Thinking Skills and Creativity 48 (2023) 101305

Contents lists available at ScienceDirect

Thinking Skills and Creativity

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

Development and validation of the Complex Thinking Assessment Instrument (CTAI)

Cristián Silva Pacheco[a,b,*], Carolina Iturra Herrera[a,b]

[a] PIA Cognitive Sciences, Center for Research in Cognitive Sciences, Faculty of Psychology and Institute of Humanistic Studies, Chile
[b] Universidad de Talca, Chile

ARTICLE INFO

Keywords:
Complex thinking
Creative thinking
Critical thinking
Higher order thinking
Metacognition

ABSTRACT

This study proposes the development of an instrument to measure complex thinking skills, which was applied to 256 university students to assess its psychometric properties and determine its validity on content, criterion, construct, and reliability. In this process, the instrument was submitted to the evaluation of a panel of experts, the conceptual model was verified through confirmatory factor analysis, and the criterion validity (convergent validity and discriminant validity) was determined in relation to other instruments. Finally, it is concluded that the results of the instrument validation process showed satisfactory statistical validity values. This allows establishing that the complex thinking assessment instrument (CTAI) is a psychometrically valid and reliable instrument to measure their cognitive abilities.

1. Introduction

Since the mid-twentieth century, a new vision for reality understanding began to spread, many philosophers and scientists began to question and abandon reductionist conceptions and excessive rationalization. The idea that the phenomena of reality are more complex than the abstract constructs (laws, formulas, and equations) elaborated up to that moment begins to generalize. There is the awareness that these tools would be insufficient to observe and evaluate its multidimensionality, relational aspects, uncertainty, and complexity. Especially at a scientific level, multiple theories arise in this line, among the most important we can mention: general systems theory (von Bertalanffy, 1968), complex systems theory (Levins, 1970), information theory (Ash, 1965), cybernetics (Ashby, 1956), entropy principle (Brooks & Wiley, 1988), self-organization in unbalanced systems (Prigogine & Nicolis, 1977), among others. According to Edgar Morin (1990)), this set of theories would have contributed to an epistemological paradigm shift, which he called the paradigm of complexity. This new paradigm would replace the paradigm of simplicity, which sought to bring order to the universe. The principle of simplicity (which isolates, separates; circuit breaker) would have been responsible for the fragmentation of knowledge and disciplinary hyper-specialization. Which, in turn, would have contributed to endorsing a blind and simplifying intelligence, incapable of accessing knowledge that relates to the different scopes of the phenomena as a whole and their context. In contrast, the paradigm of complexity would require complex thinking, capable of understanding the web of events, actions, interactions, feedback, determinations, and chances, that constitute our phenomenal world.

On the other hand, several authors propose the need to develop complex thinking in educational systems, especially in university education (Álvarez et al., 2019; Bustamante et al., 2018; Colina, 2020; De Melo, 2022; Estrada, 2018; Servin, 2020; Sun et al., 2022;

* Corresponding author.
E-mail address: crsilva@utalca.cl (C.S. Pacheco).

https://doi.org/10.1016/j.tsc.2023.101305
Received 5 December 2022; Received in revised form 21 April 2023; Accepted 26 April 2023
Available online 28 April 2023
1871-1871/

Yang, 2018). However, to the best of our knowledge, instruments to measure their cognitive abilities as a whole are still lacking. Having an instrument that allows measuring higher-order cognitive processes and/or abilities can be extremely beneficial, considering that assessment systems determine the practices in teaching-learning processes (López-Pastor & Sonlleva, 2019; Otero-Saborido & Vázquez -Ramos, 2019; Subheesh & Sethy, 2020). In this context, the idea arose to aiming at the developing and psychometrically validating an instrument that allows the assessment of cognitive skills of complex thinking in university students. The study hypothesis is as follows: Statistical hypothesis tests confirm the validity and reliability of the CTAI.

1.1. Complex thinking conceptualization

Nowadays, there is no unified conceptualization for complex thinking. This concept has several lines of research that are under development, the definitions of which could diverge or converge depending on the disciplinary or conceptual perspectives used to address it. Although the conceptualizations come from and/or are developed from studies related to education, the lines of research could be classified into two groups: a competency-based perspective and a cognitive perspective.

The first perspective is linked to theories of the educational field from competency –based or competency– based approaches. In this regard, we can mention the notion of complex thinking from the socio-formative approach to competencies (Tobón, 2013; Tobón & Luna-Nemecio, 2021), complex thinking as a macro-competence within the framework of Education 4.0 (Ramírez-Montoya *et al.*, 2021, 2022) and complex thinking as a meta-competence (Silva, 2020).

The second perspective, cognitivist, is aimed towards establishing the cognitive processes that compose it, based on some findings from cognitive science, neuroscience, or reasoning. Along these lines are the notions of complex thinking such as complex reasoning linked to learning progressions (Songer et al., 2009), complex thinking based on a narrative and recursive approach, to understand non-linear processes (Tsoukas & Hatch, 2001), the notion of higher order thinking by Sun et al. (2022) and the proposal of a conceptual model and operational definitions of complex thinking (Silva & Iturra, 2021). This last proposal is the one that we will address hereunder and is the one that supports the present study.

Silva and Iturra (2021) postulate that complex thinking is composed of three main cognitive processes: critical thinking, creative thinking, and metacognition. They take as a point of reference Lipman's (1997) postulates on complex thinking as higher-order

Table 1
Operational definitions of skills/indicators.

