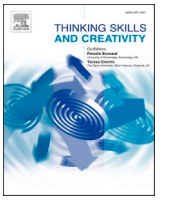
Thinking Skills and Creativity 46 (2022) 101169

Contents lists available at ScienceDirect

Thinking Skills and Creativity

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



Design and validate the Employer-Employee-Supported Critical Thinking Disposition Inventory (2ES-CTDI) for undergraduates

Yong Liu [a,*], Attila Pásztor [b,c,d]

[a] Doctoral School of Education, University of Szeged, Hungary
[b] Institute of Education, University of Szeged, Hungary
[c] MTA-SZTE Research Group on the Development of Competencies, Hungary
[d] MTA-SZTE Digital Learning Technologies Research Group, Hungary

ARTICLE INFO

Keywords:
Validation
Critical thinking disposition
Inventory
Undergraduates

ABSTRACT

Even though the framework of critical thinking (CT) comprises both skills and disposition, many studies have placed more emphasis on evaluating CT skills than disposition, probably because of the paucity of disposition instruments. On the contrary, CT disposition (CTD) is more frequently mentioned in the workforce than CT skills. Furthermore, prevalent disposition instruments are mostly based on the literature and suggestions from educators, rather than employers' or employees' perspectives. The purpose of this study is to design and validate the Employer-Employee-Supported Critical Thinking Disposition Inventory (2ES-CTDI) for college students to evaluate CTD. Intelligence accumulation (IA) was employed to synthesize the ideas from the literature, 25 employers, 43 employees, and a seven-person expert group. The interpretative structural model (ISM) was used to build the potential inventory. Validation data were collected from 661 Chinese undergraduate students (328 males, 49.6 %, Mean = 19.22, SD = 1.20 and 333 females, 50.4 %, Mean = 19.91, SD = 1.23) in 2021. Explorative factor analysis (EFA) was conducted for consistency between the conceptual framework and data. Confirmative factor analysis (CFA) and the common method variance (CMV) were adopted to check the model fit and structural validity. Partial least squares structural equation modeling (PLS-SEM) was used to identify the mediating effects among CTD components, and item response theory (IRT) was utilized to evaluate the test endorsement and item discrimination. Finally, a three-factor seven-point CT disposition inventory was confirmed. This newly-developed scale can be used to assess college students' CTD as expected by the job market. While the inventory has been validated in this study, its contents are limited to intelligence obtained from the participants from some Chinese job markets. Further validations need to be carried out and implications for wider applications need to be considered.

1. Introduction

The outbreak of the COVID-19 pandemic has caused a sharp decline in the global economy. This has been followed by increasing conflicts and slanders in politics, economy, and culture. Moreover, younger generations have been trapped by fake news. "The world needs critical thinkers now more than ever before, and not only is the sheer amount of information available to us undeniably overwhelming, but the accuracy of the information is suspect" (Butler & Halpern, 2020, p. 152). Accordingly, assessing critical

* Corresponding author at: Doctoral School of Education, University of Szeged, Petőfi S. sgt. 32-34, 6722 Szeged, Hungary.
E-mail address: yong.liu@edu.u-szeged.hu (Y. Liu).

https://doi.org/10.1016/j.tsc.2022.101169
Received 21 February 2022; Received in revised form 2 October 2022; Accepted 8 October 2022
Available online 12 October 2022

thinking (CT) to cultivate critical talents is imperative. In addition, employers throughout the world have been in search of those who exhibit CT (World Economic Forum, 2016). Ignorance at schools has resulted in inconsistency between students' CT outcomes and labor market needs (Lane & Oswald, 2016; OECD, 2017). Cruz et al. (2021) found that CT disposition is more frequently mentioned (69%, and 39% respectively) in the workforce than CT skills. Furthermore, employees, as the core of the labor market, have a clear understanding of dispositions of the market needs. However, there is a lack of instruments for assessing CT dispositions (Quinn et el., 2020). Moreover, consultation between employers and employees has not led to the development of such instruments. Consequently, it is crucial that a CT disposition assessment instrument be developed in accordance with employers' and employees' perceptions.

1.1. The framework of CT

Critical thinking, a general term with the complexity and uncertainty of its definition, experienced three phases in the framework development: one-dimensionality of only cognitive skills included, two-dimensionality of both cognitive skills and psychometric disposition, and a comprehensive model with skills and disposition as its input and problem-solving as its output.

In the first phase (the 1960s to 1980s), three shifts have been witnessed in the scope of cognitive skills in CT (Lewis & Smith, 1993): CT equaled problem-solving skills (Kemp, 1963), CT consisted of evaluation and judgment (Quellmalz, 1987), and CT combined problem-solving with evaluation (Ennis, 1987).

In the second phase (the 1990s to 2010s), the scope of CT skills has been extended broadly, with the disposition embedded into CT (Elder & Paul, 2020; Ennis, 2011; Facione, 1990b; Halpern, 1998; Ku, 2009). Accordingly, the two-component structure (skills-disposition model) starts. CT skills have been broadly defined as a facet of higher-order thinking skills (Lewis & Smith, 1993), reasoned and reflective thinking ability (Ennis, 2011), evaluative competence on accounts (Lawson, 1999), and scientific reasoning skills (Schmaltz et al., 2017). Although the scope is inconsistent, CT skills were summarized into six main dimensions: interpretation, analysis, evaluation, inference, explanation, and self-regulation as seen in Table 1 (Facione, 1990a). Subsequent classifications have included these six dimensions as well as verbal reasoning (Halpern, 1998), creative thinking, and mathematical problem-solving skills (Perkins et al., 1993).

CT disposition, a type of consolidated intellectual habit (Paul & Binker, 1990; Salomon, 1994) with the attributes of inclination, propensity, tendency, and willingness, is a prerequisite process that can activate CT skills (Ennis, 2011; Norris, 2003; Perkins et al., 1993). Despite the inconsistent contents of CT disposition, the basic seven dimensions are dominant. They are analyticity, inquisitiveness, systematicity, open-mindedness, truth-seeking, self-confidence, and maturity (Facione & Facione, 1994; Facione & Facione, 1992) as shown in Table 2. The following studies on its scope are all inspired by this (see Table 3).

Despite the acceptance of the scopes of both CT disposition and skills of the two-component structure and the effectiveness of the cultivation of CT via some teaching methods (Liu & Pásztor, 2022), due to the complexity of the world and the focus on students' learning outcomes, a theoretical framework to assess CT disposition and skills is not enough. Educators should know whether students can use CT to solve problems in daily life (practical implementation). Therefore, at the beginning of the third phase (the 2020s -), a comprehensive framework of CT was developed to clarify the relationships among all these components (Fig. 1). CT is not static but rather a dynamic process in which the disposition encompasses the means to ignite the acquisition of skills. In conjunction, both can strengthen the scenario problem-solving competence of the workforce. During this procedure, various interventions can be employed to enhance students' CT ability. This process comprises three steps, which are repeated as experience is accumulated.

To summarize, although CT skills have dominated through the whole first phase, disposition is regarded as the basis of skills. Despite mastering CT skills, it is unlikely that they can be used without disposition (Valenzuela et al., 2011). However, studies have revealed the presence of the cognition-preferred bias that most scholars employ to assess CT skills rather than disposition, which may result from the paucity of CT disposition assessment instruments (Bravo et al., 2020; Quinn et al., 2020; Sosu, 2013).

1.2. CT disposition scales

CCTDI (Facione & Facione, 1994; Facione & Facione, 1992) was the first CT disposition assessment instrument. The developers, through item and factor analysis, generated seven broader and unified dimensions from the 19 scales (see Fig. 2) of the Delphi Report (Facione, 1990b). However, in most studies, the reliability and validity of the original version appear to be inconsistent with the preliminary results. A literature review of these studies (Table 4) indicates: (a) When the original version was used, although the overall reliability can be acceptable, that of subscales was quite low, and by the factor analysis, the modifications of combining dimensions and shortening the number of items were put forward in most studies. (b) When they adopted the adapted version, although

Table 1
Contents of six main dimensions of CT skills.

Dimensions	Contents
Interpretation	To comprehend and present the category, meaning, and significance
Analysis	To check ideas, find out arguments, and analyze their components
Evaluation	To evaluate statements and arguments
Inference	To question evidence, infer choices, and make conclusions
Explanation	To demonstrate results, procedures, and arguments
Self-regulation	Self-evaluation and self-modification

Table 2
Contents of CT disposition dimensions.

Dimensions	Contents
Analyticity	Anticipate potential results or outcomes
Inquisitiveness	Curiosity to know things even though not instantly or obviously beneficial
Systematicity	Solve problems in a disciplined and organized way
Open-mindedness	Accept different opinions and bravely speak out on own
Truth-seeking	Use reason and evidence to understand any situations and avoid bias or preconceptions
Self-confidence	Trust your reflective thinking to make decisions and solve problems
Maturity	Recognize the complexity of problems and make mature decisions

Table 3
The taxonomies of CT disposition in the literature.

