

Development of a measurement index of critical thinking in professional formation

Beatriz Elena Ospina Rave¹
Edinson Gabriel Brand Monsalve²
Carlos Andrés Aristizabal Botero³

Development of a measurement index of critical thinking in professional formation

Objective. This research sought to construct and validate a measurement index of critical thinking (CT) in professional formation. **Methods.** Transversal, quantitative, test validation-type study. After reviewing scientific production in CT, which permitted defining the concept and its conceptual domains, a 65-item scale was constructed of closed questions on the analysis of cases evaluating CT. The scale was subjected to expert evaluation to then be applied to 53 undergraduate students (35 from nursing and 18 from sociology) to evaluate validity and reliability. **Results.** The 65-item scale has an explained variance of 61.3% and is comprised of five CT domains: inference, evaluation, argumentation, analysis, and interpretation. A Cronbach's alpha coefficient of 0.61 was obtained. **Conclusion.** The scale proposed to assess students' CT skills converges with concepts by known authors

with the CT theory and is adequate for use as a CT measurement index in professional formation.

Descriptors: thinking; educational evaluation; education, higher; scales.

Desarrollo de un índice de medición del pensamiento crítico en la formación profesional

Objetivo. Construir y validar un índice de medición del pensamiento crítico (PC) en la formación profesional. **Métodos.** Estudio cuantitativo transversal del tipo de validación de pruebas. Luego de revisar la producción científica en PC, la definición del concepto y sus dominios conceptuales, se construyó una escala de 65 ítems de preguntas cerradas sobre análisis de casos que evalúan PC. La escala fue sometida a valoración de expertos para luego ser aplicada a 53 estudiantes de pregrado (35 de enfermería y 18 de sociología) con el fin de evaluar la validez y la confiabilidad. **Resultados.**

1 Nurse, M.Sc. Professor, Universidad de Antioquia -UdeA. Calle 70 No. 52-21, Medellín, Colombia.
email: beospinar7@gmail.com

2 Sociologist, M.Sc. Professor, Universidad de Antioquia -UdeA. Calle 70 No. 52-21, Medellín, Colombia.
email: ebrand8@gmail.com

3 Sociologist, M.Sc. Universidad de Antioquia -UdeA. Calle 70 No. 52-21, Medellín, Colombia.
email: carlos.aristizabal@udea.edu.co

Article linked to the research: Desarrollo índice de medición del pensamiento crítico en la formación profesional. 2013-2015.

Conflicts of interest: none.

Receipt date: July 13, 2016.

Approval date: January 31, 2017.

How to cite this article: Ospina BE, Brand EG, Aristizabal CA. Development of a measurement index of critical thinking in professional formation. Invest. Educ. Enferm. 2017; 35(1): 69-77

DOI: 10.17533/udea.iee.v35n1a08

La escala de 65 ítems tiene una varianza explicada del 61.3% y está compuesta por cinco dominios de PC: inferencia, evaluación, argumentación, análisis e interpretación. Se obtuvo un coeficiente alfa de Cronbach de 0.61. **Conclusión.** La escala propuesta para evaluar las habilidades del PC del estudiante converge con conceptos de autores reconocidos con la teoría de PC y es adecuada para ser utilizada como índice de medición de PC en la formación profesional.

Descriptor: pensamiento; evaluación educacional; educación superior; escalas

Desenvolvimento de um índice de medição do pensamento crítico na formação profissional

Objetivo. Construir e validação um índice de medição do pensamento crítico (PC) na formação profissional.

Métodos. Estudo quantitativo transversal do tipo

Introduction

Research exercises to measure CT have been conducted for over three decades, during which different instruments have emerged and have been positioned for said purpose, evaluating results from specific scenarios to promote CT, as well as everyday scenarios of formative nature such as school and professional formation. The most recent development of CT measurement exercises in the last six years evidence that interest has been oriented into two lines: in the first place, measurement made by applying already existing instruments and validated within the scientific community, in which some account for CT skills;^{1,2} others for the dispositions³ and others for both,⁴ both in scenarios intended for CT promotion and in everyday formation spaces. Secondly, there are exercises dedicated to constructing measurement proposals, based on preexisting measurement instruments⁵ or on the conceptualization of CT,⁶ given that the latter sought a construction of items specific to emerging conditions for which a specific CT definition was applied. These construction exercises evidence the interest for achieving measurement instruments that adjust to the specific conditions of each case.

de validação de provas. Logo de revisar a produção científica em PC, o que permitiu a definição do conceito e seus domínios conceituais, se construiu uma escala de 65 itens de perguntas fechadas sobre análise de casos que avaliam PC. A escala foi submetida a valoração de peritos para logo ser aplicada a 53 estudantes de pré-graduação (35 de enfermagem e 18 de sociologia) com o fim de avaliar a validade e a confiabilidade.