Main cognitive processes	Sub-processes or cognitive abilities
Creative thinking	*Fluency:* it is the characteristic of creativity or the facility to generate a large number of ideas or the ability to produce a large number of responses in a given field, from verbal or figurative stimuli. *Flexibility:* it is the characteristic of creativity through which the process is transformed to reach the solution of the problem or its approach. It includes a transformation, a change, a rethinking or reinterpretation. In short, it is the constant ability to produce different ideas to move from one focus to thinking and using different problem-solving strategies. *Originality:* it is the characteristic that defines the idea, process or product as something unique or different. It refers to the ability to produce novel, unconventional responses, far from the established and habitual. *Bisociativity:* ability to establish connections or relationships between apparently dissimilar concepts, coming from different areas of knowledge and/or opposites; redefining concepts, structuring information or producing new logical alternatives to obtain novel ideas.
Metacognition	*Metacognitive knowledge:* ability to: know the characteristics and complexity of all the information included in a task, in its relation to the strategies, the proposed processes, to the general, abstract or epistemological knowledge related to a task and its context; know the monitoring and control strategies in their relation to the object of knowledge involved in specific tasks; know the application and transfer of strategies and their adaptation to the demands of the tasks. *Multilogical awareness:* it is the knowledge or awareness of the use of different strategies, forms of reasoning or logical systems (coming from different areas of knowledge) in the solution of a problem or artistic production; to examine the assumptions, perspectives, and conceptual structures that lie beneath the surface of a particular problem or task, considering different points of reference. *Regulation of cognition:* ability to: control and monitoring cognitive strategies or processes, supervising retrospectively or prospectively or making metacognitive judgments that provide feedback; observe their own beliefs, prejudices or states and their involvement in the cognitive processes in the development of the task or artistic production; to self-assess or reflect on their memory and cognitive abilities or characteristics, recognizing their cognitive strengths and weaknesses; to be aware of your own feelings, emotions, intuitions or passions and how it influences your confidence or sense of satisfaction and, in turn, your performance and achievement of results.
Critical thinking	*Inference capacity:* ability to extract judgments or conclusions from certain facts, propositions or principles, whether general or particular, observed or assumed. The inference arises from a mental evaluation between different expressions that, when related as abstractions, allow a logical implication to be drawn. From some hypotheses or arguments, it is possible to infer a conclusion. *Information interpretation:* ability to interpret and understand information from a text, speech or event and judge the proposals; or to elucidate the reasons that led a subject to act as he did, using broader contexts that include social norms, customs, or others, to give meaning to the action. *Dialogic capacity:* it is the process by which knowledge is built through dialog or contradictions. It allows us to understand the interpretations of others and to look for arguments to refute, affirm or reformulate them. Communicative action and dialog are key components in the construction of knowledge and the main instrument of social transformation. This process also focuses on assessing the strengths and weaknesses of opposing perspectives. And through this, we can arrive at consensual interpretations or better argued solutions. *Ability to make judgments:* ability to formulate judgments using criteria based on the evaluation of arguments. These judgments can be universal, particular, singular, affirmative, negative, categorical, hypothetical or disjunctive depending on the problem or object in question.

Source: Silva and Iturra (2021).

thinking. They also consider the findings from cognitive sciences and neurosciences, which suggest there is evidence of a close relationship between critical and creative thinking and metacognition, implying that this link may be associated with a higher cognitive process underlying it, which could be called complex thinking. And they propose the following operational definitions for each of the sub-processes or cognitive skills of complex thinking (see Table 1):

1.2. On the instruments to assess complex thinking

A systematic search of published articles was carried out, which included a review of the following databases: Web of Science, Scopus, ScienceDirect, Scielo, and Latindex. With the following search and selection criteria: 'complex thinking assessment', 'complex thinking measurement', 'complex thinking test', and 'complex thinking instrument' in isolation and in conjunction with 'design', 'development', 'validation' and/or studies linking two or more of the above terms. Their presence was searched in: a) original articles, b) with search terms in the title, abstract, and/or keywords, c) published in English or Spanish, d) and without time limitations. Additionally, books, doctoral theses, and master's theses were searched, applying the same criteria and search terms.

As a result, 5 articles related to the design and/or validation of instruments that assess complex thinking were found. Four of these articles propose instruments, but they do not develop nor are based on a specific construct of complex thinking. Rather, they are aimed at assessing particular areas or competencies, such as mathematics (Graf & Arieli-Attali, 2015; Iversen & Larson, 2006), sciences (Ryoo & Linn, 2015), and online learning (Botha et al., 2005), focused on educational methods or models such as learning progressions, model elicitation activities (MEAs) and/or based on Bloom's taxonomy. Just one of these articles proposes a conceptualization, the proposed instrument is a Likert-type scale called Complex-21 Scale (Tobón & Luna-Nemecio, 2021). Nonetheless, attitude scales (such as Likert scales) have limitations, they only allow to observe participant's perception of an object of study or the level of inclination towards a statement, e.g.: from 'totally agree' to 'totally disagree', from 'always' to 'never', among others. Evaluations by expert judges have established that attitude scales are not suitable for measuring cognitive or intellectual abilities (see table in De Miguel, 2005, p.117). Unlike the above-mentioned scale, the present study aims to develop an instrument that includes cognitive tasks, problem-solving, and questions with open-ended answers that actually measure cognitive skills and not just the perception of these abilities.

2. Methods

2.1. Research design

The study was carried out in four phases: In the first one, the CTAI was designed and developed based on a literature review of instruments related to cognitive processes or complex thinking skills; In the second phase, the CTAI underwent an expert judges evaluation; In the third one, a pilot test was applied to adjust the CTAI to the characteristics of the selected sample; And in the last one, the CTAI was applied to a final sample of 256 university undergraduates to study its psychometric properties. The collection of total survey data took place over the course of March 2021 to November 2022.

2.1.1. Instrument design and construction

For the elaboration of the CTAI, a literature review of studies proposing tests or measurement tasks that had a conceptual affinity with the proposed theoretical construct was carried out. The instruments selected and used as a reference for each of the cognitive abilities were the following: for the creative thinking items: the Alternative Uses Task (Guilford, 1967), the Torrance Test of Creative Thinking (Torrance, 1966) and the Bi-association Task (Benedeck *et al*., 2020); for the critical thinking items: the Watson-Glaser Critical Thinking Test (Watson-Glaser, 1964) and the dialogic dimension of the Critical Thinking Questionnaire (Santiuste, 2001); for the metacognition items: the Metacognitive Calibration Self-Report (Negretti, 2017)(see Annex 1). The items or tasks were adapted and constructed considering the guidelines suggested by the International Test Commission (2017) and the Standards for Educational and Psychological Testing (AERA et al., 2014).

2.1.2. Review of items by an expert panel

The collaboration of 9 expert judges was requested, who were selected on the basis of the following criteria: having at least 10 years of professional experience in one of the areas corresponding to the CTAI dimensions (evaluation, critical thinking, creative thinking, or metacognition); and to have postgraduate studies, with at least a Master's degree. Three general assessment specialists, two creative thinking specialists, two critical thinking specialists, and two metacognition specialists participated. Two of the specialists had Master's degrees and seven of them had Doctorates. The CTAI was evaluated according to the following criteria: relevance, related to whether the item assesses a central aspect of the purpose, and it is consistent with the theoretical construct and dimensions of the CTAI. Writing, whether the item is understandable to potential users and meets with the grammatical rules of the language. Subsequently, the Lawshe (1975) content validity index was calculated. Additionally, the Aiken (1980) V coefficient was estimated. And to determine the hypothesis test (*P* value), the binomial test was applied using the IBM-SPSS software.