Scholar	Taxonomy
Perkins et al. (1993)	The inclination: (1) to be broad and adventurous; (2) to possess curiosity; (3) to clarify and build understanding; (4) to make plans and be strategic; (5) to be intellectually careful; (6) to seek and evaluate reasons; (7) to be metacognitive
Halonen (1995)	The disposition to be: (1) tentative; (2) a skeptic; (3) tolerant of ambiguity; (4) indulgent of individual differences; (5) careful of ethical issues
Ennis (1996)	The tendency to (be): (1) seek alternatives and be open to them; (2) well-informed; (3) endorse a position when it is justified to do so; (4) consider others' viewpoints; (5) clear about intended meaning; (6) determine and maintain focus on the conclusion or question; (7) seek and offer reasons; (8) take into account others' feelings and level of understanding; (9) reflectively aware of own beliefs; (10) discover and listen to others' view and reasons; (11) take into account the total situation; (12) concerned about others' welfare
Halpern (1998)	The tendency in: (1) habitual use of plans and the suppression of impulsive activity; (2) willingness to engage in and persist in a complex task; (3) flexibility or open-mindedness; (4) willingness to abandon non-productive strategies in an attempt to self-correct; (5) awareness of social realities so that thoughts can become actions

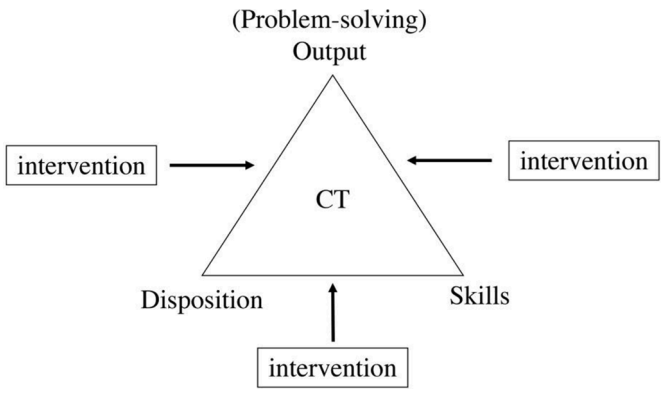


Fig. 1. The framework of CT.

the reliabilities of both overall scale and subscales were good, they, based on the results of factor analysis, also suggested combining dimensions and shortening the number of items to get a better validity. Researchers have also summarized the overlaps among the dimensions and lack of demographic information and factor loadings of CCTDI (Walsh & Hardy, 1997; Walsh et al., 2007). Despite its disadvantages, CCTDI contains recommendations for subsequent instruments and discussions on disposition dimensions. However, suitable adaptation and further exploration are essential.

Based on CCTDI, Yoon (2004) developed the YCTD scale for Korean nursing students. The dimensions are similar to those in the Delphi Report, including 27 items in the following seven subscales: objectivity, prudence, systematicity, intellectual curiosity, intellectual fairness, healthy skepticism, and self-confidence. Although the initial reliability was measured as .84, construct validity and measurement invariance were not assessed. Subsequently, Shin et al. (2015) confirmed Cronbach's Alpha was .84. Once items were deleted, Cronbach's Alpha ranged between .83 and .85. The model fit by Confirmative Factor Analysis (CFA) and measurement invariance was both acceptable. This instrument appears to be an adaptation of CCTDI, thus supporting the notion that disposition instrument adaptation and improvement are both crucial.

Accordingly, Sosu (2013) developed and validated CTDS whose dimensions were constructed under the Delphi Report and CCTDI as well as the taxonomies related to CT dispositions and two other researchers' advice. The initial item bank comprised 98 items on all the dimensions in previous taxonomies. After discussions, 46 items were retained. Subsequently, a pilot study on these 46 items revealed flaws in several items. To ensure validity, 24 items were finalized for the main study. Explorative Factor Analysis (EFA) and CFA recommended 11 items to formulate two factors, namely, critical openness and reflective skepticism, with a good model fit. However, the model checks Bravo et al. (2020) and Yockey (2016) conducted revealed that although the model fit of Sosu's two-factor

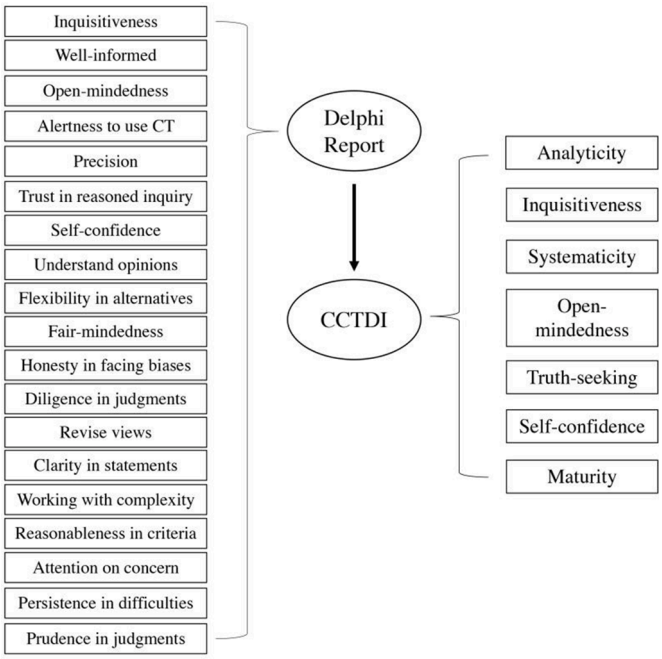


Fig. 2. Dimensions of Delphi report and CCTDI on CT disposition.

Table 4
The summary of research on CCTDI in previous studies.

Researchers	Context	Initial Adaptation	Initial Reliability Cronbach's Alpha		Necessity for Improvement Suggested after Factor Analysis	
			Overall	Subscales	Unify Dimensions	Shorten Items
Ip et al. (2000)	China	No	.85	.34–.76	No	No
Walsh and Hardy (1997)	USA	No	.91	.57–.78	Yes	Yes
Bondy et al. (2001)	USA	No	.89	.39–.77	No	Yes
Bondy et al. (2001)	USA	No	.90	.49–.79	No	Yes
Kakai (2001, 2003)	USA	No	.89	.59–.77	Yes	Yes
Walsh et al. (2007)	USA	No	.91	.53–.84	Yes	Yes
Yeh (2002)	China	No	.71	.34–.73	Yes	Yes
Yeh (2002)	China	No	.71	.52–.73	Yes	Yes
İskifoğlu and Ağazade (2013)	Turkey	No	.87	.81–.90	Yes	Yes
Wang et al. (2019)	China	Yes	.92	.82–.86	Yes	Yes
Peng et al. (2004)	China	Yes	.90	.54–.77	Yes	Yes
Kokdemir (2015)	Turkey	Yes	.88	.61–.78	Yes	Yes

model was not superior to the one-factor model, it was acceptable. Cronbach's Alpha demonstrated the reliabilities of the two-factor models were .78 and .79 respectively, which may be regarded as acceptable. However, because this model of CT disposition does not appear to be stable, an in-depth discussion is important.

To rectify these imperfections, Quinn et al. (2020) developed and validated SENCTDS in an endeavor to address the limitations of other instruments that failed to consider students' perceptions. This development procedure, which comprises four collective intelligence (CI) sessions (Dwyer et al., 2016), employed an interactive management methodology (Warfield & Cárdenas, 1994) and generated 13 categories with 101 items initially for validation. Six dimensions with 21 items emerged after EFA and CFA had been conducted, namely, reflection, attentiveness, open-mindedness, organization, perseverance, and intrinsic goal motivation. While the overall reliability is .77, that of its subscales ranges from .59 to .82 with an acceptable model fit. The convergent and predictive validity of SENCTDS is positive. However, despite these positive attributes, there are overlaps in the inventory. Although attentiveness is unique, it may be regarded as a form of perseverance. Intrinsic goal motivation, as a broad subscale, is grouped with subscales that have a narrow scope, thus resulting in conceptual vagueness. Although student participation was emphasized in the development process, young students with little knowledge of CT tend to choose skills that they may lack but not need in the future.

In brief, the dimensions in Delphi Report formed the basis of the CT disposition from which subsequent instruments were borrowed. However, ongoing discussions on components of CT dispositions are still prevalent. Differences related to categories of sub-dimensions and structure are evident. Although progress in the field has been made, the availability of reliable and valid CT disposition instruments remains limited. The procedure to develop CT disposition inventory can generally be divided into three perspectives: literature-based, educator-based, and educator-student-based. While the first two perspectives have been dominant, instruments developed in consultation with employers and employees are lacking.

Accordingly, the purpose of this study was to design and validate a new CT disposition scale: The Employer-Employee-Supported Critical Thinking Disposition Inventory (2ES-CTDI). Based on the aforementioned gaps and aims, the research questions in this study are:

(a) What kind of CTD framework will be identified by the expert group based on the intelligence from the literature, employers, and employees?
(b) Is the conceptual framework consistent with the statistical data?
(c) Is the 2ES-CTDI reliable and valid for assessing undergraduate students' CTD?
(d) What is the relationship among the components of CTD?
(e) What is the endorsement of the 2ES-CTDI and its items according to item response theory?

2. Materials and methods

2.1. Participants

The participants in this study included three groups: employers and employees, Interpretative Structural Model (ISM) group, and college students. The expected dispositions for college students from a total of 25 employers and 43 employees in different fields in China (Table 5). The classification of economic activities was based on the theories of the European Statistical Office (2008). The ISM group comprised seven persons: two scholars who were professionals in CT research, one coordinator in charge of organization and ISM technical issues, one employer from the field of human resources and one from the education system, and two employees from the education system. Finally, 661 college students from four universities in southwest China voluntarily and anonymously participated in the assessment (Mean = 19.57; SD = 1.26). Specifically, there were 328 male (49.6 %; Mean = 19.22; SD = 1.20) and 333 female students (50.4 %; Mean = 19.91; SD = 1.23).

The recruitment of employers, employees, ISM members, and undergraduate students was up to a non-probability sampling approach. Employers were recruited via volunteer sampling. A human resource company was invited to release the online survey to all employers and employees who cooperated with it. ISM members were purposely selected since we should find two experts, one who can use ISM software, two employers who can know much about most fields, and two employees who have much experience in their jobs. For the recruitment of college students, we asked student counselors in some universities to release online questionnaires to their students.

2.2. Procedure

This study comprised two phases: first, the finalization and item development phase, and second, the statistical study validation phase (Fig. 3). An in-depth elucidation of these phases follows.

Table 5
The distribution of employers and employees.

