

## Translation, adaptation, and validation of the Halpern Critical Thinking Assessment to Portugal: Effect of disciplinary area and academic level on critical thinking

Amanda Rodrigues Franco<sup>1\*</sup>, Patrício Soares Costa<sup>2,3</sup>, and Leandro da Silva Almeida<sup>1</sup>

<sup>1</sup> Instituto de Educação - Universidade do Minho (Portugal).

<sup>2</sup> Faculdade de Psicologia e Ciências da Educação - Universidade do Porto (Portugal).

<sup>3</sup> Escola de Medicina - Universidade do Minho (Portugal).

**Título:** Traducción, adaptación y validación del *Halpern Critical Thinking Assessment* en Portugal: Efecto del área disciplinaria y nivel académico en el pensamiento crítico.

**Resumen:** Actualmente el pensamiento crítico es fundamental, lo dicen las instituciones de enseñanza superior, las empresas y la sociedad. El objetivo de este estudio fue traducir, adaptar y evaluar las propiedades psicométricas de la versión portuguesa del test *Halpern Critical Thinking Assessment* (Halpern, 2012), administrado a 333 estudiantes universitarios. Además, se quiso analizar si hay diferencias en los scores totales en función del área científica de estudios y del nivel académico de los estudiantes. Los resultados apoyan la validez de la versión portuguesa del test e indican diferencias en las dos variables. Estos resultados tienen interés para investigación futura conducida en Portugal y para estudios transculturales utilizando este test para evaluar el pensamiento crítico de estudiantes universitarios. Como el pensamiento crítico combina las competencias y disposiciones necesarias al pensar *bien*, y que, además otros beneficios en los contextos académicos y profesionales, se anticipa que una calidad superior de pensamiento crítico esté asociada a menos resultados negativos cotidianos, esta línea de investigación es de gran relevancia.

**Palabras clave:** pensamiento crítico, Halpern Critical Thinking Assessment; enseñanza superior; evaluación; validación.

**Abstract:** Critical thinking is essential today, state higher education institutions, organizations, and society. This study aimed to translate, adapt, and assess the psychometric proprieties of the Portuguese version of *Halpern Critical Thinking Assessment* (Halpern, 2012), administered to 333 college students. Also, it aimed to analyze if there were differences in total scores according to disciplinary area and academic level. Data support the validity of the Portuguese version of this instrument, and indicate differences in both variables. These findings are of interest in future research conducted in Portugal, and in cross-cultural studies using this test to assess college students' critical thinking. Seeing that critical thinking combines the skill and will that are needed for *good* thinking, and that, alongside benefits in academic and work contexts, a higher quality of critical thinking is expected to be associated with less negative outcomes in everyday life, this line of research is of most relevance.

**Keywords:** critical thinking; Halpern Critical Thinking Assessment; higher education; assessment; validation.

### Introduction

In today's world of information, there is a set of factors that are essential to succeed: understand complex, ambiguous, contradictory information; sort from the massive amount of data which information is legitimate, grounded, and reusable; make *good* decisions (i.e., decisions that are adapted to the problem in hands and that enable an individual to solve it efficiently, avoiding negative outcomes); solve problems efficiently; participate democratically and actively in one's community; enter and stay in an ever more competitive and unusual job market; interact and work with people who have a dissimilar background, culture, language, political party, and/or religion; update skills and forever be a lifelong learner; and more. Here, critical thinking (CT) is claimed to be fundamental (Franco, Butler, & Halpern, 2015; Halpern, 2014; Kay, 2015; Vieira, Tenreiro-Vieira, & Martins, 2011).

Being a hot topic, there are many definitions and theoretical models of CT. From a revision of the literature (Finn, 2011; Franco & Almeida, 2015; Franco et al., 2015; Krupat et al., 2011), CT may be defined as having the ability to use a particular set of higher-order cognitive skills, and being in the disposition to use those higher-order cognitive skills in everyday life, to tilt the odds of success in one's favor.

Hence, thinking becomes *custom-made* thinking, since it is customized to the task, circumstances, and context.

As relevant as it is to define CT, to identify its skills and dispositions, and as imperative as it is to understand the impact that CT has in personal, academic, and professional life spheres, it is also indispensable to consider valid and reliable ways to assess CT (Saiz & Rivas, 2008). There is a set of tests designed to do exactly so, each one focused on (a) particular dimension(s), and using a specific format. There are a few examples of well-known CT assessment measures. The *Watson-Glaser Critical Thinking Appraisal* (Watson & Glaser, 1980, 2005), aimed at grade nine through adulthood, assesses inference, recognition of assumptions, deduction, interpretation, and evaluation of arguments, using multiple-choice items. The *Ennis-Weir Critical Thinking Essay Test* (Ennis & Weir, 1985) is open-ended and aimed at grades seven through college, assessing general CT ability in argumentation. The *California Critical Thinking Skills Test: College Level* (Facione, 1990) is aimed at college students, and assesses analysis, inference, evaluation, deductive reasoning, and inductive reasoning using multiple-choice items. The *Cornell Critical Thinking Tests* have a multiple-choice item format and two levels of assessment: *Level X* (Ennis & Millman, 2005a), aimed at grades 4-14, that assesses induction, deduction, credibility, and identification of assumptions; *Level Z* (Ennis & Millman, 2005b), specially aimed at college students and adults, that assesses induction, deduction, credibility, identification of assumptions, fallacies, definition, prediction, and experimental planning. Also, there is the *Halpern Critical*