2.1.3. Pilot study

A pilot test was applied to a sample of 27 university students. The group size is within the guidelines specified by Hertzog (2008) and Johanson and Brooks (2010), who propose from 25 to 40 participants for pilot test samples. Students answered the CTAI through the SurveyMonkey virtual platform from their smartphones or laptops, by accessing a link sent to them via email. Following the suggestions of Muñiz and Fonseca-Pedreros (2019), this test was intended to examine the general functioning of the instrument in a sample of participants with similar characteristics to the target population of interest. Each of the items was preliminarily evaluated, both qualitatively, and quantitatively. Aspects related to the length of the CTAI and time response were observed to detect, avoid and correct possible errors.

2.1.4. Psychometric evaluation

2.1.4.1. Participants. The sample consisted of 256 university students. The sample size was calculated considering the number of items and the statistical analyses used to extract the validity of the CTAI (Pituch & Stevens, 2016). Taking as a reference the Confirmatory Factor Analysis (CFA) through the structural equation model (SEM, the calculation for this type of analysis was performed with *the A-priori Sample Size Calculator for Structural Equation Models* software (Soper, 2021), which is based on Westland's work (2010). The following estimates were considered: predicted effect size = 0.3; desired power level = 0.80; the number of latent variables = 3; number of observed variables = 11; probability level = 0.05; calculation result: $n = 123$. This value is within the range suggested for exploratory factor analysis, where it is proposed that it is preferable to apply it to samples of at least 100 subjects (Hair *et al.*, 2014; Streiner, 1994). And it is also within the range suggested by authors who recommend that each item administered should be applied to at least 5 or 10 participants (Ferrando & Anguiano, 2010; Muñiz & Fonseca-Pedrero, 2019), considering that the CTAI was composed by 24 items.

As for the characterization of the sample, the university students belonged to different degrees: 21.5% Engineering, 27% Design, 4.7% Psychology, 27.3% Nursing, and 19.5% Law. Their ages ranged from 18 to 25 years old, of which 64.8% (166) of them were female and 35.2% (90) were male. In order to select the participants, the following inclusion criteria were established: to be a regular undergraduate student at a Chilean university and to have completed most of the subjects at the corresponding level. The exclusion criteria were as follow: having some mental health or neurological pathology confirmed by a medical certificate.

2.1.4.2. Instruments. ***Complex Thinking Assessment Instrument (CTAI):*** It is the instrument proposed in this study, which measures the abilities of 3 cognitive processes: critical thinking, creative thinking, and metacognition. The version used in this stage consisted of 24 items that were determined from the results analysed in the pilot test and distributed as follows: 3 items for creative thinking, made up of cognitive tasks; 19 items for critical thinking, made up of multiple-choice questions and problem-solving; and 2 items for metacognition, made up of 2 open self-assessment questions.

COMPLEX-21 Complex Thinking Scale: It is a Likert scale proposed by Tobón and Luna-Nemecio (2021); It measures participants' perception of their levels of complex thinking skills and it is composed of five factors or dimensions: problem-solving, critical analysis, metacognition, systemic analysis, and creativity. The scale consists of 25 questions where the participant must choose one alternative among the following frequencies: never, hardly ever, sometimes, usually, always. This instrument was applied and validated on 626 Peruvian university students (Tobón & Luna-Nemecio, 2021). It was used in this study to extract the convergent validity of the CTAI.

Parental Bonding Instrument (PBI): It is a questionnaire developed by Parker et al. (1979) based on John Bowlby's attachment theory. The PBI is aimed at people over 16 years old (both genders). It measures the perception of the attitude and behavior of one or both parents in relation to the subject during their childhood and adolescence. The test consists of 25 statements composed of two scales or dimensions: Caring (12 items) and Overprotection (13 items); each item is scored using the Likert method. Thus, each response is scored in a range of 0 to 3 points, leaving the Caring scale with a maximum score of 36 points and the Overprotection scale with 39 points. This instrument was adapted in Chile by Albala and Sepúlveda (1997), and standardized and validated for the Chilean population (aged 16–64) by Melis *et al.* (2001). The paternal bonding test was used to extract the divergent or discriminant validity of the CTAI in this study.

2.1.4.3. Procedures. This study was reviewed and approved by the Scientific Ethical Committee of the University of Talca Subsequently, authorization was requested from the deans and directors of the faculties and careers that made up the sample of this study to access and apply the CTAI during some lessons. In order to have the possibility of delivering the instructions and applying it in a more controlled way. In addition, all participants gave prior informed consent to participate in this study.

The CTAI application procedure consisted of handing out the instructions and QR codes so that students could access the SurveyMonkey platform and answer from their smartphones or laptops. Along with the QR codes, they were given a sheet with drawings on it -–printed with figures– where they had to develop the task requested in item 1. Five minutes into the assessment, students were asked to photograph that sheet and hand it back to the evaluator. They then had to continue answering the rest of the questions or tasks. They had a maximum of one hour to complete the entire evaluation.

Subsequently, through an email invitation, links to two surveys were sent to the students for them to answer via the SurveyMonkey virtual platform. Each survey took an approximate of 5 min to complete. These surveys were used to obtain convergent and divergent validity, Complex-21 and PBI, respectively.

2.1.4.4. Data analysis. Once the responses had been collected and tabulated, they were transferred from the SurveyMonkey virtual platform to an external database. For the analysis of the qualitative data, referring to the answers to the open-ended questions, the NVivo PRO software was used, which allowed categorizing and/or coding the answers and extracting the frequencies of the concepts, which was very useful in the activities related to some dimensions associated to creative thinking. Subsequently, scores –numerical values– were assigned to all the items, according to the correction and scoring guidelines proposed for the CTAI.

For the analysis of the psychometric properties of the instrument, the difficulty index, discrimination index, and/or the homogeneity of the items were extracted. For the reliability analysis, the software proposed by Zhang and Yuan (2016) was used to calculate the Cronbach's Alpha and McDonald's Omega coefficients. In parallel, McDonald's Omega was calculated using the IBM-SPSS software, this estimate is based on the factor loading of a single-factor forced maximum likelihood factor analysis using the SPSS Embedded Factor procedure (Hayes & Coutts, 2020).

For construct validity, Confirmatory Factor Analysis (CFA) was applied through the AMOS-SPSS software. The CFA made possible to contrast the proposed conceptual model based on the correlations between the latent factor and the variability of the observed variables. The adequacy of CFA was tested using the significance calculated from the chi-square test with the Satorra-Bentler correction (S-B $\chi 2$) (Bentler, 2004; Satorra & Bentler, 1994). The adequacy of the suggested models was verified with additional coefficients, such as the discrepancy between $\chi 2$ and degrees of freedom (CMIN/DF), with values below 5 being considered acceptable (Byrne, 2009; Carmines & McIver, 1981). The Goodness-of-Fit Index (GFI), the Adjusted Goodness of Fit Index (AGFI), the Comparative Fit Index (CFI), the Incremental Fit Index (IFI), the Normalized Fit Index (NFI), and the Non-Normalized Fit Index (NNFI or TLI) were also tested; for these indicators, values >0,90 are considered a good fit (Hu & Bentler, 1999; MacCallum & Austin, 2000; Marsh & Hau, 1996; Schumacker & Lomax, 2010). The Root Mean Square Error of Approximation (RMSEA) was also tested, whose scores had to be <0,08 to be considered a good fit (Browne & Cudeck, 1993; Hox, 2021; Kline, 2015).

