psychology*, *92*, 938–956. http://dx.doi.org/10.1037/0022-3514.92.5.938

Dwyer, C. P., Hogan, M. J., & Stewart, I. An evaluation of argument mapping as a method of enhancing critical thinking performance in e-learning settings. *Metacognition and Learning*, under review.

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2010). The evaluation of argument mapping as a learning tool: Comparing the effects of argument map reading versus text reading on comprehension and recall of arguments. *Thinking Skills and Creativity*, *5*, 16–22.

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2011). The promotion of critical thinking skills through argument mapping. In C. P. Hovarth, & J. M. Forte (Eds.), *Critical thinking*. New York: Nova.

Ennis, R. H. (2005). *Supplement to the test/manual entitled The Ennis-Weir Critical Thinking Essay Test*. Urbana, IL: The Illinois Critical Thinking Project. Retrieved from http://faculty.ed.uiuc.edu/rhennis/supplewmanual1105.htm

European Higher Education Area. (2011). *The official Bologna Process website 2010–2012*. Retrieved from the European Higher Education Area website: http://www.ehea.info/

Galagan, P. (2010, February). *Bridging the skills gap: New factors compound the growing skills shortage*. Retrieved from the American Society for Training and Development website: http://www.astd.org/TD/Archives/2010/Feb/Free/1002_BridgingSkillsGap.htm

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

Halpern, D. F. (2010). *Halpern Critical Thinking Assessment*. SCHUHFRIED (Vienna Test System). http://www.schuhfried.com/vienna-test-system-vts/all-tests-from-a-z/test/hcta-halpern-critical-thinking-assessment-1/

Halpern, D. F. (Ed.). (2010). *Undergraduate education in psychology: A blueprint for the future of the discipline*. Washington, DC: American Psychological Association.

Hau, K., Halpern, D. F., Marin-Burkhart, L., Ho, I. T., Ku, K. Y. L., Chan, N., et al. (2006). Chinese and United States students' critical thinking: Cross-cultural construct validation of a critical thinking assessment. In *Paper presented at the American Educational Research Association Annual Meeting* San Francisco, CA.

Hunt, E. (1995). *Will we be smart enough? A cognitive analysis of the coming workforce*. New York, NY: Russell Sage Foundation.

Ku, K. Y. L. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking Skills and Creativity*, *2*, 70–76. http://dx.doi.org/10.1016/j.tsc.2009.02.001

Ku, K. Y. L., Ngai-Man, C., Miu-Chi Lun, V., Halpern, D. F., Marin-Burkhart, L., Hau, K.-T., et al. (2006). Chinese and U.S. undergraduates' critical thinking skills: Academic and dispositional predictors. In *Paper presented at the Annual Meeting of the American Association for Education Research* San Francisco, CA.

Marin, L., & Halpern, D. F. (2011). Pedagogy for developing critical thinking in adolescents: Explicit instruction produces greatest gains. *Thinking Skills and Creativity*, *6*, 1–13. http://dx.doi.org/10.1016/j.tsc.2010.08.002

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

Nisbett, R. E. (1992). *Rules for reasoning*. Hillsdale, NJ: Lawrence Erlbaum.

Redden, E. (2010, June). *Bologna beyond Europe*. Retrieved from the Higher Education News website: http://www.insidehighered.com/news/2010/06/07/bologna

Stanovich, K. E. (2009, November/December). *Rational and irrational thought: The thinking that IQ tests miss*. Scientific American Mind., pp. 34–39. Retrieved from http://www.scientificamerican.com/sciammind/

Stanovich, K. E., & West, R. F. (2008). On the relative independence of thinking biases and cognitive abilities. *Journal of Personality and Social Psychology*, *94*, 672–695. http://dx.doi.org/10.1037/0022-3514.94.4.672

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

U.S. Department of Education. (2006). *A test of leadership: Charting the future of U.S. Higher education*. (DOE Publication No. ED-06-C0-0013). Retrieved from http://ed.gov/about/bdscomm/list/hiedfuture/reports/final-report.pdf

van Gelder, T. J. (2007). The rationale for Rationale^TM. *Law, Probability and Risk*, *6*, 23–42.