**\* Correspondence address [Dirección para correspondencia]:**

Amanda Rodrigues Franco. Instituto de Educação - Universidade do Minho. Campus de Gualtar, 4710-057 Braga, (Portugal).  
E-mail: [amanda.hr.franco@gmail.com](mailto:amanda.hr.franco@gmail.com)

*Thinking Assessment* (HCTA; Halpern, 2012), which will be presented ahead in further detail.

Such a variety of instruments creates a wide range of possibilities for cognitive assessment. Yet, tests have limitations. Some, given their format: although easier to grade, multiple-choice may prevent the possibility of grasping beyond the basic cognitive processes of memorization and recall, or recognition and selection, since this format asks if the solution presented is right or wrong, or requests the selection of one answer from various alternatives. Here, the thinking process may not be (fully) assessed, since respondents are not asked to elaborate their thinking by answering questions or providing constructed answers. On the other hand, some tests present items that seem too artificial to respondents, failing to activate the relevant thinking processes that are applied to solve everyday problems, or even the motivation to invest intellectual effort in responding (Franco, Almeida, & Saiz, 2015; Saiz & Rivas, 2008).

The HCTA test is considered a turning point in the CT assessment scenario, mainly because of its format, the presentation of real-life scenarios, and seeing that it comprises the two components of CT – skills and dispositions. The HCTA test assesses five dimensions that correspond to the five skills that the theoretical model presented by Halpern postulates as core to CT: *verbal reasoning*, *argument analysis*, *thinking as hypothesis testing*, *likelihood and uncertainty*, *decision making and problem solving*. This test combines multiple-choice questions (which demand recognition) and open-ended questions (which demand free recall and elaboration), thus enabling a more comprehensive evaluation of CT: while preventing a long consumption of time when taking the test, it allows the test administrator to analyze the thinking process of respondents. The HCTA test is also attractive for its real-life situations, with which the respondents can easily relate to. This feature is particularly relevant considering that CT is thinking that is contextually applied, very much reliant on the circumstances faced by individuals. Also, it comprises the behavioral and attitudinal components of CT, making it possible to assess both skills and disposition to be a critical thinker (Ku, 2009). Overall, validation studies of the HCTA test consider academic performance and the impact of CT in daily life. Indeed, the relevance of CT becomes clearer when considering the links connecting CT to the quality of decisions made on a daily basis. This real-world validity is one of the main assets of the HCTA test, since it allows to predict if CT results in decisions and action, and to analyze if such outcomes are positive/negative for the performer (Butler et al., 2012; Halpern, 2012).

In this paper, our aim was to analyze the level of CT of the average college student. We describe the process of translation and adaptation of the HCTA test to Portuguese, and present its validation study. Considerations are made about the psychometric proprieties of the Portuguese version, the (dis)similarity with the original version, and the impact of culture to help explain our data, as well as the concept of CT itself. Also, we examined if there are differences

in the quality of CT according to disciplinary area and academic level, seeing that both variables are associated to CT in the literature (e.g., Badcock, Pattison, & Harris, 2010; Brint, Cantwell, & Saxena, 2012; Cisneros, 2009; Drennan, 2010; Mathews & Lowe, 2011).

## Methods

### Participants

A total of 333 students enrolled in a public university located in the North of Portugal participated in this study. The inclusion criteria/requirements were: (i) to be at least 18 years or about to turn 18 (in this study, two participants, having late birthdays, entered university with 17 years); (ii) attend the first year of a Graduate Degree or a Master's Degree; (iii) provide voluntary participation. Participants ranged in age from 17 to 51 ( $M = 22.0$ ,  $SD = 5.65$ ), and a majority was female ( $n = 252$ , 75.7%). This convenience sample was composed of students in the first year of a Graduate Degree ( $n = 179$ , 53.8%) or a Master's Degree ( $n = 154$ , 46.2%) in a diversity of majors (e.g., Biological Sciences; Biomedical Sciences; Communication; Computer Science; Economics; Education; Engineering; Foreign Languages, Literatures and Linguistics; Management; Medicine; Physics; Psychology) in the disciplinary area of Social Sciences and Humanities (SSH) ( $n = 159$ , 47.7%), or Science and Technology (ST) ( $n = 174$ , 52.3%). All academic majors assessed were equally distributed according to academic level (Graduate Degree and Master's Degree).