Economic Activities		No. of Participants		Proportion
Main Category	Sub-Category	Employer	Employee	
Education	Primary Education	0	3	63.2%
	General Secondary Education	6	6	
	Tertiary Education	3	6	
	Other Education (Cramming School)	9	10	
Social Sciences, Journalism and Information	Sociology and Cultural Studies	0	1	7.4%
	Restaurants and Mobile Food Services	2	2	
Business, Administration, and Law	General Public Administration	0	1	22%
	Foreign Affairs	0	2	
	Landscape Services	0	2	
	Central Banking	1	1	
	Non-life Insurance	0	1	
	Risk and Damage Evaluation	1	2	
	Real Estate Agencies	1	1	
	Human Resources Administration	1	1	
Information and Communication Technologies (ICTs)	Computer Programming	1	4	7.4%
Total	68	25	43	100%

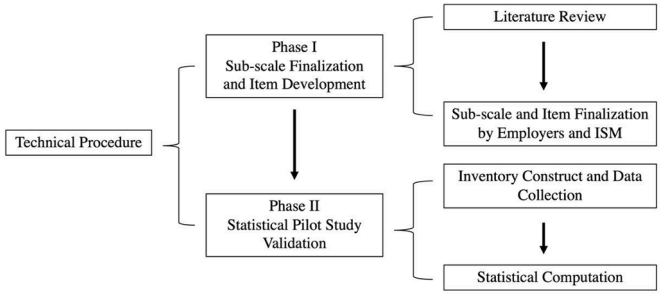


Fig. 3. The technical procedure employed.

- Phase I

Step 1: A literature review. Given all the disposition assessment tools that have been based on the dimensions recommended by the Delphi Report, although this instruction for disposition subscales was followed, additional ones from employers and employees were accepted in this study. Furthermore, the items from previous instruments were collected for evaluation.

Step 2: Employers' and employees' subscale item finalization by ISM. Employers and employees wrote additional suggestions on a document in which the subscales from the Delphi Report and the aforementioned scales were described. Specifically, the employers and employees described the dispositions they expected graduates to possess. In accordance with these expectations, the ISM process, combined with the intelligence accumulation (IA) approach was conducted to finalize the retained dimensions and items.

- Phase II

Step 1: Inventory construction and data collection. After comparisons, discussion, and finalization, the inventory was integrated into the online survey and sent to participants by email or social applications via a QR code or link. It is estimated participants need approximately five to ten minutes to complete the inventory. Furthermore, the data, recorded automatically online, anonymously, and confidentially, can be downloaded at a later stage for data analysis.

Step 2: Statistical computation. EFA was first conducted to explore the factor loadings and content validity of the matches between the ISM framework and data. CFA and the common method variance (CMV) were subsequently conducted with acceptable matches of ISM's taxonomy to identify the best model and its model fit. Then, the basic scale reliability was checked via Cronbach's Alpha, and the additional structural validity (the convergent and discriminant validity of the identified model) was examined. Partial Least Squares Structural Equation Modeling (PLS-SEM) was further employed to specify the inner effects among the factors. In addition, Item Response Theory (IRT) was utilized to investigate inventory endorsement and item discrimination when examining the causes of the ineffective items.

2.3. Data analysis techniques

2.3.1. Subscale finalization and item development techniques

The term IA was coined in this study. IA is a review technique employed to collect experience from previous or current intelligence, including the prevalent literature, questionnaires, and/or interviews with scholars and/or correspondent subjects, and experimental or natural phenomena, so as to synthesize and compare the results that could be employed in the present study. It can be used as a basic theory to support later research. IA was utilized in the first two steps of Phase I, namely, the literature review and the collection of employees' and employers' intelligence (i.e., their expert knowledge and information about the matter).

Warfield (1973a, 1973b) detailed the conceptual and theoretical interpretations of ISM. ISM, one of the qualitative-oriented analysis approaches in modern system engineering, can be employed to simplify a complex system into subsystems with hierarchical levels by employing people's experience and knowledge as well as computers. Currently, it is frequently used in the research of educational sciences research (e.g., Quinn et al., 2020). The steps conducted in ISM help one to classify and explore the relationships among sub-dimensions generated from studies and expected by employers (Fig. 4). The ISM intelligence group can comprise technical experts, an organizer or coordinator, and other intelligence participants. The correlations of each factor should be listed in both the column and row in a table: while one depicts that one factor has a direct relationship with the other one, zero indicates no direct relationship. The adjacency matrix and accessibility matrix are used to explain the one-level and two-level relationships among factors, respectively. These two methods were employed to check the content validity of the inventory.

2.3.2. Statistical techniques

The degree to which an assessment instrument is relevant to and representative of the specific construct it is supposed to evaluate is referred to as content validity (Rusticus, 2014). This was evaluated from three perspectives: domain definition, domain representation, and domain relevance (Sireci, 1998). This was fulfilled previously when the literature reviewed dimensions and items generated from previous studies and representatives of the ISM group. EFA was employed to check the consistency between ISM and dimension

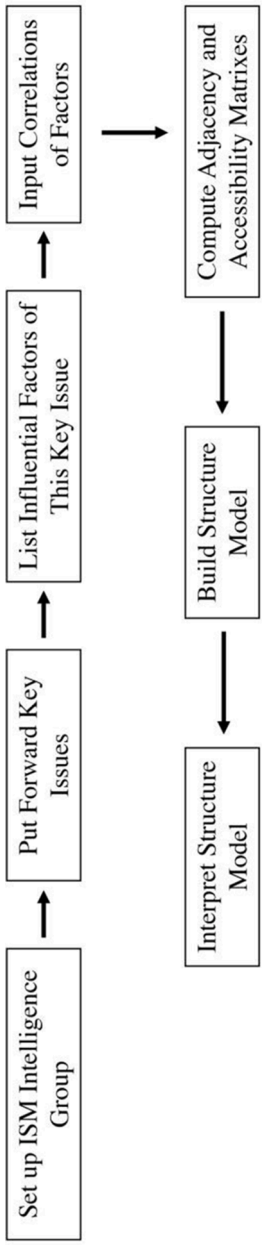


Fig. 4. The steps of conducting ISM.

structure (domain relevance) and monitor the rotated factor loadings, retaining those above 0.4 (Worthington & Whittaker, 2006). In addition, the Kaiser-Meyer Olkin (KMO) test (Kaiser, 1970) and Bartlett's test (Snedecor & William, 1989) were conducted to examine the common variance and homogeneity of the variances of the data, respectively.

The common-method variance (CMV, Podsakoff et al., 2003), which may be employed to explore if the structure is influenced by indicators, instruments, and other factors, often requires combining all the items into one construct and comparing the fit between the one-factor model and the target model. If the one-factor model fit is less superior to the target model, the credibility of the latter may be concluded. In addition, CFA, a component of the Structural Equation Model (SEM), can be deployed to examine the model fit. Chi-square (Pearson, 1900), the Root Mean Square Error of Approximation (RMSEA; Steiger, 1990), the Comparative Fit Index (CFI; Bentler, 1990), and the Tucker–Lewis index (TLI; Bentler & Bonett, 1980; Tucker & Lewis, 1973) were employed for this purpose.

Cronbach's Alpha (Fornell & Larcker, 1981), evidence of internal consistency, is often employed to indicate scale reliability. Two aspects of construct validity, proposed by Campbell and Fiske (1959), have been regarded as powerful evidence that CFA can collect to further confirm the model fit:

(a) Convergent validity, which indicates to what degree the construct has been measured by indicators, can be determined by average variance extracted (AVE) and composite reliability (CR). A value above .50 is acceptable for AVE, which captures the variance level of construct versus measurement error. A level of more than .70 is regarded as acceptable for CR. Dillon-Goldstein's rho (DG), another type of CR was tested by employing the PLS-SEM approach, which is more dependable because it is generated by the weight of each indicator. Chin (1998) noted that the value level should be higher than .70.
(b) Discriminant validity, which signifies the level of non-relation among different indicators, can be assessed by the square root of AVE (Fornell & Larcker, 1981) and the Heterotrait-Monotrait ratio of the correlations (HTMT; Henseler et al., 2015). The value of the square root of the AVE of each indicator should be larger than their correlations (Fornell & Larcker, 1981). Although the HTMT value should be less than .90 (Teo et al., 2008), it is recommended that it is less than .85 (Kline, 2011).

To enhance the validity of the model and avoid the influence of the possible non-normality of the data, PLS-SEM (Wold, 1985) was employed to examine the mediation effect among factors in conjunction with CFA. However, although classical test theory was employed in many studies to identify the validity of the instrument, it can just output the fit between model and data instead of among model, data, test, and student's ability. Thus, IRT (Samejima, 1969) was employed to determine inventory endorsement (information that the test can provide and the match between the scale and students' competence), item discrimination, the predictive direction, and the path for further inventory modification.

Table 6
Subscales generated from IA.

Subscales	No. of Choices	Proportion	Final Choice
From the Literature			
Well-informed	51	75%	Yes
Precision	37	55%	Yes
Self-confidence	44	65%	Yes
Understand opinions	51	75%	Yes
Flexibility in alternatives	20	30%	No
Fair-mindedness	24	35%	No
Honesty in facing biases	34	50%	Yes
Diligence in judgments	44	65%	Yes
(Self-)Revise views	41	60%	Yes
Clarity in statements	27	40%	No
Reasonableness in criteria	24	35%	No
Attention to concerns	41	60%	Yes
Persistence in complexities	37	55%	Yes
Prudence in judgments	34	50%	Yes
Systematicity	51	75%	Yes
(Self-)Reflection	51	75%	Yes
Objectivity	47	70%	Yes
Analyticity of possible results	24	35%	No
Healthy skepticism	34	50%	Yes
Careful about ethical issues	41	60%	Yes
Concern for others' feelings	27	40%	No
Concern for others' benefit	14	20%	No
Awareness of social realities	44	65%	Yes
From Employers and Employees			
Cooperation and collaboration	-	-	Yes
Independent thinking	-	-	Yes
Retaining original aspiration	-	-	Yes
Global insights	-	-	Yes

3. Results

3.1. Construction of the inventory

To avoid repeated choices from employers, ISM group members synthesized the Delphi Report and scales of other instruments and deleted the duplicates. Finally, 23 sub-dimensions were approved (Table 6). The collection of the employers' and employees' expectations demonstrated that all the dispositions from studies had been referred to and four additional CT dispositions, namely, cooperation and collaboration, independent thinking, keeping original aspirations, and global insights proposed. Subsequently, those with 50% or more of the employers' and employees' expectations were chosen as the potential subscales. Although the four dispositions that were proposed by employers and employees were retained, they were not supported by theory. Finally, 20 subscales were approved. Through discussion and modeling in the ISM group, the initial CT disposition framework was developed (Fig. 5). It encompassed two scopes: the mindedness-oriented disposition, which included intrinsic motivation (self-efficacy) and extrinsic perception (instant judgment), and skill-oriented disposition (habitual truth-digging).