For criterion validity, the results obtained from the CTAI were compared and analysed with other instruments for convergent and discriminant validity (Hogan, 2013), COMPLEX-21 and PBI, respectively. Both analyses were calculated using Spearman's correlation coefficient using the IBM-SPSS software.

3. Results

3.1. Content validity

The results extracted from the different indices analysed show excellent validity as a whole, considering that the minimum value for the number of panellists (9) should be >0.78 (Lawshe, 1975). Lawshe's content validity index, overall CVI = 0.97; overall Aiken's V coefficient = 0.93 (relevance = 0.91 and writing = 0.95). In the Binomial Test, agreements (p) and disagreements (q) were considered to be equal to or greater than 0.5 ($p = q = 0.5$). The result of the analysis showed exact (bilateral) significance values below 0.05 in almost all items (see Table 2). Only 2 items were above the critical significance value, items 1 (sig. = 0.27) and 22 (sig. = 0.09). In item 1, the disagreement of the expert evaluators was focused on the wording (sig. = 0.51), so it was decided to improve this aspect. In item 22, the disagreement was focused on relevance (sig. = 0.18), so it was also decided to improve it according to the suggestions provided by some expert evaluators. Therefore, the 27 items initially proposed were maintained at this stage. The average significance obtained was sig. = 0.02, being less than 0.05, which means that the null hypothesis (H_0) is rejected and the alternative hypothesis (H_A) is

Table 2
Summary of results of content validity indices and binomial test.

Ítems	Lawshe's CVI		Aiken's V		Binomial Test	
	Relevance	Writing	Relevance	Writing	Relevance	Writing
1	0.89	0.69	0.89	0.70	0.04	0.51
2	1.00	0.89	0.96	0.93	0.00	0.04
3	1.00	0.89	0.96	0.93	0.00	0.04
4	1.00	1.00	0.93	0.85	0.00	0.00
5	1.00	1.00	0.93	0.96	0.00	0.00
6	1.00	1.00	0.93	0.96	0.00	0.00
7	1.00	1.00	0.93	0.96	0.00	0.00
8	1.00	10.89	0.85	0.85	0.00	0.04
9	1.00	1.00	0.89	0.96	0.00	0.00
10	1.00	1.00	0.89	0.96	0.00	0.00
11	1.00	1.00	0.89	0.96	0.00	0.00
12	1.00	1.00	0.93	0.96	0.00	0.00
13	1.00	1.00	0.89	0.96	0.00	0.00
14	1.00	1.00	0.89	0.96	0.00	0.00
15	1.00	1.00	0.89	0.96	0.00	0.00
16	0.89	1.00	0.89	1.00	0.04	0.00
17	0.89	1.00	0.89	1.00	0.04	0.00
18	1.00	1.00	0.93	1.00	0.00	0.00
19	1.00	1.00	0.93	1.00	0.00	0.00
20	1.00	1.00	0.93	1.00	0.00	0.00
21	0.89	1.00	0.85	1.00	0.04	0.00
22	0.78	1.00	0.81	1.00	0.18	0.00
23	1.00	1.00	0.96	0.96	0.00	0.00
24	1.00	1.00	0.96	0.96	0.00	0.00
25	0.89	1.00	0.89	0.96	0.00	0.00
26	1.00	0.89	0.89	0.89	0.00	0.04
27	1.00	0.89	0.93	0.89	0.00	0.04
Averages =	0.97	0.97	0.91	0.95	0.01	0.02

accepted. That is, overall the content validity of the instrument is accepted.

3.2. Reliability

Once the score had been assigned to each of the answers given by the students, the first step was to analyze the difficulty and discrimination indices. Once this analysis had been carried out, it was decided to eliminate 5 items since they showed low discrimination rates. The items eliminated were: n° 4 ($D = 0{,}20$); n° 8 ($D = 0{,}19$); n° 13 ($D = 0{,}16$); n° 14 ($D = 0{,}07$); n° 18 ($D = 0{,}15$). As a result, the CTAI was composed by 19 items (see final version of the instrument in the annex). Based on these 19 items, the following statistical analyses were conducted.

For the reliability analysis as well as for the subsequent analyzes (CFA, convergent and divergent validity) the skills of each of the cognitive processes or factors that make up complex thinking were considered as observed variables, as has been done in the analysis of instruments similar, for example, in studies of instruments that measure creative thinking (Kim, 2006; Said-Metwaly, Kindt & Van den Noortgate, 2020). The scores of the 19 items were redistributed across 11 observed variables.

To determine reliability, the software proposed by Zhang and Yuan (2016) was used because it provides robust procedures for estimating coefficients in samples with non-normal distributions, as is the case in this study, whose normality tests showed asymptotic significance values below 0.05 (see Table 3). The reliability coefficients were as follows: Cronbach's Alpha $\alpha = 0.691$; and McDonald's Omega $\omega = 0.813$. The McDonald's Omega ($\omega = 0.849$) was also calculated with IBM-SPSS software. This latter result is consistent with the omega coefficient calculated with the previous software and is within the acceptable range (0.70 – 0.90) (Campo-Arias & Oviedo, 2008; Zhang & Yuan, 2016) and is higher than the range proposed for the early stages of a research (0.50–0.60) (Nunnally, 1967; Streiner, 2003).

3.3. Construct validity

Numerous authors have considered applying Exploratory Factor Analysis (EFA) and CFA to determine to construct validity; however, other authors consider that this procedure may be redundant, advising that an exploratory approach (from a perspective of assumed total lack of knowledge) may lead to random and unexpected results (Pérez-Gil *et al.*, 2000). The application of AFE was not considered because a pre-defined model or construct of complex thinking was already available (Silva & Iturra, 2021). In addition, it should be considered that the items were adapted and/or elaborated taking as reference tasks or problems extracted from widely known instruments and validated instruments in multiple psychometric studies.

3.3.1. Confirmatory factor analysis (CFA)

The chi-square value was not considered for the CFA since it is extremely sensitive to sample size (>200) (Byrne, 1998). Some authors propose that with large samples it is almost certain that the test will be significant and the factorial model will be rejected. On the contrary, with small samples, the model will almost always be accepted, even if it fits poorly. Therefore, fit indices less dependent on sample size have been formulated, such as Goodness-of-Fit Index, which are a function of the chi-square and degrees of freedom (Hox, 2021).