### Instrument

*Halpern Critical Thinking Assessment.* The HCTA test was originally developed by Halpern in the USA, and has been tested in several other countries, such as Belgium (Verburgh, François, Elen, & Janssen, 2013), Ireland (Dwyer, Hogan, & Stewart, 2012), Spain (Nieto & Saiz, 2008), or even China (Hau et al., 2006; Ku & Ho, 2010). Its reliability and validity have been ascertained in different countries, testing a diversity of participants in concern to their academic level, and using varied methodologies. The HCTA test is composed of two parts that are correlated, thus creating a general CT factor: one pertaining constructed response items, one concerning forced choice items; this structure is confirmed using confirmatory factor analysis in the original study (.51), as well as in other studies in different countries, such as Belgium (.42). As for reliability, the HCTA test showed adequate internal consistency in the original study ( $\alpha = .88$ ), and again, in other studies in different countries, such as Spain ( $\alpha = .78$ ) (for further detail, see Halpern, 2012). This standardized test assesses five dimensions: *Verbal Reasoning* (VR) – to recognize how thought and language influence each other, and to detect techniques that are present in daily language to avoid being manipulated by them; *Argument Analysis* (AA) – to analyze the validity of the arguments that are daily used in

support of a certain decision or action; *Thinking as Hypothesis Testing* (THT) – to keep an empirical attitude when processing information to explain life events and test hypotheses; *Likelihood and Uncertainty* (LU) – to mediate decisions with estimates about the probability of success/failure when performing that decision; *Decision Making and Problem Solving* (DMPS) – to analyze a problem from different angles, to generate alternatives of action, and to select the one with higher chances of success. Twenty-five everyday life scenarios from diverse areas (e.g., education, politics, health, finances) are presented; for each, the respondent must first respond to open-ended questions, and then to multiple-choice questions. The HCTA test combines items that demand free recall and elaboration (open-ended questions), and recognition (multiple-choice questions). Besides, it appeals to both behavioral and motivational components of CT by assessing CT skills and the disposition to be a critical thinker (Halpern, 2012; Ku, 2009). Both components are strongly connected, and for this reason they are not separately evaluated in the Factor Analysis of the HCTA test's items. Construct validity suggests two first-order factors associated to the format of the items (open-ended and multiple-choice items), from which a general second-order factor emerges, identified as CT. This computer-based test is aimed at subjects aged 15 years and older, and can be used for Educational Psychology or in the job market for personnel selection. Its completion time varies between 60 to 80 minutes, and both its administration and grading are computerized. Higher the score in the test – ranging from 0 to 194 points – higher the level of CT.

### Procedures

The author was contacted and permission (from author and editor) was granted to use the HCTA test for our research. This test was translated to Portuguese by a native speaker; the back-translation to English and the comparison between the two versions was undertaken with the assistance of two university teachers and a professional translator. The goal was to identify inaccuracies, and to accomplish an accurate translation and a culturally-fit adaptation. This process required a close collaboration with the author, given the specificity of a few idiomatic expressions and/or due to cultural particularities in the original form of the test, hence, the need to find a proper equivalent in Portuguese. Minor adaptations were made to convey authenticity and familiarity to all items, securing linguistic equivalence while respecting cultural singularities.

A qualitative analysis was undertaken with 14 undergraduates in the third year of a psychology major using the *think aloud* method (Ericsson & Simon, 1993); this analysis with this incidental group aimed to guarantee the clarity, comprehensibility, and relevance of all items. Since the HCTA test is computerized, it was necessary to collaborate with the Portuguese company associated to the company that mar-

kets the HCTA test, to program it and have it operational to administer.

We contacted faculty teachers from several departments to ask them a few minutes of class time to approach participants. In class, we presented the concept and relevance of CT to students, informed them about our study goals, and asked for their voluntary participation. On the other hand, we offered to share with each participant her/his results, so they could find out more about their CT. The students who were interested in participating provided their email to be later on contacted, to arrange a convenient day and schedule for their participation. Participants could also decide if they took the test individually or at the same time as other peers, seeing that the administration of the HCTA test occurred in a classroom specifically designated for this study, equipped with seven computers. Each participant spent between one to two hours taking the test.

All tests were rated by one same grader: open-ended items were graded by analyzing each answer and rating it according to a grading chart presented by the computer program; the multiple-choice items were automatically graded by the computer program. After grading the protocols of the complete sample, data was exported to a SPSS data base, and for each participant we printed a results' sheet produced automatically by the computer program. Following the administration of the HCTA test and its grading, each participant was again contacted to set a convenient day and schedule to provide them their results.

### Statistical analysis

For the sensitivity estimation, descriptive statistics (Min, Max, Me, *M*, *SD*, skewness and kurtosis) were analyzed by CT dimension and CT total score. Absolute values higher than three (for skewness) and 10 (for kurtosis) were considered to challenge the normality of the distribution (Kline, 2000). For the validity estimation, the five-factor model presented for this test was tested through a Confirmatory Factor Analysis (CFA). The model's goodness-of-fit was assessed using a set of indices: Chi Square ( $\chi^2/df$ ), Comparative Fit Index (CFI), Parsimony Comparative Fit Index (PCFI), Goodness-of-Fit Index (GFI), Parsimony Goodness-of-Fit Index (PGFI), Root Mean Square Error of Approximation (RMSEA), and Standardized Root Mean Square Residual (SRMR). The model was considered to have acceptable or good fit, respectively, if:  $\chi^2/df$  was less than 5 or 2; CFI and GFI were higher than 0.8 or 0.9; PCFI and PGFI were higher than 0.6 or 0.8; and RMSEA was lower than 0.10 or 0.08 (Arbuckle, 2008; Wheaton, 1987). For the reliability estimation, we computed omega reliability coefficients for each format dimension and for the total score, given the ordinal response format of items. Values equal to/higher than .70 were considered acceptable (Gadermann, Guhn, & Zumbo, 2012). The significance cutoff point considered for all statistical analyses was  $p < .05$ . Statistical analyses were conducted using the software IBM SPSS Statistics for Win-

dows (version 22.0), AMOS statistical package (version 21.0), and R (version 1.5.8).