In accordance with the CT Disposition framework, the ISM group selected and adapted the suitable items from the item bank of the previous instruments to be employed in the 2ES-CTDI. It comprises 12 items for critical-mindedness-oriented disposition, specifically eight items for the intrinsic motivation dimension and four for extrinsic perception as well as eight items for the critical skill-oriented disposition (Table 7). In addition, given the response designs of the previous questionnaires in which CCTDI employed a 6-point scale, YCTD and CTDS a 5-point scale, and SENCTDS a 7-point scale, a 7-point Likert scale was deemed suitable to assess participants' true evaluations in a wide-scale without forcing them to choose one side (Finstad, 2010; Taherdoost, 2019), ranging from 1 (strongly disagree) to 7 (strongly agree). Some reverse items were applied to check students' concentration. The English version was then translated into Chinese by two professional interpreters who had been certified by the China Accreditation Test for Translators and Interpreters.

3.2. Empirical examinations of the inventory

3.2.1. Examining the content validity

After the finalization of the inventory, data were collected from 678 respondents. We deleted the data of 67 respondents because of missing values and inconsistent answers between positive and negative items. Therefore, data from 611 respondents were analyzed. EFA with principal component analysis was conducted to determine the relevance between the theoretical framework and statistical data. The results revealed the KMO was .925 and Bartlett's test: $\chi^2 = 7163.071$, df = 325 ($p < .001$), thus indicating the EFA was reliable and additional EFA could be conducted. The initial eigenvalues suggested three factors with a 50.75% variance. The varimax rotated component matrix which is consistent with our theoretical framework developed by the ISM group is displayed in Table 8. A good AVE score requires almost each factor loading and communality to be more than .70 and .20, respectively. Accordingly, to ensure an enhanced structure, the items with factor loadings below .6 were to be deleted: A1, A3, A6, B4, C6, and C8. Moreover, some items such as B4 and C3 were also to be deleted because of the higher factor loadings listed in other factors that were inconsistent with the expected ones and high cross-loadings. Interviews with some participants revealed that the descriptions of these items were related to two factors, which may have misled the participants. For example, while Item A6, *I believe in my ability to make rational and informed*

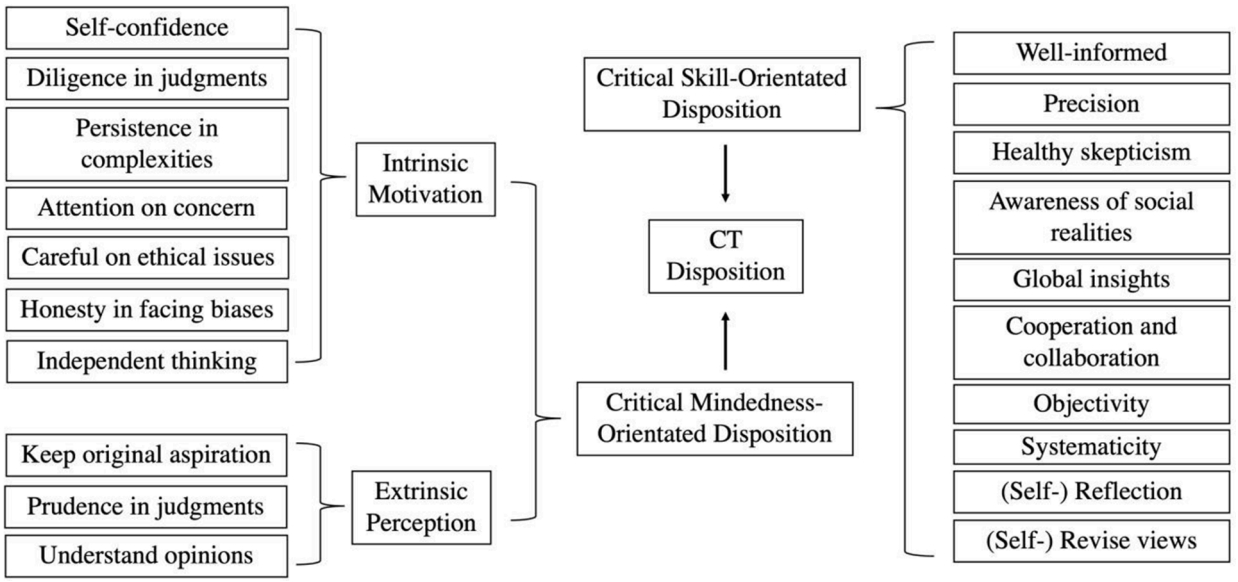


Fig. 5. Initial CT disposition structure generated by ISM.

Table 7
Items finalized by ISM.

IDs	Items	Adapted Source
Critical Mindedness-Oriented Disposition		
Intrinsic Motivation: Self-Efficacy (L1)		
A1	I would not like to spend time exploring the truth without any benefit to me. (R)	ISM
A2	If, for one thing, most people agree with the conclusion and few reject it, I tend to choose to agree. (R)	CCTDI
A3	Although I have found the evidence violating my thoughts, I tend to persist in mine. (R)	CCTDI
A4	When dealing with complicated issues, I often tend to give up. (R)	CCTDI
A5	I'm easily distracted when dedicated to a task. (R)	SENCTDS
A6	I believe in my ability to make rational and informed judgments.	SENCTDS
A7	When making judgments, I am easily influenced by ethical values while ignoring reasoning thinking. (R)	ISM
A8	When making decisions, I tend to be impulsive. (R)	CCTDI
Extrinsic Perception: Instant Judgment (L2)		
B1	The decisions of people with power will be right. (R)	CCTDI
B2	In my mind, critical thinking focuses on criticizing others' opinions. (R)	ISM
B3	When discussing in a group, if someone's opinion is regarded as wrong by others, he should not have any chance to express his thoughts. (R)	CCTDI
B4	Facing inconsistent issues, I tend to agree with the one who is the last to give me advice. (R)	CCTDI
Critical Skill-Oriented Disposition		
Habitual Truth-Digging (L3)		
C1	If I need to reject others' opinions, I tend to propose my reasons for such.	CCTDI
C2	Whenever I read or hear an assertion or conclusion, I tend to think about possible alternatives.	ISM
C3	When I hear or read international news in my mother tongue, I tend to search for the original version to confirm my opinion.	ISM
C4	Before deciding, I tend to search for all the potentially relevant materials.	CCTDI
C5	I tend to consider the social context and reality when dealing with problems.	ISM
C6	When in group work, I am ready to cooperate with others well.	ISM
C7	I prefer proposing suspensions while others are in a calm mood.	ISM
C8	I tend to question conclusions I hear or read to decide if they are convincing.	ISM
C9	I like to analyze any statements or opinions rather than just read and let them go.	ISM
C10	I like to deal with problems logically (step by step or with a good plan).	CCTDI
C11	When judging others' statements, I tend to use reasonable evidence or criteria rather than basing them on my own experience.	ISM
C12	I like to reflect on my own or others' thinking procedures.	CTDS
C13	When I find myself or others using less-persuasive evidence to support opinions, I tend to drive myself or them to adjust the evidence.	ISM
C14	I am accustomed to predicting possible results when I solve problems.	CTDS

Note: R = reversed scoring.

Table 8
The Varimax rotated component matrix.

Item	Communalities	Factor One	Factor Two	Factor Three	To be Deleted
A1	.204	.349			yes
A2	.460	.659			
A3	.302		.348		yes
A4	.637	.785			
A5	.512	.700			
A6	.375			.599	yes
A7	.470	.659			
A8	.519	.713			
B1	.657		.617		
B2	.630		.665		
B3	.681		.754		
B4	.521	.513			yes
C1	.632			.777	
C2	.544			.727	
C3	.577		.641		yes
C4	.609			.752	
C5	.543			.730	
C6	.304			.543	yes
C7	.417			.604	
C8	.330			.477	yes
C9	.526			.661	
C10	.607			.756	
C11	.471			.614	
C12	.567			.724	
C13	.513			.680	
C14	.588			.707	

judgments, demonstrated both the self-confidence and inclination to judge rationally. Item A9, *I prefer proposing suspensions while others are in a calm mood to anytime*, included questioning others' opinions and self-control when caring for others emotionally.

3.2.2. Examining the model fit, reliability, and construct validity

To ensure the model fit that EFA identified, CFA on this three-factor model was also examined (Table 9, first row, EFA Model). After the deletion of Items A1, A3, A6, B4, C3, C6, and C8 based on the aforementioned reasons, in accordance with the theoretical framework, to test CMV, the model fits of the one-, two-, and three-factor model were compared with one another (Table 9). The results showed that the three-factor model was the best, with $X^2/df < 5$, RMSEA $< .08$, CFI and TLI $> .90$, SRMR $< .08$, and almost the smallest AIC and BIC.