The models analysed in Table 4 only differ in the number of observed variables, since the latent variables (creative thinking, critical thinking, and metacognition) were pre-defined from the proposed theoretical model. Three error covariances (e1-e4; e2-e4; e5-e7) were applied to the initially proposed model (model 1a) following the modification rates suggested by AMOS, resulting in model 1b. Two additional models (models 2 and 3, from which the observable variables with the lowest factor loadings were extracted) were also for comparison with the previous ones in order to determine the best fitting model. According to the values extracted from the different indices, it is observed that Model 1b (see Figure 1) is the one with the best fit, since all the values fall within the expected ranges, and even the CFI, GFI, and IFI values are equal to 0.95, indicating a very good fit of the model (Marsh & Hau, 1996). Additionally, the factor values indicate that all the observed variables have loadings greater than 0.3 with their respective factor (Howard, 2016). Therefore, these results would allow to establish on a preliminary basis that the proposed theoretical model for complex thinking would be confirmed, at least from the results extracted from the analysis of this instrument.

Table 3
Kolmogorov-Smirnov test for the variables creative thinking, critical thinking, metacognition, and complex thinking.

		Creative thinking	Critical thinking	Metacognition	Complex thinking
N		256	256	256	256
Normal parameters[a,b]	Mean	47.95	27.93	11.70	163.46
	Dev. St.	27.296	9.599	4.444	63.382
Maximum extreme differences	Absolute	0.124	0.067	0.372	0.098
	Positive	0.124	0.067	0.372	0.098
	Negative	−0.105	−0.044	−0.273	−0.073
Test statistic		0.124	0.067	0.372	0.098
Asymptotic sig. (two-tailed)		0.000[c]	0.007[c]	0.000[c]	0.000[c]

[a] The test distribution is normal.
[b] Calculated from data.
[c] Lilliefors' significance correction.

Table 4
Adjustment indices for the CFA of the proposed models.

Adjustment indices	Expected values	Model 1a	Model 1b (error covariances: e1-e4; e2-e4; e9-e11)	Model 2 (without Bisociativity)	Model 3 (without bisociativity and multilogical awareness)
CMIN/DF	<3/5	3.50	2.14	2.49	2.61
GFI	0.90 – 1	0.91	0.95	0.94	0.95
AGFI	0.90 – 1	0.86	0.91	0.90	0.908
RMSEA	<0.08	0.10	0.07	0.08	0.08
CFI	0.90 – 1	0.88	0.95	0.94	0.95
IFI	0.90 – 1	0.88	0.95	0.94	0.95
NFI	0.90 – 1	0.84	0.91	0.90	0.92
NNFI/TLI	0.90 – 1	0.84	0.93	0.92	0.93

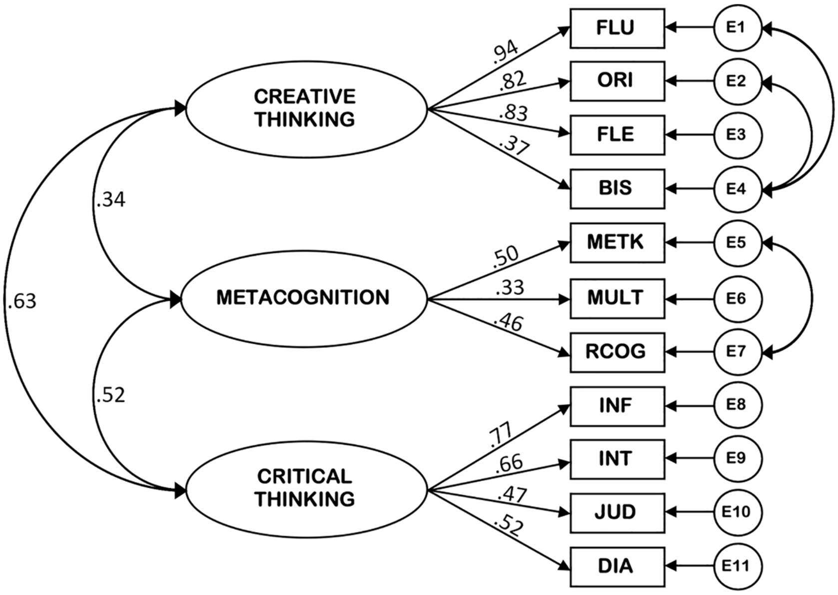

Fig. 1. Model (1b) of the complex thinking construct.

FLU Fluency, ORI Originality, FLE Flexibility, BIS Bisociativity, METK Metacognitive knowledge, MULT Multilogic awareness, RCOG Regulation of cognition, INF Inference capacity, INT Information interpretation, JUD Ability to make judgments, DIA Dialogic capacity.

3.4. *Criterion validity*

3.4.1. *Convergent validity*

This analysis was determined using the Spearman-Rho coefficient, since the sample values were not normally distributed (Gauthier, 2001; Puth *et al.*, 2015), which could be established through the Kolmogorov-Smirnov Normality Tests. (Complex-21 = 0.200, CTAI = 0.002, and difference = 0.024). Spearman's Rho correlation coefficient calculated from the total scores obtained from CTAI and Complex-21 was 0.309, which corresponds to a positive and moderate relationship (Dancey & Reidy, 2004), and higher than the critical value calculated for this sample. (0.261) (Zar, 1984). The calculated *P* value was 0.050, being significant because it was equal to 0.05, this means that H_0 is rejected and H_A is accepted, indicating that there is a relationship between the overall results of both instruments (see Table 5). Therefore, these results allow us to determine the convergent validity referred to external criteria.

Table 5
Results of the Spearman-Rho correlation coefficient between complex-21 and CTAI.

			Complex-21	CTAI
Spearman-Rho	Complex-21	Correlation coefficient	1.000	0.309*
		Sig. (two-tailed)	.	0.050
	CTAI	Correlation coefficient	0.309*	1.000
		Sig. (two-tailed)	0.050	.

* Correlation is significant at the 0.05 level (two-tailed).

3.4.2. Divergent validity

For the analysis of divergent validity, the total scores obtained on the CTAI and PBI instruments were correlated. This analysis was also determined with the Spearman-Rho correlation coefficient after applying the Kolmogorov-Smirnov Normality Test (PBI = 0.200, CTAI = 0.001, and difference = 0.001). The calculated Spearman's Rho coefficient was 0.135, which corresponds to a null or insignificant relationship (Dancey & Reidy, 2004), and lower than the calculated critical value (0.254) for this sample size (256) (Zar, 1984). The calculated *P* value was 0.387 (not significant because it is greater than 0.05), which means that the H_0 is confirmed and the H_A is rejected. Therefore, it can be determined that there is no relationship between the overall results of both instruments (see Table 6). This result allows us to determine that there is divergent validity referring to external criteria.