Resultados. A escala de 65 itens tem uma variável explicada de 61.3% e está composta por cinco domínios de PC: inferência, avaliação, argumentação, análise e interpretação. Se obteve um coeficiente alfa de Cronbach de 0.61. **Conclusão.** A escala proposta para avaliar as habilidades do PC do estudante converge com conceitos de autores reconhecidos com a teoria de PC, e é adequada para ser utilizada como índice de medição de PC na formação profissional.

Descriptor: pensamento; avaliação educacional; educação superior; escalas.

As a result of the review of these studies, as well as of others registered since 1960, various situations were found that currently permit problematizing the CT measurement. In the first place, with respect to the structure, two aspects were evidenced, one referring to the scales with essay-type open questions, which pose greater difficulty for their reproducibility – considering the elements that must be evaluated within the texts constructed, requiring from evaluators a high level of CT knowledge, without this guaranteeing a standardized criterion in scoring these types of tests. The other aspect of the structure has to do with the Likert-type questions that seek a self-evaluation of those evaluated with respect to their behavior or their behavioral trend regarding a given situation, evidencing that these types of questions lead more to a self-evaluated projection of CT skills and dispositions, based on the concept of each of those evaluated on the situation proposed.

A second situation corresponds to the composition of the scale, that is, to the concepts measured in the different scales, observing that similar skills are evaluated – although denominated with different concepts and, in some cases, these

skills are assessed in disaggregated manner and in others in grouped manner, evidenced in the definitions made of each of the concepts assessed. This situation has led to many proposals that generate a situation of high fragmentation in measurement and, consequently, to a lower probability of achieving its standardization, which permits reproducibility in different contexts. Finally, there is a third situation that refers to the lack of proposals in Colombia that aim to respond to the two previous situations, having theoretical reflections and methodological construction to promote CT in school classrooms and universities, but not on its measurement to evaluate progress in this subject. In response to these problematic situations, this text presents the construction and validation process of a measurement index of CT in professional formation to propose a first approach to generating tools in the country, as well as of sensitizing on the importance of consolidating a measurement instrument that responds to the specific conditions in which formative processes are developed, achieving a proposal that can be replicated in different scenarios, easily read by institutions and teachers in general.

For this design, integrating the best definitions proposed in the academic field,⁵⁻¹³ researchers understand CT as the “rational cognitive process of higher order, developed from tendencies to thinking in a certain form and doing something under given conditions, which integrates conceptual knowledge with experience through contextualized problematizing, leading to construction hypotheses supported on evidence to reach judgment. This thinking implies inference, evaluation, argumentation, analysis, and interpretation.” In nursing, it is necessary to advance in the promotion and development of critical thinking in the scientific, technical, human, and social formation, which favors the construction of the horizon of care, and a significant teaching learning experience for nursing professionals within the social context. The aim of this study was to construct and validate a CT measurement index in professional formation.

Methods

This was a test validation descriptive study. To measure CT, multiple instruments have been developed among which there are qualitative, quantitative and mixed proposals, having a variety of scales that try to evaluate different components, among them: the California Critical Thinking Disposition Inventory (CCTDI),⁸ which tries to measure dispositions to thinking critically through a scale of 71 items, distributed in seven subscales, and assessed through a Likert scale per level of agreement through self-evaluation; the California Critical Thinking Skills Test (CCTST)¹⁴ that seeks to evaluate CT skills with 34 multiple-response items, distributed into five subscales; the Watson-Glaser Critical Thinking Appraisal (WGCTA) composed of 80 items distributed into five subscales, reduced to three subscales in its second version (WGCTA II),⁹ the Halpern Critical Thinking Assessment (HCTA)¹⁵ poses 25 everyday scenarios, distributed into five subscales, five scenarios per each, where those evaluated are first approached from open questions or questions of construction, followed by mandatory selection questions (multiple-response, range, or classification of alternatives); finally, with less cases of application, there is the Health Sciences Reasoning Test (HSRT)¹⁶ that is comparable to the CCTDI, but with a more specific application for health sciences, being derived specifically from CCTST but aiming to recognize elements of the context; this instrument is composed of 33 multiple-response items distributed into five subscales. Besides these scales, used in different parts of the world, others are elaborated based on the CT definitions they apply, with most using the structure of Likert-type scales through self-evaluation of behaviors, as well as open questions of essay-type construction.