## Results

The descriptive statistics are presented in Table 1. Given the computerized grading of each item (the open-ended items manually by one same grader, and the multiple-choice items automatically by the program) and their automatic computation into global scores by the software, items were not analyzed individually (in order to test their difficulty), and for this reason only the descriptive statistics are presented for each dimension and for total score. Additionally, scores are shown according to the items' format – elaboration and recognition – seeing that this double format accounts for singular, yet correlated, skills. Indeed, in the manual of the

HCTA test (Halpern, 2012), the author adopts this same procedure, in order to enable the assessment of these distinct cognitive processes of recall and recognition in participants' performance, hence creating the possibility of a more comprehensive assessment of critical thinking, and also, an easier format for test-taking and test-grading. For each dimension, all items assessed in an open-ended format created a total score concerning elaboration (e.g., VRf) and all items assessed in a multiple-choice format created a total score concerning recognition (e.g., VRr). The maximum points possible is specified for each dimension and for total score. Considering the total score of each dimension, skewness (negatively skewed, ranging between  $-.700$  and  $-.004$ ) and kurtosis (leptokurtic, ranging between  $-.093$  and  $1.570$ ) values are acceptable (Kline, 2005). The average CT total score was  $111.57$  ( $SD = 13.51$ ).

**Table 1.** Descriptive statistics by CT dimension and CT total score.

CT dimensions	Min	Max	Me	<i>M</i>	<i>SD</i>	sk	ku
VR (22 points)	4	20	12.00	12.05	3.01	-.004	-.093
VRf (15 points)	0	14	8.00	7.94	2.54	-.045	-.164
VRr (7 points)	1	7	4.00	4.10	1.14	-.214	-.041
AA (41 points)	5	37	24.00	23.44	4.57	-.379	.576
AAf (22 points)	0	20	12.00	12.16	3.04	-.237	.744
AAr (19 points)	2	17	12.00	11.28	2.68	-.539	.239
THT (46 points)	12	36	24.00	24.47	4.33	-.049	.150
THTf (19 points)	0	16	10.00	9.62	2.45	-.080	.138
THTr (27 points)	5	23	15.00	14.85	2.98	-.070	-.081
LU (24 points)	4	19	13.00	12.41	2.92	-.367	.162
LUf (17 points)	0	13	8.00	8.18	2.34	-.322	.087
LUR (7 points)	1	7	4.00	4.24	1.16	-.109	.002
DMPS (61 points)	14	51	40.00	39.20	5.05	-.700	1.570
DMPSf (22 points)	0	19	13.00	12.59	3.22	-.731	1.151
DMPSr (39 points)	14	33	27.00	26.61	3.12	-.704	1.136
CT (194 points)	61	148	113.00	111.57	13.51	-.352	.473
CTfree (95 points)	4	73	51.00	50.50	8.95	-.576	1.744
CTrec (99 points)	39	81	61.00	61.08	6.71	-.236	.155

*Note.* Minimum (Min) and maximum (Max) values, Median (Me), Mean (*M*), Standard deviation (*SD*), skewness (sk) and kurtosis (ku) for CT and its dimensions - verbal reasoning (VR), argument analysis (AA), thinking as hypothesis testing (THT), likelihood and uncertainty (LU), and decision making and problem solving (DMPS), by total score, free recall (e.g., ...f) and recognition (e.g., ...r).

Concerning the validity of the Portuguese version, we tested the five-factor model presented in the manual of the HCTA test through a CFA to analyze its factorial structure (see Figure 1). Similarly to the original study (Halpern, 2012), we tested a measurement model (M1) and analyzed its goodness-of-fit. In M1, scores for the free-recall items in all five dimensions were hypothesized to load on a latent factor

designated "CT free-recall" (CTfree), while scores for the multiple-choice items in all five dimensions were hypothesized to load on a latent factor designated "CT recognition" (CTrec). Both latent factors – CTfree and CTrec – were correlated to test the hypothesis that the two response formats, free-recall and recognition, were differentiated yet associated aspects of CT.