4. Discussion and conclusions

The objective proposed for this study was satisfactorily fulfilled. The hypothesis tests confirm the validity of the overall CTAI. In content validity, the *P* value was calculated using the Binomial Test, whose resulting significance value was 0.02 (being less than 0.05), which means that the content validity of the instrument, in general, was accepted in general, whose validity indices were excellent (Lawshe's ICV = 0.97; Aiken's $V = 0.93$).

In construct validity, the *P* value was calculated using the CFA Goodness-of-Fit Indices, the chi-square value was not considered for this analysis (which is the indicator usually used to calculate the *P* value in CFA) as it is extremely sensitive to sample size (>200) (Byrne, 1998). The values of the extracted goodness indices (CMIN/DF = 2.14; GFI = 0.95; AGFI = 0.91; RMSEA = 0.07; CFI = 0.95; IFI = 0.95; NFI = 0.91; NNFI = 0.93) of the proposed model (Model 1b) fall within the expected ranges. And even, the CFI, GFI, and IFI values were values equal to 0.95, indicating a good fit for the model (Marsh & Hau, 1996). Therefore, these results allowed us to accept the construct validity of the instrument.

For criterion validity, the *P* value was calculated using the Spearman-Rho correlation test to determine both convergent and divergent validity. In convergent validity, the CTAI and Complex-21 were analysed, whose correlation coefficient was 0.309, corresponding to a positive and moderate relationship (Dancey & Reidy, 2004). The calculated *P* value (0.050) allowed to determine convergent validity. In divergent validity, the CTAI and PBI were analysed, whose correlation coefficient was 0.135, which corresponds to a null or insignificant relationship (Dancey & Reidy, 2004), and lower than the critical value calculated for this sample (0.254) (Czar, 1984). The calculated *P* value (0.387) was higher than 0.05, indicating that there is divergent validity. Both results together allow us to establish that the criterion validity is acceptable.

On the other hand, the reliability indices of the instrument were adequate: $\alpha = 0.691$ y $\omega = 0.813$ y 0.849. It is important to highlight that a higher value was obtained in the calculation of omega (ω), considering that some authors claim that this coefficient provides a more accurate approximation of reliability than Cronbach's Alpha (Hayes & Coutts, 2020; Peters, 2014). It is argued that Cronbach's alpha coefficient is affected by the number of items, the number of response alternatives, and the proportion of variance in the test (Domínguez-Lara & Merino-Soto, 2015). It also has the disadvantage of working with continuous variables, something that hardly occurs in social sciences, which underestimates reliability (Elosua & Zumbo, 2008). On the other hand, the omega coefficient does not depend on the number of items and it is calculated based on factor loadings (Gerbing & Anderson, 1988; Ventura-León & Caycho-Rodríguez, 2017). From these results, it can be established that the CTAI is a psychometrically valid and reliable instrument for measuring the cognitive skills of complex thinking.

Although the results of the research were positive, some limitations need to be acknowledged. One of these is related to the representativeness of the sample since it was limited to students belonging to one university. The methodological design established the inclusion of participants from different Chilean universities and programmes. Requests for authorization were sent to the authorities of different universities, but were rejected. The sample included participants whose universities agreed to authorize the application of the instruments. However, since the sample was estimated for the validation of an instrument, it is suggested that convenience sampling is more suitable for these studies than probability sampling (Viswanathan, 2005). The size was calculated in relation to the number of items and the statistical analyses applied (24 items and CFA), as suggested by some authors (Pituch & Stevens, 2016; Westland, 2010). Though, for the confirmation of the conceptual model, future research considering larger and more heterogeneous samples is needed.

Regarding future research, it would be necessary to apply the CTAI to students of other age groups (high school, postgraduate, etc.) in order to re-evaluate and/or confirm its validity and, at the same time, to see if the model proposed by CFA is confirmed. Another research possibility would be to measure complex thinking skills in students from different degree programmes in order to compare and characterize their levels of development of these skills. This would allow, in turn, to characterize each of the proposed cognitive skills with the purpose of improving the CTAI and/or the conceptual model.

Declaration of Competing Interest

The authors declare they have no conflicts of interest.

Data availability

I have shared the link to my data at the attach files step.

Table 6
Results of the Spearman-Rho correlation coefficient between PBI y CTAI.

			PBI	CTAI
Spearman-Rho	PBI	Correlation coefficient	1.000	0.135
		Sig. (two-tailed)	.	0.387
	CTAI	Correlation coefficient	0.135	1.000
		Sig. (two-tailed)	0.387	.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tsc.2023.101305.

References

Aiken, L. (1980). Content validity and reliability of single items or questionnaires. *Educational and Psychological Measurement, 40*(4), 955–995.

Albala, J., & Sepúlveda P. (1997). *Adaptación del cuestionario parental bonding instrument (PBI) de Parker, Tupling y Brown, para la población consultante mayor de dieciséis años del gran Santiago* (tesis de grado). Universidad Central, Santiago de Chile. https://doi.org/10.4067/S0717-92272001000200005.

Álvarez, L., Pérez, C., & Lara, R. (2019). Las tareas problematizadoras como propuesta para el desarrollo del pensamiento complejo. *Revista Metropolitana de Ciencias Aplicadas, 2*(2), 75–83.

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing.* American Educational Research Association.

Ash, R. (1965). *Information theory*. Interscience Pub.

Ashby W. (1956). *Introduction to cybernetics*. Methuen.

Benedek, M., Jurisch, J., Koschutnig, K., Fink, A., & Beaty, R. (2020). Elements of creative thought: Investigating the cognitive and neural correlates of association and bi-association processes. *NeuroImage, 210*, Article 116586.

Bentler, P. (2004). *EQS 6 structural equations modeling software. Multivariate software*. Encino, CA.

Botha, J., van der Westhuizen, D., & De Swardt, E. (2005). Towards appropriate methodologies to research interactive learning: Using a design experiment to assess a learning program for complex thinking development. *International Journal of Education and Development using ICT, 1*(2), 105–117. https://www.learntechlib.org/p/42228/.

Brooks, D., & Wiley, E. (1988). *Evolution as entropy: Toward a unified theory of biology*. University of Chicago Press.

Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen, & J. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.

Bustamante, L., Ayllón, S., & Escanés, G. (2018). Abordando la trayectoria universitaria desde el pensamiento complejo. *Praxis Educativa, 22*(3), 64–70.