From this balance, this construction exercise was guided toward a scale that, in the first place, used a structure of closed questions with single responses on specific cases presented through readings, which permits its reproducibility without requiring a very high level of CT knowledge, as well as the direct evaluation of CT skills and not

their projection; secondly, triangulate concepts in which the different measurement proposals present common elements in their definition to validate a scale that gathers the different proposals to contribute to greater unification, given that a fundamental finding was that several of the best positioned scales name differently skills that are equal, which generates a high degree of compatibility rather than differentiation; with the highest objective being that of having a measurement scale that, through an index, evaluates the conditions of higher education students in the region and the country.

This process of construction and validation of a CT measurement scale was developed between 2013 and 2015 in the city of Medellín-Colombia, specifically at Universidad de Antioquia, with an application to 53 students, 35 from Nursing and 18 from Sociology; as population of principal interest because these programs seek the promotion of CT skills. The work started with the descriptive phase, which constructed the current state of CT knowledge, identifying gaps and achievements in its definition, research trajectories and conceptualization on the theme, and knowledge of the proposals by different authors, to comprehend their integrality as object of study, achieving the construction of the conceptual and methodological base of the investigation. A search was made in the scientific databases of Health Sciences, Education, and Social Sciences, using databases, like Science Direct, Embase – Elsevier, Medline, Dialnet, EBSCO, ISI, MD consult, Blackwell, SciELO, Redalyc, JSTOR, Wilson Web, SpringerLink, Current Contents, and PubMed. In the analytical and interpretative phase, synthesis and elaboration of conclusions was carried out by relating data obtained through different techniques, identifying categories, authors, and currents of thought, which led to more precise identification of trends, convergences, and contradictions.

The final process was the selection of domains, from the conceptual base matrix, defining as CT skills object of measurement for the project: inference, evaluation, argumentation, analysis,

and interpretation, for which their definition was constructed from the most relevant authors. The five domains constructed were evaluated by experts to validate their conceptual construction to measure the skills of the CT construct. These voices qualified based on renowned experience, issued a concept on the work developed by the research team. The group of experts was selected bearing in mind their expertise on the CT topic and on the method to construct and apply measurement scales. The call process was conducted considering the consultation of databases specialized on the theme, which indicated authors, origin, articles, or scales. In total, ten national and international experts were consulted, obtaining response from four of them. The responses obtained were aimed at the analysis of the validity of content and the validity of construct. One of the experts considered that the indicators correspond to the characteristics clearly stated by the authors referenced in the investigation, the instrument corresponds to an adequate design, measures what it seeks to measure, and displays validity of construct and content.

Also, another expert considered that the instrument may be excellent in logical and verbal terms, but the usefulness must be clear for the purposes it pursues, the users, and the context. Its evaluation alerts that three conditions must be considered on the instrument: *cognitive-linguistic adequacy*, referring that the indicators are expressed so that they turn out in their vocabulary and syntactic construction comprehensible and applicable by its users (professors and students) because it corresponds to their levels of cognitive-linguistic development. *Systematicity*, referring to the indicators being expressed in a way to permit evaluating to what extent students proceeded in flexible or systematic manner in executing the skill being evaluated, and *criticity*. Finally, another expert manifested the need make it explicit, as part of the theoretical framework, if CT skills are independent or if they are dimensions of a single construct.

The recommendations of the experts were included in the scale prior to running the pilot test, with this being the next step that consisted in applying the

scale to first and last semester students, aiming to evaluate the situation when starting and ending the higher formation process. This test had the fundamental purpose of testing comprehension of the questions by those interviewed. Application was made to groups of students from a specific academic course in each level, securing informed consent from each participant, where their voluntary participation was explicit and each was informed of the confidentiality of the information and its academic use. This investigation did not contemplate risks to participants and adhered to resolution 008430 of 1993 by the Ministry of Health of the Republic of Colombia.

The scale was responded directly in self-reported manner for which students were provided a booklet with readings and questions along with an independent answer sheet. This application took an average of one hour and 40 minutes, within the range of time of the international tests reviewed; the final evaluation of the participants showed that the readings achieved their attention and were easily understood; in general, the structure of the questions was clear and the language was within reach of those being evaluated.