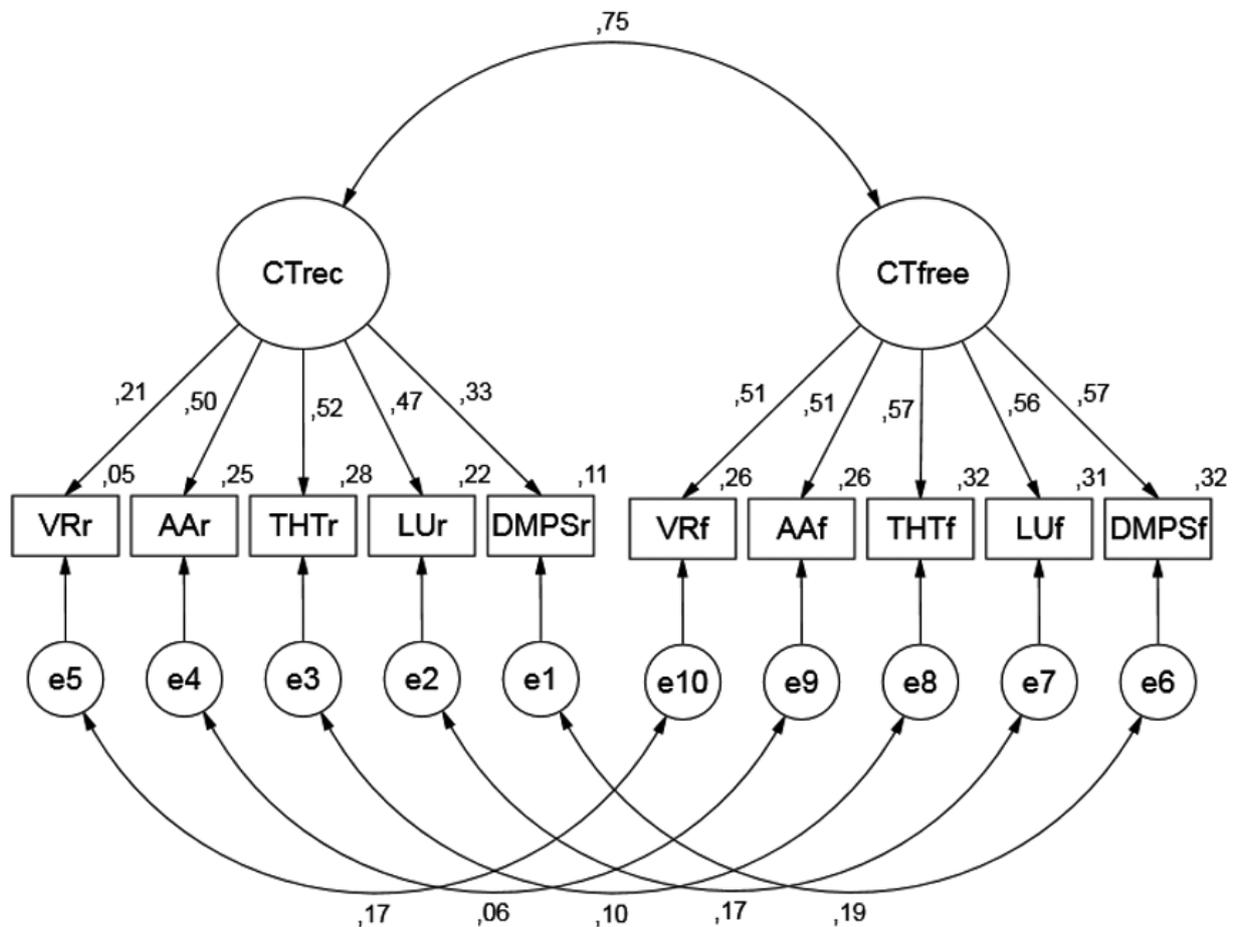


Figure 1. Standardized factor loadings of the HCTA test measurement model.

The indices considered to decide on the model's goodness-of-fit supported the five-factor structure presented for the original HCTA test with acceptable (PCFI and PGFI) to good ( $\chi^2/\text{df}$ , CFI, GFI, and RMSEA) fit. The goodness-of-fit was slightly lower in our sample, yet, accomplished ( $\chi^2[29, n = 334] = 58.6, p = .001; \chi^2/\text{df} = 2.02; \text{CFI} = .929; \text{PCFI} = .599; \text{GFI} = .966; \text{PGFI} = .509; \text{RMSEA}[\text{CI}90\%] = .055[.035; .076]$ ). For test of close fit ( $\text{RMSEA} < .05$ ) = .308,  $\text{SRMR} = .0438$ . Overall, M1 explains 56.3% of shared variance between CTfree and CTrec, when compared to the 75.7% obtained in the original study.

Concerning the reliability estimation for total CT, CTrec and CTfree, we found reliability coefficients to be acceptable for CT ( $\Omega = .75$ ) and CTfree ( $\Omega = .70$ ), yet not as satisfactory for CTrec ( $\Omega = .58$ ).

