

A cross-national assessment on critical thinking disposition between Hungarian and Chinese undergraduate students

Yong Liu ^{a,*} , Attila Pásztor ^{b,c} , Gyöngyvér Molnár ^{b,c} 

^a School of Foreign Languages, Chongqing Technology and Business University, China

^b Institute of Education, University of Szeged, Hungary

^c MTA-SZTE Digital Learning Technologies Research Group, Hungary

ARTICLE INFO

Keywords:

Critical thinking disposition
Undergraduate students
Cross-national assessment

ABSTRACT

Critical thinking (CT), a higher-order cognitive skill, has gained urgency amid AI advancements. Despite its recognized importance, cross-cultural assessments of critical thinking disposition (CTD) remain scarce, particularly comparisons between non-Western and Western educational contexts. This study addressed this gap by comparing CTD between Hungarian ($N = 1090$) and Chinese ($N = 1535$) undergraduates using Confirmative Factor Analysis (CFA), Measurement Invariance (MI), Item Response Theory, Differential Item Functioning (DIF) and descriptive statistics. Results revealed distinct correlation patterns: Hungarians exhibited strong inter-correlations among self-efficacy, instant judgment, and habitual truth-digging (all p -values < 0.001), whereas Chinese students showed no self-efficacy–habitual truth-digging link ($p > .05$). MI analysis indicated the acceptable configural and metric model (CFI and TLI > 0.90 , SRMR < 0.08 , and RMSEA < 0.10), but failing to achieve the scalar MI requirement ($\Delta\text{CFI} > 0.01$), partially attributable to DIF in specific items, specializing in habitual truth-digging (favoring Hungarian respondents), self-efficacy (favoring Chinese respondents), and mixed pattern for instant judgment. These findings challenge the assumption of CTD universality, suggesting that cultural-educational systems (e.g., Hungary's interdisciplinary focus vs. China's rote-innovation blend) may drive divergent CTD development, underscoring the need for context-specific pedagogical interventions.

1. Introduction

1.1. Framework of critical thinking

The modern concept of critical thinking (CT) was defined in the Delphi Report (Facione, 1990) as the deliberate and self-regulating judgment that resulted in interpretation, analysis, evaluation, and inference, as well as an explanation of the evidentiary, conceptual, methodological, criteriological, or contextual factors that underpinned that judgment, which was quite abstract and complex for educational practice. Accordingly, the concrete CT framework was developed for its assessment.

As shown in Fig. 1, the development of the CT framework went through three stages (a, b, and c). The framework expanded as society evolved, necessitating increasingly complex characteristics of competence for students to adapt to the new circumstances. The

* Corresponding author.

E-mail address: yong.liu@ctbu.edu.cn (Y. Liu).

first phase, shown in Fig. 1 as section (a), began in the mid-to-late twentieth century (the 1960s to 1980s) when CT was uniquely viewed as equal to critical thinking skills (CTS), meaning, CTS established the whole CT framework at the time (Lewis & Smith, 1993).

The second phase, section (b) in Fig. 1, updating from the first single-factor CT framework into a two-dimensional structure, laid a solid foundation for almost all of the following frameworks. This phase began in the 1990s and ended in the 2010s with the publication of the Delphi Report (Facione, 1990). CTS was related to cognitive skills while CTD was more connected to the tendency, inclination, or habit of thinking critically (R.H. Ennis, 1996). Furthermore, CTS and CTD were placed on opposite sides of the balance, and the weight of CTS was greater than that of CTD, the evidence of which was based on a review of literature showing that researchers thought CTS was more important than CTD, and they preferred doing research on CTS (Bravo et al., 2020; Quinn et al., 2020; Sosu, 2013). Despite receiving CTS, they are unlikely to be implemented in the absence of CTD (Valenzuela et al., 2011). It may be argued that, even though CTD was interwoven within the CT framework throughout this time period, indicating its value, researchers still preferred CTS.

In addition, during the second phase, the PSs were removed from CT. Thus, with the openness and acceptance of the debates from all researchers, the third period, section (c) beginning in the 2020s. As a consequence, based on the two-dimensional structure, the framework for the third period was conceptually developed, incorporating PSs and instructional intervention perspectives into the model (Liu & Pásztor, 2022a). As noted, CTS and CTD were positioned at the bottom of the recycling triangle model, with PSs at the top, indicating that CTS and CTD were the primary inputs for CT development and PSs were the output genuine capacity to solve practical everyday life challenges. As each ability increases, more talents emerge as a result of positive recycling. Meanwhile, effective instructional interventions may be implemented in each step to improve the culture of each component, which will be considered as a dynamic recycling CT cultivation framework (Liu & Pásztor, 2022b).

1.2. Framework of critical thinking disposition

As the attitudinal disposition had been included in the concept of CT by Dewey (1910), CTD was also brought into the CT framework as aforementioned at almost the same time. However, this preliminary “CTD framework” only emphasized the disposition of analyzing problems within one’s experience (Dewey, 1910) as a certain concept without any assessments for practice.

On the inspiration of this concept, the quite fundamental and vital CTD framework was put forward in the Delphi Report with 19 sub-dimensions. They were the disposition of prudence in judgments, persistence in difficulties, attention on concerns, reasonableness in criteria, working with complexity, clarity in statements, views revision, diligence in judgments, honesty in facing biases, fair-mindedness, flexibility in alternatives, understanding opinions, self-confidence, trust in reasoned inquiry, precision, alertness to use CT, open-mindedness, well-informed judgments, and inquisitiveness (Facione, 1990). Later on, this 19-factor framework was induced into seven general categories with the creation of the California Critical Thinking Disposition Inventory (CCTDI; Facione & Facione, 1992), which were the basis and reference of almost all the following CTD frameworks. The contents comprised analyticity, inquisitiveness, systematicity, open-mindedness, truth-digging, self-confidence, and maturity. With the development of CTD research, some additional taxonomies of CTD (e.g., R.H. Ennis, 1996; Halonen, 1995; Halpern, 1998; Perkins et al., 1993) were unfolded but were not developed as frameworks into instruments for assessments.