Results

The definitions taken for the five CT domains object of measurement were the following: **Inference:**^{7-9,16} consists in drawing reasonable conclusions, like conjectures or hypotheses, based on pertinent and relevant information available, following a logical path. **Evaluation:**⁷⁻⁹ assesses the credibility, pertinence, and relevance of the information and in assessing its logical relations, whether real

or assumed. **Argumentation:**^{11,12} consists of the process of selecting and presenting coherently the results of its own reasoning. **Analysis:**⁷⁻⁹ identifies the logical force of the inferential relations between approaches or questions from reasoned justification over evidence. **Interpretation:**^{7,12,17} selects the best alternatives to express meaning.

The measurement index. The concepts of these five domains were operationalized through a methodological base matrix, identifying variables, indicators, and items, guaranteeing the validity of content in the measurement index. As a result, 8 variables, 13 indicators, and 65 assessment items were constructed to structure the scale. From this structure, a scale was constructed with five levels with which those assessed were classified from the number of items answered correctly, presenting an index from 0 to 100, with the following values: very high (80 to 100), high (60 to 79.9), medium (40 to 59.9), low (20 to 39.9), and very low (0 to 19.9). The index presents a global measurement, that is, it generates a CT measurement that evaluates the five domains; and a marginal or partial measurement that evaluates each domain individually, with the same scale from 0 to 100. In this index, the domains present equal proportional weight, given that the literature reviewed did not permit establishing weighting criteria, that is, the knowledge produced shows each domain at the same level in conceptualization and CT measurement. The instrument required the construction of single-response closed items. These items were derived from short texts or fragments of texts on different themes from everyday life, as well as statements that pose problems to be solved by those being assessed. The following illustrates two of the items constructed:

1. Based on the text, **indicate the option that shows the logical process to solve a problem:**

Option A

- a. Symbolize
- b. Find methodology
- c. Characterize
- d. Define
- e. Comprehend
- f. Find solution

Option B

- a. Characterize
- b. Find methodology
- c. Identify strategies
- d. Model
- e. Apply solution mechanisms
- f. Reuse solution mechanisms

Option C

- a. Characterize
- b. Define
- c. Find methodology
- d. Identify strategies
- e. Model
- f. Apply solution mechanisms

Option D

- a. Symbolize
- b. Characterize
- c. Model
- d. Find methodology
- e. Identify strategies
- f. Apply solution mechanisms

2. Indicate in the following question the two **statements that relate poorly or do not relate to this conclusion.**

- A. Although certain logical discernment exists, personal subjectivity prevails
- B. Judgment is a necessary instrument in examining all types of issues
- C. The essay tends to be “a deep cove on a subject not intended to be exhausted
- D. “essay” has an origin in the French “essai”, which implies a test

Each of the questions constructed on the texts selected were derived from the methodological base matrix, according to the structure of the 13 indicators that integrate the index, and arranged in a work booklet that guides their response from reading 22 texts.

To assess validity an analysis of main components was carried out, given that it is a statistical technique that permits, among other applications, evidencing from experimental data if a grouping is or is not fulfilled of variables that belong to a general element sought to being measured. Thus, this technique permits extracting from a set of variables a lower number of components (non-observable theoretical variables) that explain the higher variability observed in the data, so that when applied to the 13 indicators that group the 65 items permitted seeing if the five domains were effectively indicated. In this regard, it was

noted that indicators exist with statistically significant correlations, with the KMO test of 0.59, showing the sample’s adequacy for factor analysis and Bartlett’s test permitted rejecting the null hypothesis that the correlations matrix is the identity matrix ($p = 0.025$), confirming the pertinence of the analysis. The result permitted observing that, in fact, the information gathered pointed to five main components that managed to explain 61.3% of the data variability, being quite significant in that it confirms the hypothesis that five CT domains are being measured (Table 1).

It is worth noting that the emerging components have some differences against the theoretical grouping, which could be described in general from the processes that are involved in the following manner: *Component 1*: Deduction, induction, identification of alternatives, conjecture, identification of sequences; *Component 2*: Ordering and correspondence; *Component 3*: Conjecture and real relationship; *Component 4*: Induction and real relationship; and *Component 5*: Selection of information and assumed relationship. This emerging grouping permits verifying an aspect noted from theory in that development of skills depends on the development of others, which is why they are strongly linked, finding cases where the indicators of a skill point to the evaluation of another, without this compromising the latter’s indicators. The internal consistency analysis revealed that the 65-item scale had a Cronbach’s alpha of 0.61.