As shown in Table 10, all the Cronbach's Alphas of each dimension and overall inventory are more than .70, indicating the good internal consistency and good reliability of the inventory. Besides, the factor loadings of nearly all the items were above .70, CR and DG values were more than .7, and all the AVEs were above .5, thus signifying the inventory has strong convergent validity.

The discriminant validity was shown in Table 11. While HTMT values between every two latent variables were listed at the top right corner with all values $< .85$, the correlations between the variables were in the bottom left corner with all values $< \sqrt{AVE}$s, which demonstrated that the three indicators, to some extent, were separate from each other, thus ensuring discriminant validity.

Subsequently, the model was developed (Fig. 6). The model illustrates that the correlations between L2 (instant judgment) and L1 (self-efficacy) as well as between L1 and L3 (habitual truth-digging) were significant. However, the correlation between L3 and L2 was not significant. To analyze the insignificant correlation between L2 and L3, IRT was conducted for all the items. The results revealed that all the items in L2 had low item discrimination (B1, Coef. $= -.14$; B2, Coef. $= .03$; B3, Coef. $= .03$ respectively, all *p-values* $> .05$). This indicated that with regard to instant judgment (L2), most students tended to judge without deep thought but expected positive results.

3.2.3. The mediating effects among the components of CTD

Subsequently, the PLS-SEM model (Fig. 7) was developed to explore further relationships (Table 12) via the paths L2, L1, and L3 as well as L3, L1, and L2. It may be concluded that (a) the strong disposition to make a correct instant judgment (L2) will accelerate the level of self-efficacy (L1). In addition, although strong self-efficacy will promote the tendency of habitual truth-digging (L3), enhanced instant judgment has a significantly negative effect on habitual truth-digging, with self-efficacy as the competitive partial mediation (path coefficient $-.197$, $p < .01$; indirect effect $p < .001$, $z > 1.96$). Thus, the strong instant judgment disposition may decrease the tendency to uncover truth habitually by a partial influence of self-efficacy. (b) In contrast, for the second path, habitual truth-digging has a positive effect on self-efficacy, which may also have a positive effect on the enhancement of instant judgment. However, habitual truth-digging has no significant effect on instant judgment (path coefficient $p > .05$) but can influence the instant judgment disposition via the complete mediation of self-efficacy (indirect effect $p < .01$, $z > 1.96$). Similar results were obtained from IRT analysis. For instance, in Fig. 8, the Item B1 category characteristic curves were generated from IRT to demonstrate the relationship between the probability to obtain scores and students' CTD level, indicating that while the students with a CT disposition level of $\theta < -16.5$ (a lower-level ability) are likely to acquire seven points (a higher-level of instant judgment), those with $\theta > -6$ (a higher-level ability) will probably obtain one point (a lower-level of instant judgment). This also shows that the habitual truth-digging inclination will be cut down by the increase of instant judgment tendency.

3.2.4. Examining the test endorsement

The micro model fit with IRT was checked concerning test characteristics and item information. Fig. 9 shows the relationships between students' CTD level and expected inventory score, which may be noted that with the students' CT disposition level (θ) increasing, the expected score increased, thus indicating that the inventory was reliable and could be used to examine students' real CT disposition level.

In relation to the item information plot (Fig. 10), it shows the relationship between the amount of item information and students' CTD level. Most parts of the curves are located on the negative sides ($\theta = 0$) of the students' level (θ). The two peaks, $\theta = -2$ and 1, illustrate that the test can provide more information for low-level students, and it is more suitable for students with the CT disposition level θ between -2 and 1. One can conclude from the second plot that for average-level students, the information provided by the test is somewhat lower than that for students in the range. This may suggest that, if the aim is to test average level, it is preferable to add more items for average level students. Furthermore, all the items in L1 (As) and L2 (Bs) provided less information than those in L3 (Cs), thus indicating that the critical-mindedness-oriented disposition may not be essential in CT disposition components. However, one may

Table 9
Comparisons of model fits.

Model	X^2	X^2/df	RMSEA	CFI	TLI	SRMR	AIC	BIC
EFA	1417.97***	4.79	.076***	.839	.823	.097	55521	55885
One-	2358.32***	15.52	.148***	.587	.535	.161	41432	41688
Two-	885.26***	5.86	.086***	.863	.844	.071	39961	40221
Three-	583.18***	3.91	.066***	.919	.907	.054	39663	39932

Note:
*** $p < .001$

Table 10
Reliability and convergent validity of the inventory.

Item	Mean ± SD	Factor Loadings			Cronbach's Alpha	CR	DG	AVE
		L1	L2	L3				
A2	4.12 ± 1.57	.608			.796	.831	.859	.50
A4	3.79 ± 1.63	.762						
A5	4.13 ± 1.56	.709						
A7	4.34 ± 1.44	.700						
A8	4.05 ± 1.49	.738						
B1	2.56 ± 1.71		.740		.824	.814	.859	.60
B2	3.03 ± 1.67		.782					
B3	3.05 ± 1.79		.787					
C1	5.32 ± 1.37			.759	.905	.919	.919	.51
C2	4.89 ± 1.34			.732				
C4	4.88 ± 1.34			.770				
C5	5.03 ± 1.24			.730				
C7	4.99 ± 1.34			.602				
C9	4.66 ± 1.31			.690				
C10	4.83 ± 1.31			.781				
C11	4.59 ± 1.31			.650				
C12	4.95 ± 1.27			.733				
C13	4.87 ± 1.27			.680				
C14	5.07 ± 1.26			.704				
Overall Cronbach's Alpha	.858							

Note: L1: Intrinsic Motivation: Self-Efficacy; L2: Extrinsic Perception: Instant Judgment; L3: Habitual Truth-Digging; CR: Composite Reliability; DG: Dillon-Goldstein's rho; AVE: Average Variance Extracted

Table 11
Correlations among latent variables, square root of AVE, and HTMT.

	L1	L2	L3
L1		.66	.24
L2	.669		.03
L3	.210	.008	
$\sqrt{AVE}$	.706	.770	.713

Note: L1: Intrinsic Motivation: Self-Efficacy; L2: Extrinsic Perception: Instant Judgment; L3: Habitual Truth-Digging; AVE: Average Variance Extracted

deduce that the inventory is dependable and informative.

4. Discussion

4.1. Discussions on research questions a, b, and d related to CTD framework

Research questions a, b, and d are related to the CTD framework identified by the ISM group based on the intelligence from the literature, employers, and employees, its consistency with statistical data, and the relationship among these components.

The three-factor (instant judgment, self-efficacy, and habitual truth-digging) framework, though, tested consistent with statistical data, is different from those in previous studies. Those in previous studies were mostly comprised of self-efficacy and habitual truth-digging disposition without any explorations of instant judgment. However, the ISM group identified this factor by the ISM approach to be discussed in this study.

Instant judgment, which was generally ignored previously, was a new perspective employed in the inventory. While most studies (e.g., Holloway et al., 2021) have examined the positive relationship between CT and decision-making, which may be regarded as a form of habitual judgment in general, this type of decision-making was divided into two parts as suggested by Rutter and Brown (2019) and Kahneman (2011) in this study: instant and habitual ones. The conclusion that instant judgment is necessary to reflect on issues quickly but more thought is needed to specify the nature of these judgments was arrived at in this study (Rutter & Brown, 2019). The inclination of habitual truth-digging is necessary to weaken such. This procedure also proves CT is a form of competence to find a balance to solve problems well and make the right decision but is neither a short- nor long-term thinking process. These mutual restrictions between instant judgment and habitual truth-digging were also identified in this study. However, self-efficacy was found to have a mediating effect on these restrictions, which means it can be used to keep the balance for students' CTD improvement.

Besides, although the inclusion of self-efficacy was also in dispute in previous studies, any CT disposition inventories, which have also examined self-efficacy and skill-oriented disposition, obtained similar results to those of this study, namely, self-efficacy did not appear linked to the CT disposition (Quinn et al., 2020) and was thus deleted from the inventory. However, conclusions were drawn from the perspective of low factor loadings without any powerful correlation and regression evidence. However, IRT and PLS-SEM

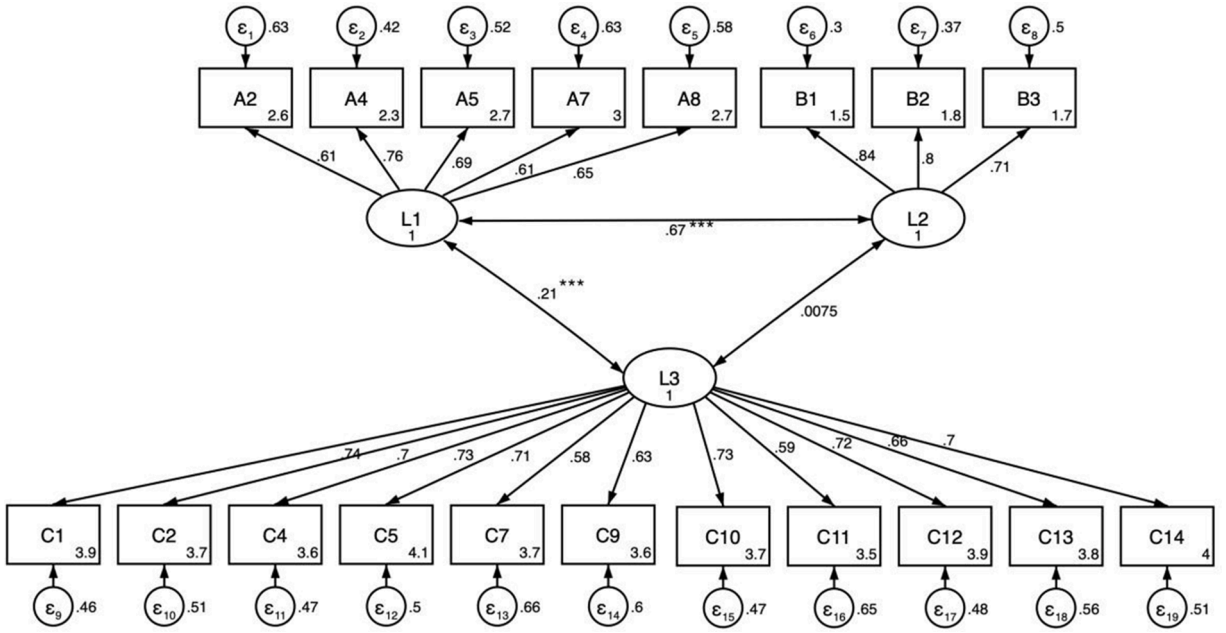


Fig. 6. Structural model of the inventory (L1: Intrinsic Motivation: self-efficacy; L2: extrinsic perception: instant judgment; L3: habitual truth-digging).
Note: ***$p < .001$