We performed a one-way MANOVA for disciplinary area and for academic level to test if there were differences in the total HCTA scores concerning CTfree and CTrec. Statistically significant differences were found for disciplinary area,  $F(2, 330) = 26.00, p < .0005; \text{Wilk's } \Lambda = 0.864, \text{partial } \eta^2 = .14$ . Disciplinary area had a statistically significant effect on both CTrec ( $F(1, 331) = 20.50; p < .0005; \text{partial } \eta^2 =$

.06) and CTfree ( $F(1, 331) = 49.44; p < .0005; \text{partial } \eta^2 = .13$ ), with ST students scoring higher in CTrec ( $M = 62.62, SD = 6.36$ ) and CTfree ( $M = 53.57, SD = 8.19$ ) when compared to SSH students in each CT score ( $M = 59.38, SD = 6.68$ , and  $M = 47.13, SD = 8.54$ , respectively). Also, we found a statistically significant difference in CT scores according to academic level,  $F(2, 330) = 2.92, p = .05; \text{Wilk's } \Lambda = 0.983, \text{partial } \eta^2 = .02$ . Academic level had a statistically significant effect on CTfree ( $F(1, 331) = 5.62; p < .05; \text{partial } \eta^2 = .02$ ), with Master students scoring higher ( $M = 51.74, SD = 8.53$ ) than Graduate students ( $M = 49.42, SD = 9.18$ ), yet not on CTrec ( $F(1, 331) = 2.41; p = .12; \text{partial } \eta^2 = .01$ ).

## Discussion

This study adds to the HCTA test validation literature and its value to assess CT. We found fit indices for descriptive statistics, pointing to the normality of the distribution. Concerning validity, our model's goodness-of-fit was slightly lower when compared to the original study, yet, accomplished, with a model explaining 56.3% of shared variance

between CT<sub>free</sub> and CT<sub>rec</sub>. As for precision, it can be accepted. Perhaps CT itself has a different meaning across cultures, hence, its manifestation in dispositions/behavior is also singular to each culture. This hypothesis gains strength considering data about the HCTA test validity across cultures: for instance, in the Spanish version, with a sample of 355 adults, the correlation between free recall and recognition formats was .49; in the Dutch version, with a sample of 173 college students in Belgium, this correlation between the two formats was .42 (Halpern, 2012). Indeed, the cross-cultural validity of the HCTA test, by comparing its different translated versions, is a pertinent development and requires further research.

Performing a one-way MANOVA for disciplinary area and academic level, statistically significant differences were found. Disciplinary area had a statistically significant effect, with ST students scoring higher in CT<sub>rec</sub> and CT<sub>free</sub> when compared to SSH students. As for academic level, it showed a statistically significant effect on CT<sub>free</sub> alone, with Master students scoring higher than Graduate students. In light of our data, CT scores vary according to disciplinary area; yet, it may be necessary to go beyond such label and identify the variables that are truly illustrative of the reasons why students (that happen to be) from different disciplinary areas are unlike in regard to CT. On the one hand, disciplinary differences could be due to a particular emphasis set on skills/dispositions related to CT in the form of learning contexts, classroom characteristics, course content, or pedagogical practices (Badcock et al., 2010; Brint et al., 2012; Mathews & Lowe, 2011). For instance, Heijltjes, Van Gog, and Paas (2014) found that learning CT skills is potentiated by the combination of explicit CT instruction and practice. On the other hand, disciplinary differences could be due to individual characteristics that impact when time comes to enroll in a particular course from a certain disciplinary area (Badcock et al., 2010), seeing that neurocognitive systems, in interaction with life experiences, are responsible for individual differences in CT (Bolger, Mackey, Wang, & Grigorenko, 2014). Also, it is necessary to account for the impact of educational background, as suggested by Evens, Verburgh, and Elen (2013), who found that Secondary Education impacts the choice of disciplinary area, which relates to students' CT performance. Overall, CT differences may be mostly due to pedagogical practices, so "educators will need to study the practices of exemplary instructors rather than exemplary disciplines, because no disciplines appear to stand out in this domain" (Brint et al., 2012, p. 22).

As for academic level, it appears to differentiate students when CT is assessed in a free-recall format, with Master students scoring higher than Graduate students. In regard to this academic variable, such differences may be contextualized in light of the cumulative impact of learning, which may also affect the quality of CT. Concurrently, students in more advanced academic years would have a wider understanding of the college experience concerning academic learning, but also a deeper personal growth that accompanies the process

of attaining college (Cisneros, 2009; Drennan, 2010). Altogether, this could have an impact on the quality of thinking, with higher education exerting a positive influence on CT, requiring students to become more active, and at the same time, making them more prone to think things through, to make sound decisions, to solve problems. Also, CT takes time and practice to develop, so a higher level of CT could be expected as students move along college (Halpern, 2014). Nevertheless, it is risky to elaborate on such considerations, seeing that it is not a same sample of students who was assessed in the first year of a Graduate Degree, and three years later in the first year of a Master's Degree.