Table 1. Total variance explained from the CT model

Component	Total	% of variance	% accumulated
1	2.5	19.0	19.0
2	1.7	13.2	32.1
3	1.6	12.5	44.6
4	1.1	8.8	53.5
5	1.0	7.9	61.4

Discussion

The results obtained permitted establishing that the index constructed shows an acceptable degree of reliability within the context of Colombian formation, given that it permitted advancing on three levels against the currently existing scales; the first conceptual, upon constructing a definition of CT that involves and articulates contributions from representative and contemporary authors, which can be used to comprehend and evaluate CT in the Colombian and Latin American context.

To said extent, it should be expected for the index to permit accounting in more contextualized manner for the conditions of formation in CT, with a conceptual wager that is fed by different authors and whose interest was the articulation of the aspects proposed in the domains worked, involving elements that range from a dispositions component to a procedural component and where the characteristics of these procedures are involved as dialogical components of thought:⁷ logical,¹⁵ contextual,⁷ and pragmatic. These final five domains were ratified from the results of the expert evaluation, evidencing that the conceptual dimensions addressed have conceptual consistency, which permitted having a conception of CT consolidated to be measured, accounting for the posture assumed of combining the different theoretical elements being adequate; differentiating this proposal from the authors who assume a single conceptualization, constituting an innovation factor in the matter.

Secondly, the instrument constructed to approach these dimensions at empirical level becomes

a contribution to developing measurement instruments in the national context, given that current measurements are carried out through instruments, like the California Critical Thinking Disposition Inventory, the Critical Thinking Appraisal, Halpern's Critical Thinking, and other less known instruments; some of which specify a cost for their use and others designed in specific contexts. In that sense, this contribution, pioneer in the region, proposes the generation of a field of discussion that enables reflecting on how to measure CT in relation to the formation conditions and proposals existing in specific contexts.

A last element, considered a contribution to the discussion of the measurement, is the use of a structure of questions that directly evaluates CT skills through narrative strategies and closed indicators, which had an acceptable Cronbach's alpha coefficient based on three situations: firstly, this is a high-complexity theme for operationalization and, hence, for its measurement because there is no single agreement in scientific production with respect to its definition; secondly, the measurement scales reviewed have alpha coefficients quite similar to this,^{5,18} showing different arguments that allow seeing the high level of complexity to construct measurement scales of these types of constructs, especially in this proposal where a single author was not followed; seeking, rather, the articulation of several authors. Finally, the third argument corresponds to this being the first investigation conducted to construct a proposal to measure critical thinking in Colombia, with the characteristics exposed, and one of the few in Latin America, representing an exploratory approach, which is why we consider

acceptable the index obtained on the arguments that “during the first phases of the investigation a reliability value of 0.6 or 0.5 can be sufficient”,¹⁹ given that “the reliability value in exploratory research must be equal to or above 0.6”.²⁰

These narrative strategies promote the application of different cognitive processes related to the five CT domains worked and which account for the development and implementation of the skill. This is considered a contribution because other instruments that appeal to self-assessment, especially through Likert scales, lead to collecting perceptual information, which does not evaluate the skill, seeking information on what the subject considers it has and how the subject could apply in a possible future situation.

The immediate horizon of the scale constructed will work on broadening the versions of the items that constitute it to generate greater versatility in its application to a bigger population of students, expecting to perform follow up measurements in Social Sciences and Health with regard to the formulation of strategies to develop CT from the results obtained; seeking, thereafter, to broaden said application to other areas of knowledge that orient their profiles to CT in the formative process of higher education. Future phases of the investigation expect to model the scale to bring it to basic formation levels, especially secondary formation, given that its structure was designed only for the higher education population, specifically for the Social Sciences and Health.

It is important to consider that the main limitation for this study was its application to a small population, and although other measurement cases are registered prior to 2010 with a population below 100 cases, it is clear that this first study constitutes an exploratory approach whose principal value is to develop a conceptual and methodological proposal that discusses CT measurement in the country. Likewise, it must be considered that the work was done with population from Social Sciences and Health, which is why using the instrument in other areas of knowledge must be subjected to a validation process, before

being applied for the final measurement.

For Nursing, the theme object of investigation is vitally important within the current context of higher education because its professional and disciplinary practice demands knowledge on caring for human beings from an objective and inter-subjective vision, which should be guided by an educational process with the intention of favoring the development of cognitive and attitudinal skills for inquiry, problematization, identification, and analysis of care needs, their interpretation and argumentation, which give sense to care practices from the professional and disciplinary, both for the subject care for and for society.