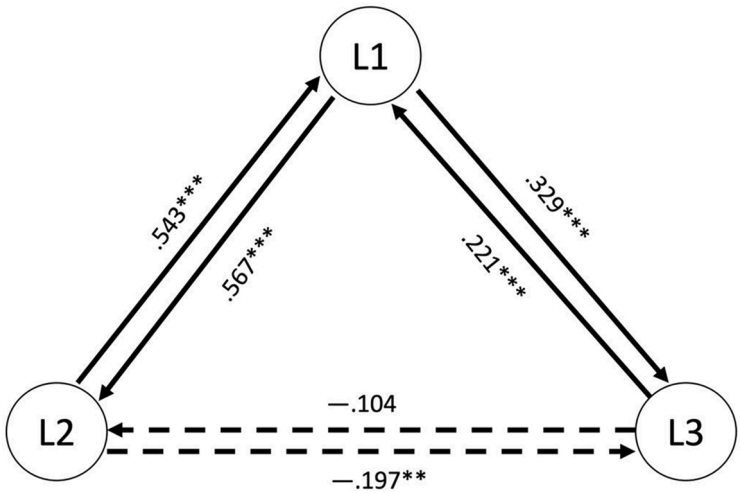


Fig. 7. PLS-SEM mediate effect model (L1: Intrinsic Motivation: Self-Efficacy; L2: Extrinsic Perception: Instant Judgment; L3: Habitual Truth-Digging).
Notes: **$p < .01$; ***$p < .001$

Table 12
Bootstrapping mediation effect test.

Model Path	Indirect Effect	S.E.	*z*	CI Low	Up
L2-L1-L3	.179***	.039	4.565	.109	.261
L3-L1-L2	.125**	.037	3.347	.053	.198

Notes: L1: Intrinsic Motivation: Self-Efficacy; L2: Extrinsic Perception: Instant Judgment; L3: Habitual Truth-Digging
** $p < .01$;
*** $p < .001$;

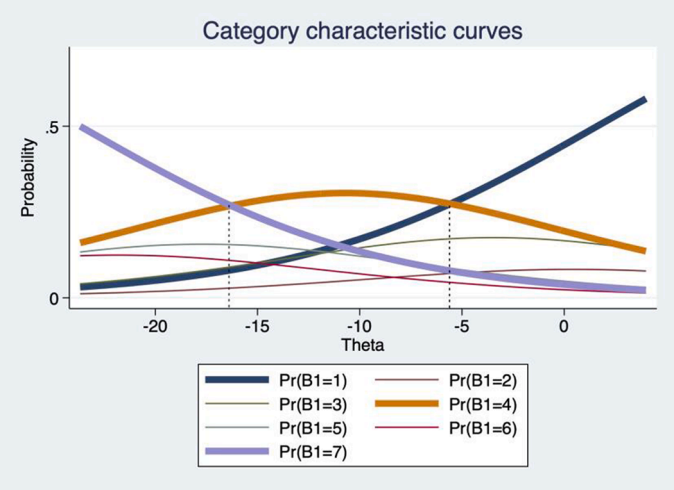


Fig. 8. The relationship between the probability of obtaining scores and students' CTD Level of Item B1.

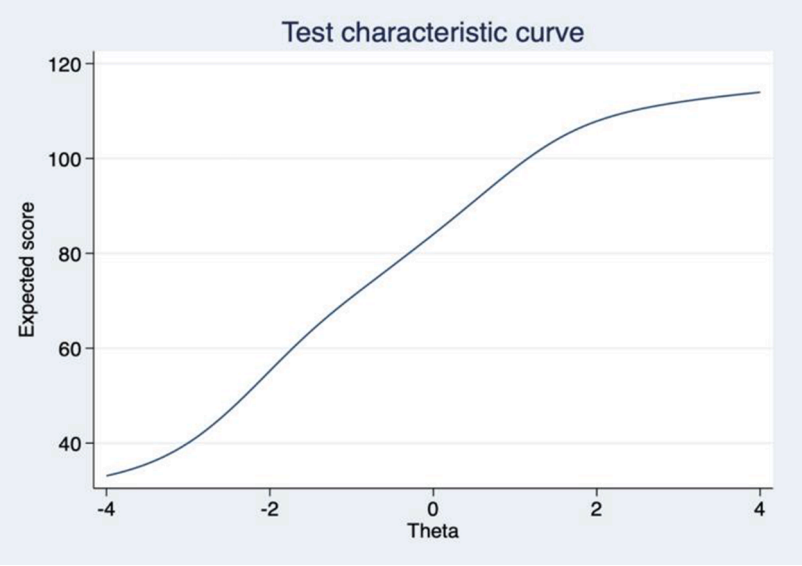


Fig. 9. The relationships between students' CTD level and expected inventory score.

were employed in this study to demonstrate that the critical-mindedness-oriented disposition has less of an effect on CT disposition than skill-oriented ones. However, its mediating effect was strong enough to demonstrate that this removal was not adequate to assess CT disposition because the model fit and test characteristic curve showed that what was tested was what was required and thus, self-efficacy was retained in the inventory.

4.2. Discussions on research questions c and e related to the inventory and its items

Research questions c and e are related to the endorsement of the overall inventory and its items. The previous studies on the validation of scales were mostly based on EFA, CFA, and classical test theory (CTT), which can also demonstrate the model fit and test endorsement, but not enough. The basic results generated from EFA, and CFA cannot further explore if the inner factors gather and separate from each other well. Thus, additional convergent and discriminant validity checks were employed. Besides, CTT, considering individual true scores as the expected ones, cannot generate dependent assumptions (Rusch et al., 2017). Thus, this study made an additional IRT analysis on the whole inventory and items to show more details on the endorsement.

Reliability, EFA, and CFA results indicated the inventory was reliable and valid for assessing undergraduates' CTD, though as noted previously, some items were deleted because of low factor loadings and high cross-loadings, which concurs with previous studies. However, IRT showed some unexpected results on item discrimination. Although items with low discrimination are regarded as

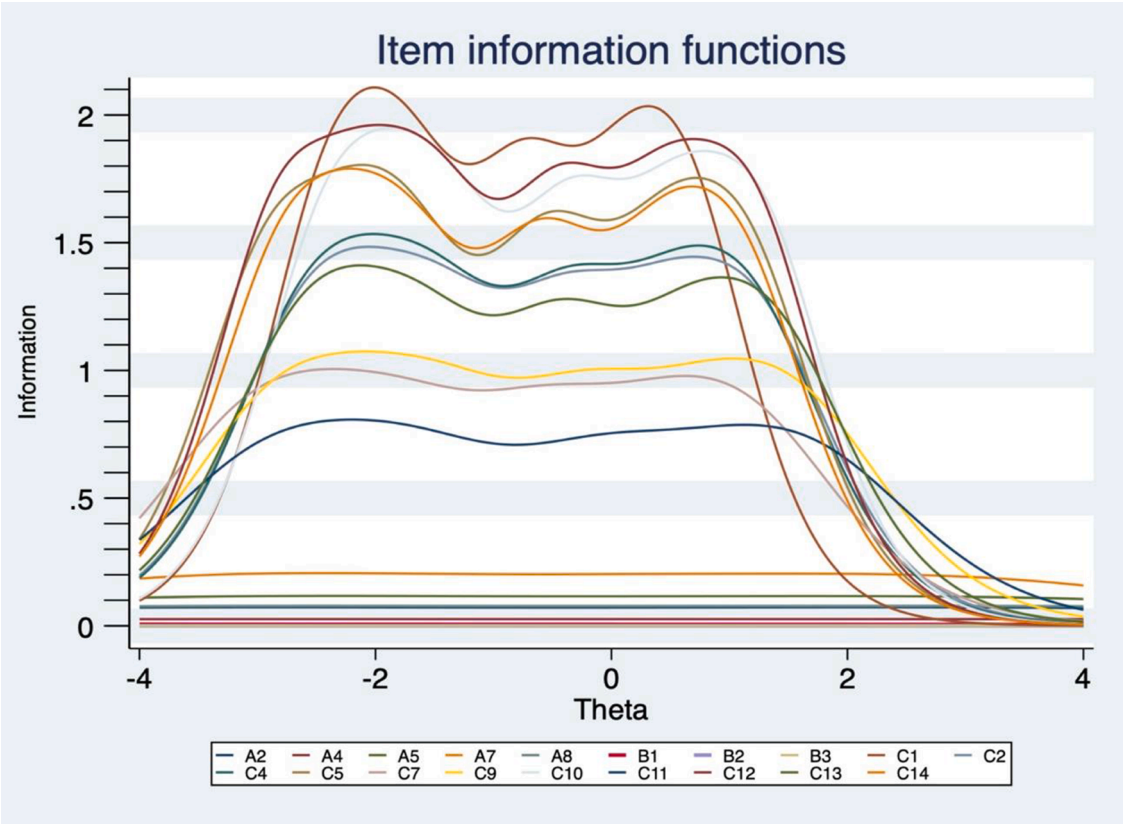


Fig. 10. The relationship between the amount of item information and students' CTD Level
Note: As, Bs and Cs represent the items in L1, L2, and L3.

ineffective in tests, it is preferable to have items with good discrimination in inventories, and thus, emphasis is placed on students' inclination. Consequently, despite their low discrimination, items B1, B2, and B3 were not removed because of the information they provided on students' instant judgment tendencies.