Byrne, B. (1998). *Structural equation modeling with LISREL, PRELIMS, and SIMPLIS: Basic concepts, applications, and programming.* Lawrence Erlbaum Associates Publishers.

Byrne, B. (2009). *Structural equation modeling with Amos: Basic concepts, applications, and programming.* Routledge.

Campo-Arias, A., & Oviedo, H. (2008). Propiedades psicométricas de una escala: La consistencia interna. *Rev Salud Pública, 10*(5), 831–839.

Carmines, E., & McIver, J. (1981). Analyzing models with unobserved variables: Analysis of covariance structures. In G. Bohrnstedt, & E. Borgatta (Eds.), *Social measurement: Current issues* (pp. 65–115). Sage Publications, Inc.

Colina, A. (2020). La educación superior desde la visión del pensamiento complejo. *Revista Científica Ecociencia, 7*, 1–18.

Dancey, C., & Reidy, J. (2004). *Statistics without maths for psychology: Using SPSS for windows*. Prentice Hall.

De Melo, A. (2022). Conceptualizing and designing (Co)augmented intelligence(s) enacting complex thinking: Morphological constraints, challenges, and implications. *Proceedings, 81*(1), 85. https://doi.org/10.3390/proceedings2022081085

De Miguel, M. (2005). *Modalidades de enseñanza centradas en el desarrollo de competencias. Orientaciones para promover el cambio metodológico en el espacio europeo de educación superior*. Ediciones Universidad de Oviedo: Asturias. https://www2.ulpgc.es/hege/almacen/download/42/42376/modalidades_ensenanza_competencias_mario_miguel2_documento.pdf.

Domínguez-Lara, S., & Merino-Soto, C. (2015b). Sobre el reporte de confiabilidad del Clarp-TDAH, de Salamanca (2010). *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud, 13*(2), 1316–1317.

Elosua, P., & Zumbo, B. (2008). Coeficientes de fiabilidad para escalas de respuesta categórica ordenada. *Psicothema, 20*(4), 896–901.

Estrada, A. (2018). El pensamiento complejo y el desarrollo de competencias transdisciplinares en la formación profesional. *Revista Científica RUNAE, 3*(2), 177–193.

Ferrando, P., & Anguiano, C. (2010). El análisis factorial como técnica de investigación en Psicología. *Papeles del Psicólogo, 31*, 18–33.

Gauthier, T. (2001). Detecting trends using Spearman's rank correlation coefficient. *Environmental Forensics, 2*(4), 359–362.

Gerbing, D., & Anderson, J. (1988). An update paradigm for scale development incorporating unidimentionality and its assessment. *Journal of Marketing Research, 25*(2), 186–192.

Graf, E., & Arieli-Attali, M. (2015). Designing and developing assessments of complex thinking in mathematics for the middle grades. *Theory Into Practice, 54*(3), 195–202. 10.1080/00405841.2015.1044365.

Guilford, J. (1967). *The nature of human intelligence*. New York: McGraw-Hill.

Hair, J., Black, W., Babin, B., & Anderson, R. (2014). *Multivariate data analysis*. Edinburgh Gate: Pearson Education Limited.

Hayes, A., & Coutts, J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. But…. *Communication Methods and Measures, 14*(1), 1–24. 10.1080/19312458.2020.1718629.

Hertzog, M. (2008). Considerations in determining sample size for pilot studies. *Research in Nursing & Health, 31*, 180–191.

Hogan, T. (2013). *Psychological testing: A practical introduction*. Wiley.

Howard, M. (2016). A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human–Computer Interaction, 32*, 51–62.

Hox, J. (2021). Confirmatory factor analysis. In J. Barnes, & D. Forde (Eds.), *The encyclopedia of research methods in criminology and criminal justice* (pp. 830–832). John Wiley & Sons.

Hu, L., & Bentler, P. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55.

International Test Commission. (2017). *The ITC guidelines for translating and adapting tests* (2nd edition). [. www.InTestCom.org.].

Iversen, S., & Larson, C. (2006). Simple thinking using complex math vs. complex thinking using simple math—A study using model-eliciting activities to compare students' abilities in standardized tests to their modeling abilities. *Zentralblatt für Didaktik der Mathematik, 38*, 281–292. https://doi.org/10.1007/BF02652811

Johanson, G., & Brooks, G. (2010). Initial scale development: Sample size for pilot studies. *Educational and Psychological Measurement, 70*(3), 394–400.

Kim, K. (2006). Is creativity unidimensional or multidimensional? Analyzes of the Torrance tests of creative thinking. *Creativity Research Journal, 18*, 251–259.