References

1. Butlera HA, Dwyerb CP, Hoganb MJ, Francoc A, Rivasd SF, Saizd C, et al. The Halpern Critical Thinking Assessment and real-world outcomes: Cross-national applications. *Think. Skills Creat.* 2012; 7:112–21.
2. Pardamean B. Measuring Change in Critical Thinking Skills of Dental Students Educated in a PBL Curriculum. *J. Dent. Educ.* 2012; 76(4):443-53.
3. Atay S, Karabacak U. Care plans using concept maps and their effects on the critical thinking dispositions of nursing students. *Int. J. Nurs. Pract.* 2012; 18:233–39.
4. Valenzuela-González JR, Molina-Patlán C, Morales-Martínez GP. Competencia transversal pensamiento crítico: Su caracterización en estudiantes de una secundaria de México. *Rev. Electron. Educare* [Internet]. 2016 [cited 20 Jul 2016]; 20 (Enero-Abril): Available from: <http://www.redalyc.org/articulo.oa?id=194143011011>
5. Rivas SF, Sainz C. Validación y propiedades psicométricas de la prueba de pensamiento crítico PENCRIASAL. *Rev. Electron. Metod. Apl.* 2012; 17(1):18-34.
6. Black B. An overview of a programme of research to support the assessment of Critical Thinking. *Think. Skills Creat.* 2012; 7: 122–33.
7. Facione PA. Critical thinking: what it is and why it counts [Internet]. Hermosa Beach, CA: Measured Reasons LLC; 2015 [Cited 20 Jun 2016]. Available from: www.insightassessment.com/pdf_files/what&why2006.pdf

8. Facione NC, Facione EA, Sanchez C. A. Critical thinking disposition as a measure of competent clinical judgment: The development of the California Critical Thinking Disposition Inventory. *J. Nurs. Educ.* 1994; 33(8):345-50
9. Watson G, Glaser EM. Watson- Glaser II, Critical Thinking appraisal [Internet]. Estados Unidos: Pearson; 2012 [Cited 20 Jun 2016]. Available from: <https://www.talentlens.co.uk/assets/news-and-events/watson-glaser-user-guide-and-technical-manual.pdf>
10. Paul R, Scriven, M. Defining critical thinking: A draft statement for the National Council for Excellence in Critical Thinking [Internet]. [Cited 21 Jun 2014]. Available from: <http://www.criticalthinking.org/pages/defining-critical-thinking/410>
11. Ennis RH. Critical thinking assessment. *Theory Pract.* 1993; 32(3):179-86.
12. Halpern DF. *Thought and Knowledge: An Introduction to Critical Thinking* (5th ed). New York: Psychology Press; 2014. p.637
13. Paul R, Elder L. *Critical thinking: concepts and tools*. Santa Rosa, CA: The Foundation for Critical Thinking Press; 2009. p. 2-23.
14. Facione PA, Facione NC, Blohm SW, Giancarlo C.A. *The California Critical Thinking Skills Test: Test Manual*. CA: California Academic Press, Milbrae. 2002
15. Halpern, DF. *The Halpern critical thinking assessment: Manual*. [internet]. Vienna: Schuhfried; 2010. [Consultado en 2016 jul 15]. Disponible en: https://drive.google.com/file/d/OBzUoP_pmwy1gdEpCR05PeW9qUzA/view
16. Insight Assessment. *Health Sciences Reasoning Test (HSRT)* [Internet]. California: The California Academic Press; 2013. [Cited 15 Jul 2016]. Available from: <http://www.insightassessment.com/Products/Products-Summary/Critical-Thinking-Skills-Tests/Health-Sciences-Reasoning-Test-HSRT>
17. Saiz C, Rivas SF. Intervenir para transferir en pensamiento crítico. *Praxis.* 2008; 10(13):129-49.
18. Hatlevik IK. The theory-practice relationship: reflective skills and theoretical knowledge as key factors in bridging the gap between theory and practice in initial nursing education. *J. Adv. Nurs.* 2012; 68(4):868-77.
19. Nunnally JC. *Psychometric theory*. New York: McGraw Hill, 1967.
20. Huh J, DeLorme DE, Reid LN. Perceived third-person effects and consumer attitudes on prevetting and banning DTC advertising. *J. Consumer Affairs.* 2006;40(1):90-116.