Overall, the three-factor framework, identified by the ISM group based on intelligence from the literature, employers, and employees, is consistent with statistical data. The relationship among these components demonstrates the solution to mediate the restrictions between instant judgment and habitual truth-digging by self-efficacy. The statistical reliability, validity, and test and item endorsement enhanced the truth that this new inventory is ready to be used for assessing undergraduates' CTD.

5. Limitations, educational implications, and future recommendations

5.1. Limitations

Although this study demonstrated positive results for the model fit and acceptable 2ES-CTDI content, it has several limitations. First, the sampling method used in this study is a non-probability sampling that, though having advantages, does have drawbacks in that it can have selection bias and cannot estimate the sampling errors. Second, the framework was subjective. Despite most of the subscales being selected from classical frameworks in previous studies, some parts of the content of our CT disposition framework were a combination of intelligence from employers and employees in China. Furthermore, the framework was developed by the ISM group. Thus, different intelligence groups' contributions to the various categories of the framework should be considered critically. Third, in relation to a theoretical foundation, none of the previous studies that were consulted examined the instant judgment factor. Although this is a new perspective, it lacks firm theories that can be employed to explain it, even though its good model fit and clear relationship with CT disposition were evident. Fourth, although the inventory provided more information for the disposition level θ between -2 and 1 (lower than average), which may be the most effective for the level of students within this interval, if the target participants are different, the validity may vary. Finally, the quality of items needs to be enhanced in relation to discrimination, for example, the instant judgment items.

5.2. Educational implications

This inventory, with a good model fit, can be employed to assess students with lower-average CT disposition levels effectively. This will provide information on students' CT disposition and assist instructors and administrators to enhance their teaching focus and policies, respectively. Furthermore, with regard to the relations among instant judgment, self-efficacy, and habitual truth-digging, it is imperative that teachers guide students to consider a balance of the three factors to realize an optimal CT disposition level. Moreover, this study can be employed to assist educators to develop and test inventories (using ISM) as well as explore the reliability and validity of their instruments with IRT and PLS-SEM.

5.3. Future recommendations

First, in relation to the limitation that the IA and ISM group members were from one part of China, it is recommended that in future studies, participants should include other age groups such as middle school students from other parts of China as well other countries. Second, since the effects of instant judgment are not consistent, it is recommended that future studies place more attention on this issue to determine the relationship among the factors. Third, untraditional approaches such as IRT and PLS-SEM were introduced to enhance the instrument's validity rather than only EFA and CFA. Thus, it is proposed that future studies also use these to provide more evidence for the validity of instruments. Finally, this study is a preliminary study on the instrument. It is recommended that future studies employ this inventory to assess students' CT disposition level but also explore the potentially influential factors such as some demographic factors on CT disposition as well as the relationship with CT skills and outcomes in accordance with the proposed framework in Fig. 1.

6. Conclusions

Intelligence was gathered from employers and employees and the newly developed CT disposition instrument 2ES-CTDI was designed and validated in this study by employing IA, ISM, EFA, CFA, PLS-SEM, and IRT. The results revealed that the content and construct validities, as well as the model fit, were suitable for future use. The results further revealed that the inventory was appropriate to assess lower-average students' disposition. In addition, although the correlations between instant judgment and habitual truth-digging were not significant, they were mediated by self-efficacy. It is, however, recommended that future studies should place more attention on developing a reliable and valid instrument, balancing the competitive mediating effect, and what the influential factors of CT disposition are.

CRediT authorship contribution statement

Yong Liu: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Resources, Software, Visualization, Writing – original draft. **Attila Pásztor:** Supervision, Funding acquisition, Writing – review & editing.

Declaration of Competing Interest

We have no conflicts of interest to disclose.

Data Availability

The data that has been used is confidential.

Acknowledgement

This is to acknowledge that this work is supported by China Scholarship Council (Grant Number: 202008500108), Tempus Public Foundation of Hungary government (Grant Number: SHE-16610-004/2020), the Research Program for Public Education Development of the Hungarian Academy of Sciences (Project Number: KOZOKT2021-16), and University of Szeged (Open Access Grant Number: 5759).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.tsc.2022.101169.

References

Bentler, P. M. (1990). Comparative fit indexes in structural models. *Psychological Bulletin, 107*(2), 238–246. https://doi.org/10.1037/0033-2909.107.2.238

Bentler, P. M., & Bonett, D. G. (1980). Significance tests and goodness of fit in the analysis of covariance structures. *Psychological Bulletin, 88*(3), 588. https://doi.org/10.1037/0033-2909.88.3.588

Bondy, K. N., Koenigseder, L. A., Ishee, J. H., & Williams, B. G. (2001). Psychometric properties of the California critical thinking tests. *Journal of Nursing Measurement, 9*(3), 309–328. https://pubmed.ncbi.nlm.nih.gov/11881271/.

Bravo, M. J., Galiana, L., Rodrigo, M. F., Navarro-Pérez, J. J., & Oliver, A. (2020). An adaptation of the critical thinking disposition scale in Spanish youth. *Thinking Skills and Creativity, 38*, Article 100748. https://doi.org/10.1016/j.tsc.2020.100748

Butler, H. A., & Halpern, D. F (2020). Critical thinking impacts our everyday lives. In D. F. Halpern, & R. J. Sternberg (Eds.), *Critical thinking in psychology* (2 ed., pp. 152–172). Cambridge University Press. https://doi.org/10.1017/9781108684354.008.

Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait-multimethod matrix. *Psychological Bulletin, 56*(2), 81. https://doi.org/10.1037/h0046016

Chin, W. W. (1998). The partial least squares approach to structural equation modeling. *Modern Methods for Business Research, 295*(2), 295–336. https://www.researchgate.net/publication/311766005_The_Partial_Least_Squares_Approach_to_Structural_Equation_Modeling.

Cruz, G., Payan-Carreira, R., Dominguez, C., Silva, H., & Morais, F. (2021). What critical thinking skills and dispositions do new graduates need for professional life? Views from Portuguese employers in different fields. *Higher Education Research & Development, 40*(4), 721–737. https://doi.org/10.1080/07294360.2020.1785401

Dwyer, C., Hogan, M., Harney, O., & Kavanagh, C. (2016). Facilitating a student-educator conceptual model of dispositions towards critical thinking through interactive management. *Educational Technology Research and Development*. https://doi.org/10.1007/s11423-016-9460-7

Elder, L., & Paul, R. (2020). *Critical thinking: Tools for taking charge of your learning and your life* (third ed.). Rowman & Littlefield https://books.google.cg/books?id=yBzzDwAAQBAJ&lpg=PP1&hl=fr&pg=PR5#v=onepage&q&f=false.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J. B. Baron, & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice*. W H Freeman/Times Books/ Henry Holt & Co.

Ennis, R. H. (1996). Critical thinking dispositions: Their nature and assessability. *Informal Logic, 18*(2), 165–182. https://doi.org/10.22329/il.v18i2.2378

Ennis, R. H. (2011). Critical thinking: Reflection and perspective. In Inquiry: Critical thinking across the disciplines. *Prentice-Hall, 26*, 4–18. https://doi.org/10.5840/inquiryctnews20112613

European Statistical Office. (2008). *NACE Rev. 2: Statistical classification of economic activities in the European community*. Luxembourg: Office for Official Publications of the European Communities. https://bit.ly/1EKGS03.

Facione, N., & Facione, P. (1994). Critical thinking disposition as a measure of competent clinical judgment: The development of the California Critical Thinking Disposition Inventory. *The Journal of Nursing Education, 33*, 345–350. https://doi.org/10.3928/0148-4834-19941001-05

Facione, P. (1990a). The California Critical Thinking Skills Test-College level. Technical report# 1. *Experimental validation and content validity*. https://files.eric.ed.gov/fulltext/ED327549.pdf.

Facione, P. (1990b). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction—The Delphi report*. Millbrae, CA: California Academic Press. https://philarchive.org/archive/FACCTA.

Facione, P. A., & Facione, N. C. (1992). *California critical thinking disposition inventory*. Millbrae, CA: California Academic Press.

Finstad, K. (2010). Response interpolation and scale sensitivity: Evidence against 5-point scales. *Journal of Usability Studies, 5*(3), 104–110. https://uxpajournal.org/response-interpolation-and-scale-sensitivity-evidence-against-5-point-scales/.

Fornell, C., & Larcker, D. F. (1981). *Structural equation models with unobservable variables and measurement error: Algebra and statistics*. Los Angeles, CA: Sage Publications Sage CA.

Halonen, J. S. (1995). Demystifying critical thinking. *Teaching of Psychology, 22*(1), 75–81. https://doi.org/10.1207/s15328023top2201_23

Halpern, D. (1998). Teaching critical thinking for transfer across domains: Dispositions, skills, structure training, and metacognitive monitoring. *The American Psychologist, 53*, 449–455. https://doi.org/10.1037/0003-066X.53.4.449

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Holloway, J. A., Johnsen, D. C., & Syrbu, J. (2021). Student performance comparisons for a critical thinking skill set (technology decision-making) for classroom and remote (zoom) facilitation. *Journal of Dental Education, 85*(3), 379–382. https://doi.org/10.1002/jdd.12443

Ip, W. Y., Lee, D. T. F., Lee, I. F. K., Chau, J. P. C., Wootton, Y. S. Y., & Chang, A. M. (2000). Disposition towards critical thinking: a study of Chinese undergraduate nursing students. *Journal of Advanced Nursing, 32*(1), 84–90. https://doi.org/10.1046/j.1365-2648.2000.01417.x

İskifoğlu, G., & Ağazade, A. S. (2013). Translation and validation of a Turkish version of the California critical thinking disposition inventory. *Social Behavior and Personality: An International Journal, 41*, 187–196. https://doi.org/10.2224/sbp.2013.41.2.187

Kahneman, D. (2011). *Thinking, fast and slow*. Macmillan. http://103.38.12.142:8081/jspui/bitstream/123456789/541/1/Thinking%2C%20Fast%20and%20Slow.pdf.