## Conclusions

"Today it is less important to know, and more important to know what is known", states John Kay (2015), referring to the advantages of a liberal Education, in his column on economics and business in the Financial Times. So it seems that although important to possess the "facts and figures" type of knowledge, it is very important to possess knowledge about which facts and figures are relevant and reliable, and then, about where to have access to such facts and figures, and how to (re)use them ethically and efficiently. This creates implications for higher education institutions, by implying that higher education should: (1) focus on thinking, not as overwhelmingly on "knowing"; (2) aim that students become more "generalized" concerning knowledge, not particularly more specialized; (3) develop transversal skills such as CT, not merely job-specific skills; (4) value the Social Sciences, such as Philosophy, Anthropology, or Sociology, and the skills and dispositions that are rooted in this field of knowledge, instead of predominantly supporting Science, Technology, Engineering and Mathematics. "The objective should be to equip students to enjoy rewarding employment and fulfilling lives in a future environment whose demands we can neither anticipate nor predict ... the capacities to think critically, judge numbers, compose prose and observe carefully ... will be as useful then as they are today" (Kay, 2015). Indeed, CT matters and higher education moulds (Franco, 2016). And one criteria to assess any institution of Education is its focus on the preparation of students, not only for employment, but for life. And here, "the cultivated capacities for critical thinking and reflection are crucial in keeping democracies alive and wide awake" (Nussbaum, 2010, p. 10).

**Acknowledgments:** We thank Prof. Diane F. Halpern for generously granting permission to adapt and use the HCTA test for our research purposes. We give special thanks to all 333 critical thinkers who kindly gave their time and cognitive energy to this study, as well as to their teachers, who made it possible for us to find these 333 critical thinkers.

**Funding:** Research financed by Fundação para a Ciência e a Tecnologia [Foundation for Science and Technology]

(SFRH/BD/76372/2011) (QREN Program – POPH – Typology 4.1 – Advanced Training).