Kline, R. (2015). *Principles and practice of structural equation modeling*. Guilford publications.
Lawshe, C. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x
Levins, R. (1970). Complex systems. In C. H. Waddington (Ed.), *Organization stability & process toward a theoretical biology* (pp. 73–88). Routledge.
Lipman, M. (1997). *Pensamiento complejo y educación*. Madrid: Ediciones de la Torre.
López-Pastor, V., & Sonlleva, M. (2019). Evaluación formativa y compartida en educación. *Revista Iberoamericana de Evaluación Educativa, 12*(1), 5–9.
MacCallum, R., & Austin, J. (2000). Applications of structural equation modeling in psy- chological research. *Annual Review of Psychology, 51*, 201–226.
Marsh, H., & Hau, K. (1996). Assessing goodness of fit: Is parsimony always desirable? *Journal of Experimental Education, 64*, 364–390.
Melis, F., Dávila, M., Ormeño, V., Vera, V., Greppi, C., & Gloger, S. (2001). Estandarización del P.B.I. Parental bonding instrument, versión adaptada a la población entre 16 y 64 años del Gran Santiago. *Revista Chilena de Neuropsiquiatría, 39*, 132–139. https://doi.org/10.4067/S0717-92272001000200005.
Morin, E. (1990). *Introducción al pensamiento complejo*. Barcelona: Gedisa.
Muñiz, J., & Fonseca-Pedreros, E. (2019). Diez pasos para la construcción de un test. *Psicothema, 31*(1), 7–16. 10.7334/psicothema2018.291.
Negretti, R. (2017). Calibrating Genre: Metacognitive Judgments and Rhetorical Effectiveness in Academic Writing by L2 Graduate Students. *Applied Linguistics, 38*(4), 512–539. https://doi.org/10.1093/applin/amv051
Nunnally, J. (1967). *Psychometric theory*. McGraw-Hill.
Otero-Saborido, F., & Vázquez-Ramos, F. (2019). La evaluación educativa en el currículo LOMCE en primaria: Análisis de los currículos autonómicos en educación física. *Revista Iberoamericana de Evaluación Educativa, 12*. 10.15366/riee2019.12.1.003.
Parker, G., Tupling, H., & Brown, L. (1979). A parental bonding instrument. *British Journal of Medical Psychology, 52*, 1–10. https://doi.org/10.1111/j.2044-8341.1979.tb02487.x
Pérez-Gil, J., Chacón, S., & Moreno, R. (2000). Validez de constructo: El uso de análisis factorial exploratorio-confirmatorio para obtener evidencias de validez. *Psicothema, 12*(2), 442–446.
Peters, G. (2014). The alpha and the omega reliability and validity. *The European Health Psychologist, 16*(2), 56–69.
Pituch, K., & Stevens, J. (2016). *Applied multivariate statistics for the social sciences. Analyzes with SAS and IBM'S SPSS*. New York: Routledge.
Prigogine, I., & Nicolis, G. (1977). *Self-organization in non-equilibrium systems*. Wiley.
Puth, M., Neuhäuser, M., & Ruxton, G. (2015). Effective use of Spearman's and Kendall's correlation coefficients for association between two measured traits. *Animal Behaviour, 102*, 77–84.
Ramírez-Montoya, M., Álvarez-Icaza, I., Sanabria-Zepeda, J., López-Caudana, E., Alonso, P., & Miranda, J. (2021). Scaling complex thinking for everyone through open science: A conceptual and methodological framework. In *Proceedings of the 9th International Conference on Technological Ecosystems for Enhancing Multiculturality (TEEM 2021*. University of Barcelona.
Ramírez-Montoya, M., Castillo-Martínez, I., Sanabria-Z, J., & Miranda, J. (2022). Complex thinking in the framework of education 4.0 and open innovation—A systematic literature review. *Journal of Open Innovation: Technology, Market, and Complexity, 8*(1). https://doi.org/10.3390/joitmc8010004
Ryoo, K., & Linn, M. (2015). Designing and validating assessments of complex thinking in science. *Theory Into Practice, 54*(3), 238–254. https://doi.org/10.1080/00405841.2015.1044374
Said-Metwaly, S., Kyndt, E., & Van den Noortgate, W. (2020). The factor structure of the verbal Torrance test of creative thinking in an Arabic context: Classical test theory and multidimensional item response theory analyzes. *Thinking Skills and Creativity, 35*. doi.org/10.1016/j.tsc.2019.100609.
Santiuste, B., Ayala, B., García, E., González, J., Rossignoli, J., & Toledo, E. (2001). *El pensamiento crítico en la práctica educativa*. Madrid: Fugaz Ediciones.
Satorra, A., & Bentler, P. (1994). Corrections to test statistics and standard errors in covariance structure analysis. In A. von Eye, & C. Clogg (Eds.), *Latent variables analysis: Applications for developmental research* (pp. 399–419). Sage.
Schumacker, R., & Lomax, R. (2010). *A beginners guide to structural equation modeling*. Routledge.
Servín, D. (2020). Diseño formativo interprofesional: Una estrategia para desarrollar el pensamiento complejo en estudiantes de ciencias de la salud. *Fundación Educación Médica (FEM), 23*(1), 39–44.
Silva, C. (2020). Art education for the development of complex thinking metacompetence: A theoretical approach. *International Journal Art & Design Education (IJADE), 39*(1), 242–254. 10.1111/jade.12261.
Silva, C., & Iturra, C. (2021). A conceptual proposal and operational definitions of the cognitive processes of complex thinking. *Thinking Skills and Creativity, 39*. https://doi.org/10.1016/j.tsc.2021.100794
Songer, N., Kelcey, B., & Wenk, A. (2009). How and when does complex reasoning occur? Empirically driven development of a learning progression focused on complex reasoning about biodiversity. *Journal of Research in Science Teaching, 46*(6), 610–631.
Soper, D. (2021). A-priori sample size calculator for structural equation models [Software]. Available from https://www.danielsoper.com/statcalc.
Streiner, D. (1994). Figuring out factors: The use and misuse of factor analysis. *The Canadian Journal of Psychiatry, 39*(3), 135–140.
Streiner, D. (2003). Starting at the beginning: An introduction to coefficient alpha and internal consistency. *Journal of Personality Assessment, 80*, 99–103.
Subheesh, N., & Sethy, S. (2020). Learning through assessment and feedback practices: A critical review of engineering education settings. *EURASIA Journal of Mathematics, Science and Technology Education, 16*(3). https://www.ejmste.com/download/learning-through-assessment-and-feedback-practices-a-critical-review-of-engineering-education-7786.pdf.
Sun, H., Xie, Y., & Lavonen, J. (2022). Exploring the structure of students' scientific higher order thinking in science education. *Thinking Skills and Creativity, 43*. https://doi.org/10.1016/j.tsc.2022.100999
Tobón, S. (2013). *Formación integral y competencias. Pensamiento complejo, currículo, didáctica y evaluación* (4ta. Ed.). Bogotá: ECOE.
Tobón, S., & Luna-Nemecio, J. (2021). Complex thinking and sustainable social development: Validity and reliability of the COMPLEX-21 scale. *Sustainability, 13*, 6591. https://doi.org/10.3390/su13126591
Torrance, E. (1966). *Torrance tests of creative thinking*. Lexington, MA: Personnel Press.
Tsoukas, H., & Hatch, M. (2001). Complex thinking, complex practice: The case for a narrative approach to organizational complexity. *Human Relations, 54*(8), 979–1013. https://www.researchgate.net/publication/228864629.
Ventura-León, J., & Caycho-Rodríguez, T. (2017). El coeficiente omega: Un método alternativo para la estimación de la confiabilidad. *Revista Latinoamericana de Ciencias Sociales, Niñez y Juventud, 15*(1), 625–627.
von Bertalanffy, L. (1968). General system theory: Foundations, development, applications. George Braziller.
Viswanathan, M. (2005). *Measurement error and research design*. Sage Publishing.
Watson, G., & Glaser, E. (1964). *Watson-Glaser critical thinking appraisal manual*. Psychological Corporation.
Westland, J. (2010). Lower bounds on sample size in structural equation modeling. *Electronic Commerce Research and Applications, 9*, 476–487. https://doi.org/10.1016/j.elerap.2010.07.003
Yang, H. (2018). Complex thinking and responses to the question of why female graduates fail to practice medicine. *Medical Education, 52*(7), 687–689.
Zar, J. (1984). *Biostatistical analysis*. Prentice Hall.
Zhang, Z., & Yuan, K. (2016). Robust coefficients alpha and omega and confidence intervals with outlying observations and missing data: Methods and software. *Educational and Psychological Measurement, 76*(3), 387–411.