Kaiser, H. F. (1970). A second-generation little jiffy. *Psychometrika, 35*(4), 401–415. https://doi.org/10.1007/BF02291817

Kakai, H. (2001). *The effects of independent and interdependent self-construals on the development of critical thinking dispositions: A quantitative and qualitative study*. University of Hawai'i at Manoa. https://www.proquest.com/openview/78113d8f94b2859cb82cb04489fdcf00/1?pq-origsite=gscholar&cbl=18750&diss=y.

Kakai, H. (2003). Re-examining the factor structure of the California critical thinking disposition inventory. *Perceptual and Motor Skills, 96*(2), 435–438. https://doi.org/10.2466/pms.2003.96.2.435

Kemp, C. G. (1963). Improvement of critical thinking in relation to open-closed belief systems. *The Journal of Experimental Education, 31*(3), 321–323. https://doi.org/10.1080/00220973.1963.11010784

Kline, R. B. (2011). *Principles and practice of structural equation modeling* (3rd ed). Guilford Press. https://pdfs.semanticscholar.org/9f61/4aefcc38de8aa425a190560055dff40fabed.pdf.

Kokdemir, D. (2015). *Decision making and problem solving under uncertainty: Theory and application*. MIT Press. http://web.stanford.edu/group/sisl/public/dmu.pdf.

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. https://doi.org/10.1016/j.tsc.2009.02.001

Lane, D., & Oswald, F. L. (2016). Do 45% of college students lack critical thinking skills? Revisiting a central conclusion of academically adrift. *Educational Measurement: Issues and Practice, 35*(3), 23–25. https://doi.org/10.1111/emip.12120

Lawson, T. J. (1999). Assessing psychological critical thinking as a learning outcome for psychology majors. *Teaching of Psychology, 26*(3), 207–209. https://doi.org/10.1207/s15328023top260311

Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory Into Practice, 32*(3), 131–137. https://www.jstor.org/stable/1476693?seq=1.

Liu, Y., & Pásztor, A. (2022). Effects of problem-based learning instructional intervention on critical thinking in higher education: A meta-analysis. *Thinking Skills and Creativity, 45*, Article 101069. https://doi.org/10.1016/j.tsc.2022.101069

Norris, S. P. (2003). The meaning of critical thinking test performance: The effects of abilities and dispositions on scores. In F. D. J (Ed.), *Critical thinking and reasoning: Current research, theory, and practice* (pp. 315–329). Hampton Press, Inc.

OECD. (2017). Getting skills right: Skills for jobs indicators. *OECD*. https://doi.org/10.1787/25206125

Paul, R. W., & Binker, A. (1990). Critical thinking: What every person needs to survive in a rapidly changing world. *ERIC*. https://eric.ed.gov/?id=ED338557.

Pearson, K. (1900). X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 50*(302), 157–175. https://doi.org/10.1080/14786440009463897

Peng, M., Wang, G., Chen, j., Chen, M., Bai, H., Li, S., Li, J., Cai, Y., Wang, J., & Yin, L (2004). Pipanxing siwei nengli celiangbai de xinxiaodu ceshi yanjiu [Validity and reliablity of the Chinese Critical Thinking Disposition Inentory]. *Chinese Journal of Nursing, 39*(9), 644–647. https://core.ac.uk/download/pdf/78762281.pdf.
Perkins, D. N., Jay, E., & Tishman, S. (1993). Beyond abilities-A dispositional theory of thinking. *Merrill-Palmer Quarterly-Journal of Developmental Psychology, 39*(1), 1–21. http://www.jstor.org/stable/23087298.
Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: a critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879. https://doi.org/10.1037/0021-9010.88.5.879
Quellmalz, E. S. (1987). Developing reasoning skills. In J. B. Baron, & R. J. Sternberg (Eds.), *Teaching thinking skills: Theory and practice*. W H Freeman/Times Books/ Henry Holt & Co.
Quinn, S., Hogan, M., Dwyer, C., Finn, P., & Fogarty, E. (2020). Development and validation of the student-educator negotiated critical thinking dispositions scale (SENCTDS). *Thinking Skills and Creativity, 38*, Article 100710. https://doi.org/10.1016/j.tsc.2020.100710
Rusch, T., Lowry, P. B., Mair, P., & Treiblmaier, H. (2017). Breaking free from the limitations of classical test theory: Developing and measuring information systems scales using item response theory. *Information & Management, 54*(2), 189–203. https://doi.org/10.1016/j.im.2016.06.005
Rusticus, S. (2014). Content validity. In A. C. Michalos (Ed.), *Encyclopedia of quality of life and well-being research* (pp. 1261–1262). Netherlands: Springer. https://doi.org/10.1007/978-94-007-0753-5_553.
Rutter, L., & Brown, K. (2019). *Critical thinking and professional judgement for social work*. Sage.
Salomon, G. (1994). *To be or not to be (mindful)*. New Orleans, LA: American Educational Research Association Meetings.
Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. *Psychometrika, 34*(4), 100. https://doi.org/10.1007/BF03372160
Schmaltz, R. M., Jansen, E., & Wenckowski, N. (2017). Redefining critical thinking: teaching students to think like scientists. *Frontiers in Psychology, 8*, 459. https://doi.org/10.3389/fpsyg.2017.00459. Article.
Shin, H., Park, C. G., & Kim, H. (2015). Validation of Yoon's Critical Thinking Disposition instrument. *Asian Nursing Research, 9*(4), 342–348. https://doi.org/10.1016/j.anr.2015.10.004
Sireci, S. G. (1998). Gathering and analyzing content validity data. *Educational Assessment, 5*(4), 299–321. https://doi.org/10.1207/s15326977ea0504_2
Snedecor, G., & William, G. (1989). *Statistical methods* (8 ed.). Iowa State University Press.
Sosu, E. M. (2013). The development and psychometric validation of a critical thinking disposition scale. *Thinking Skills and Creativity, 9*, 107–119. https://doi.org/10.1016/j.tsc.2012.09.002
Steiger, J. H. (1990). Structural model evaluation and modification: An interval estimation approach. *Multivariate Behavioral Research, 25*(2), 173–180. https://doi.org/10.1207/s15327906mbr2502_4
Taherdoost, H. (2019). What is the best response scale for survey and questionnaire design; Review of different lengths of rating scale/attitude scale/Likert scale. *Hamed Taherdoost*, 1–10. https://ssrn.com/abstract=3588604.
Teo, T. S., Srivastava, S. C., & Jiang, L. (2008). Trust and electronic government success: An empirical study. *Journal of Management Information Systems, 25*(3), 99–132. https://doi.org/10.2753/MIS0742-1222250303
Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika, 38*(1), 1–10. https://doi.org/10.1007/BF02291170
Valenzuela, J., Carracedo, A. M., & Saiz, C. (2011). Critical Thinking Motivational Scale: a contribution to the study of relationship between critical thinking and motivation. *Electronic Journal of Research in Educational Psychology, 9*, 823–848. https://doi.org/10.25115/ejrep.v9i24.1475
Walsh, C. M., & Hardy, R. C. (1997). Factor structure stability of the California Critical Thinking Disposition Inventory across sex and various students' majors. *Perceptual and Motor Skills, 85*(3_suppl), 1211–1228. https://doi.org/10.2466/pms.1997.85.3f.1211
Walsh, C. M., Seldomridge, L. A., & Badros, K. K. (2007). California Critical Thinking Disposition Inventory: Further factor analytic examination. *Perceptual and Motor Skills, 104*(1), 141–151. https://doi.org/10.2466/pms.104.1.141-151
Wang, X., Sun, X., Huang, T., He, R., Hao, W., & Zhang, L. (2019). Development and validation of the critical thinking disposition inventory for Chinese medical college students (CTDI-M). *Bmc Medical Education, 19*. https://doi.org/10.1186/s12909-019-1593-z. Article 200.
Warfield, J. N. (1973a). *An assault on complexity*. Battelle: Office of Corporate Communications.
Warfield, J. N. (1973b). Intent structures. IEEE Transactions on Systems, Man. *Cybernetics*, (2), 133–140. https://doi.org/10.1109/TSMC.1973.5408494
Warfield, J. N., & Cárdenas, A. R. (1994). *A handbook of interactive management*. Iowa State University Press Ames.
Wold, H. O. A. (1985). Partial least squares. In S. Kotz & N. L. Johnson (Eds.), *Encyclopedia of statistical sciences* (Vol. 6, pp. 581-591). Wiley.
World Economic Forum. (2016). *The Future of jobs: Employment, skills, and workforce strategy for the fourth industrial revolution*. https://reports.weforum.org/future-of-jobs-2016/.
Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist, 34*(6), 806–838. https://doi.org/10.1177/0011000006288127
Yeh, M. L. (2002). Assessing the reliability and validity of the Chinese version of the California Critical Thinking Disposition Inventory. *International Journal of Nursing Studies, 39*(2), 123–132. https://doi.org/10.1016/s0020-7489(01)00019-0
Yockey, R. D. (2016). Validation study of the critical thinking dispositions scale: A brief report. *North American Journal of Psychology, 18*(1). https://www.proquest.com/docview/1771238067.
Yoon, J. (2004). *Development of an instrument for the measurement of critical thinking disposition in nursing*. Seoul, Korea: [Doctoral dissertation, Catholic University of Education].