## References

- Arbuckle, J. L. (2008). *Amos 17 users' guide*. Chicago, IL: SPSS. ISBN-10: 1-56827-402-5
- Badcock, P. B. T., Pattison, P. E., & Harris, K. L. (2010). Developing generic skills through university study: A study of arts, science and engineering in Australia. *Higher Education, 60*(4), 441-458. doi: 10.1007/s10734-010-9308-8
- Bolger, D. J., Mackey, A. P., Wang, M., & Grigorenko, E. L. (2014). The role and sources of individual differences in critical-analytic thinking: A capsule overview. *Educational Psychology Review, 26*(4), 495-518. doi: 10.1007/s10648-014-9279-x
- Brint, S., Cantwell, A. M., & Saxena, P. (2012). Disciplinary categories, majors, and undergraduate academic experiences: Rethinking Bok's "underachieving colleges" thesis. *Research in Higher Education, 53*(1), 1-25. doi: 10.1007/s11162-011-9227-2
- Butler, H. A., Dwyer, C. P., Hogan, M. J., Franco, A., Rivas, S. F., Saiz, C., & Almeida, L. S. (2012). The Halpern Critical Thinking Assessment and real-world outcomes: Cross-national applications. *Thinking Skills and Creativity, 7*(2), 112-121. doi: 10.1016/j.tsc.2012.04.001
- Cisneros, R. M. (2009). Assessment of critical thinking in pharmacy students. *American Journal of Pharmaceutical Education, 73*(4), 1-7. doi: https://doi.org/10.5688/aj730466
- Drennan, J. (2010). Critical thinking as an outcome of a master's degree in nursing programme. *Journal of Advanced Nursing, 66*(2), 422-431. doi: 10.1111/j.1365-2648.2009.05170.x
- Dwyer, C. P., Hogan, M. J., & Stewart, I. (2012). An evaluation of argument mapping as a method of enhancing critical thinking performance in e-learning environments. *Metacognition Learning, 7*(3), 219-244. doi: 10.1007/s11409-012-9092-1
- Ennis, R. H., & Millman, J. (2005a). *Cornell Critical Thinking Test, Level X* [Measurement instrument]. Seaside, CA: The Critical Thinking Company (formerly Midwest Publications).
- Ennis, R. H., & Millman, J. (2005b). *Cornell Critical Thinking Test, Level Z* [Measurement instrument]. Seaside, CA: The Critical Thinking Company (formerly Midwest Publications).
- Ennis, R. H., & Weir, E. (1985). *The Ennis-Weir Critical Thinking Essay Test* [Measurement instrument]. Pacific Grove, CA: Midwest Publications.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data*. Cambridge, MA: MIT Press. http://dx.doi.org/10.1037/0033-295X.87.3.215
- Evens, M., Verburgh, A., & Elen, J. (2013). Critical thinking in college freshmen: The impact of secondary and higher education. *International Journal of Higher Education, 2*(3), 139-151. doi: 10.5430/ijhe.v2n3p139
- Facione, P. (1990). *The California Critical Thinking Skills Test: College Level* [Measurement instrument]. Millbrae, CA: California Academic Press. doi: 10.5430/ijhe.v2n3p139
- Finn, P. (2011). Critical thinking: Knowledge and skills for evidence-based practice. *Language, Speech, and Hearing Services in Schools, 42*(1), 69-72. doi: 10.1044/0161-1461(2010/09-0037)
- Franco, A. (2016). What do *Ode to Joy*, the Nobel Peace Prize, umbrellas, and cartoons have in common? Why critical thinking matters and how higher education moulds. *Higher Education for the Future, 3*(1), 108-124. doi: 10.1177/2347631115610231
- Franco, A., & Almeida, L. S. (2015). Real-World Outcomes and critical thinking: Differential analysis by academic major and gender. *Paidéia, 25*(61), 173-181. http://dx.doi.org/10.1590/1982-43272561201505
- Franco, A. R., Almeida, L. S., & Saiz, C. (2014). Pensamiento crítico: Reflexión sobre su lugar en la enseñanza superior [Critical thinking: Reflections about its place in higher education]. *Educatio Siglo XXI, 32*(2), 81-96. http://dx.doi.org/10.6018/j/202171
- Franco, A. H. R., Butler, H. A., & Halpern, D. F. (2015). Teaching critical thinking to promote learning. In D. S. Dunn (Ed.), *The Oxford handbook of undergraduate psychology education* (pp. 65-74). New York, NY: Oxford University Press. ISBN: 0199933820
- Gademann, A. M., Guhn, M., & Zumbo, B. D. (2012). Estimating ordinal reliability for Likert-type and ordinal item response data: A conceptual, empirical, and practical guide. *Practical Assessment, Research & Evaluation, 17*(3), 1-13. ISSN: 1531-7714
- Halpern, D. F. (2012). *Halpern Critical Thinking Assessment* [Measurement instrument]. Mödling: Schuhfried.
- Halpern, D. F. (2014). *Thought and knowledge: An introduction to critical thinking* (5th ed.). New York, NY: Psychology Press. ISBN: 13: 978-1848726291
- Hau, K. T., Halpern, D. F., Marin-Burkhardt, L., Ho, I. T., Ku, K. Y. L., Chan, N. M., & Lun, V. M. C. (2006, April). *Chinese and United States Students' critical thinking: Cross-cultural construct validation of a critical thinking assessment*. Paper presented at the 2006 Annual Meeting of the American Educational Research Association: Education Research in the Public Interest, San Francisco, California.
- Heijltjes, A., Van Gog, T., & Paas, F. (2014). Improving students' critical thinking: Empirical support for explicit instructions combined with practice. *Applied Cognitive Psychology, 28*(4), 518-530. doi: 10.1002/acp.3025
- Kay, J. (2015, August 25). The benefits of a liberal education do not go out of date. *Financial Times*. Retrieved from http://www.ft.com
- Kline, P. (2000). *The handbook of psychological testing* (2nd ed.). London: Routledge.
- Kline, R. B. (2005). *Principles and practice of structural equation modeling* (2nd ed.). New York, NY: Guilford Press.
- Krupat, E., Sprague, J. M., Wolpaw, D., Haidet, P., Hatem, D., & O'Brien, B. (2011). Thinking critically about critical thinking: Ability, disposition or both? *Medical Education, 45*(6), 625-635. doi: 10.1111/j.1365-2923.2010.03910.x
- Ku, K. Y. L. (2009). Assessing students' critical thinking performance: Urging for measurements using multi-response format. *Thinking Skills and Creativity, 4*(1), 70-76. doi: 10.1016/j.tsc.2009.02.001
- Ku, K. Y. L., & Ho, I. T. (2010). Dispositional factors predicting Chinese students' critical thinking performance. *Personality and Individual Differences, 48*(1), 54-58. doi: https://doi.org/10.1016/j.paid.2009.08.015
- Mathews, S. R., & Lowe, K. (2011). Classroom environments that foster a disposition for critical thinking. *Learning Environments Research, 14*(1), 59-73. doi: 10.1007/s10984-011-9082-2
- Nieto, A. M., & Saiz, C. (2008). Evaluation of Halpern's "structural component" for improving critical thinking. *The Spanish Journal of Psychology, 11*(1), 266-274. doi: https://doi.org/10.1017/S1138741600004303
- Nussbaum, M. C. (2010). *Not for profit: Why democracy needs the humanities*. Princeton, NJ: Princeton University Press. ISSN: 1534-6104
- Revelle, W. (2015). psych: Procedures for Personality and Psychological Research, Northwestern University, Evanston, Illinois, USA, http://CRAN.R-project.org/package=psych Version = 1.5.8.
- Saiz, C., & Rivas, S. F. (2008). Evaluación en pensamiento crítico: Una propuesta para diferenciar formas de pensar. *Ergo, Nueva Época, 22-23*, 25-66.
- Verburgh, A., François, S., Elen, J., & Janssen, R. (2013). The assessment of critical thinking critically assessed in higher education: A validation study of the CCTT and the HCTA. *Education Research International, 2*(1), 1-13. doi: http://dx.doi.org/10.1155/2013/198920
- Vieira, R. M., Tenreiro-Vieira, C., & Martins, I. P. (2011). Critical thinking: Conceptual clarification and its importance in science education. *Science Education International, 22*(1), 43-54. ISSN: ISSN-2077-2327
- Watson, G., & Glaser, E. M. (1980, 2005). *Watson-Glaser Critical Thinking Appraisal* [Measurement instrument]. London: Pearson Assessment.
- Wheaton, B. (1987). Assessment of fit in overidentified models with latent variables. *Sociological Methods & Research, 16*(1), 118-154. doi: 10.1177/0049124187016001005

(Article received: 22-10-2016; revised: 12-12-2016; accepted: 20-12-2016)