In addition, some later frameworks adapted that of CCTDI. For example, Yoon (2004) condensed and renamed the CCTDI

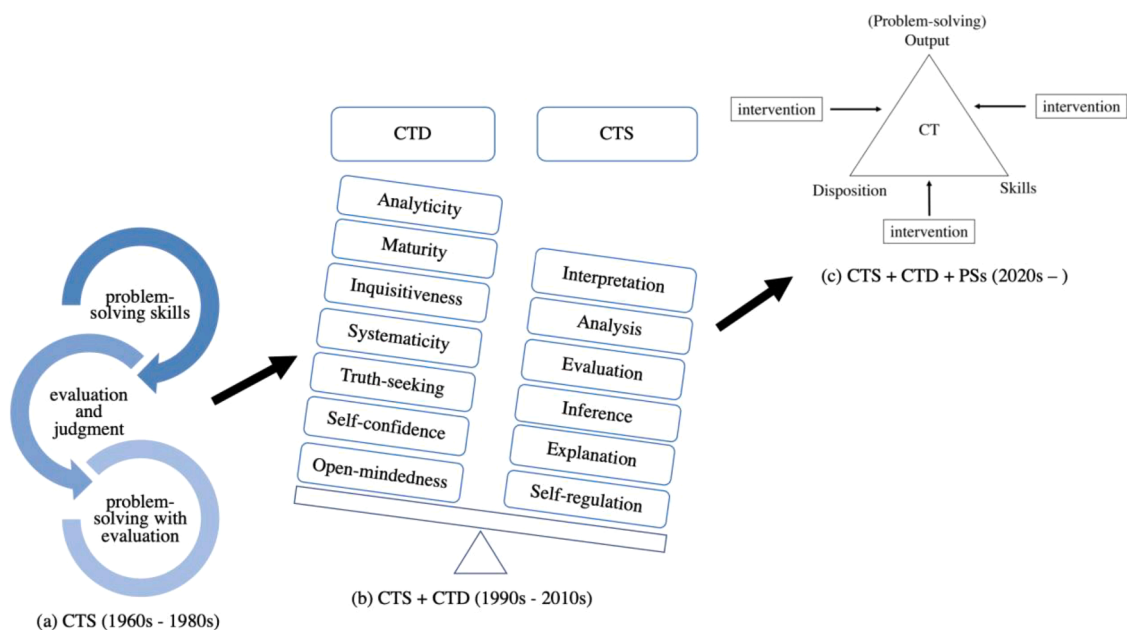


Fig. 1. The development of CT framework.

framework into 27 measures that fall into seven categories: intellectual curiosity, systematicity, prudence, objectivity, self-confidence, healthy skepticism, and intellectual fairness; [Sosu \(2013\)](#) used previous research to shorten and unify the seven factors into two dimensions (critical openness and reflective skepticism), with a total of 11 items, whereas the model fit of a one-factor model appeared better than this two-factor model ([Bravo et al., 2020](#); [Yockey, 2016](#)); and [Quinn et al. \(2020\)](#) integrated the intelligence from teachers and students instead of just previous literature, setting up a CTD framework with open-mindedness, reflection, attentiveness, organization, intrinsic goal motivation, and perseverance.

However, these frameworks were built with the supports from literature, teachers, and students, without the perspectives of the workforce. As undergraduate students are the fresh blood for the energy of the workforce, and there is a competence gap between the outcome of university students and the need of the job market, a CTD framework with the need of the workforce is essential. [Liu and Pásztor \(2022a\)](#), with the inspiration of previous studies, built a three-factor CTD framework under their designed Employ-Employee-Supported Critical Thinking Disposition Inventory (2ES-CTDI) from the intelligence of both educators and students, and employees and employers. As it is the newest and covered most of the angles, this study will adopt it as the assessment instrument.

This framework comprised the disposition of self-efficacy, instant judgment, and habitual truth-digging. As described in [Liu and Pásztor \(2023\)](#), self-efficacy is a disposition that reflects students' intrinsic motivation: self-confidence, self-regulation, persistence, diligence, and maturity, which are connected to individual views about one's ability to control motivation, cognitive abilities, and behaviors ([Bandura & Jourden, 1991](#)). The extrinsic perception in the instant judgment is represented by using fairness, systematicity, and open-mindedness to assess the statement swiftly and accurately—that is, to make rapid judgements. This capacity to make rapid judgements may be handed on to educators as a safe bet when measuring critical thinking ([Leslie & Moilanen, 2010](#)). The tendency to utilize CTS to seek truth avidly and ask questions objectively, for example, to think step by step, examine with evidence, assess ideas, and so on, is referred to as habitual truth-digging ([Facione et al., 1997](#)).

1.3. CTD assessments in and between Hungary and China

The selection of China and Hungary as comparative cases is motivated by their contrasting educational philosophies, cultural values, and limited representation in existing CT research. Chinese education, shaped by Confucian traditions, has historically emphasized rote learning and examination success, though recent reforms aim to foster creativity and critical thinking ([Xie et al., 2025](#)). In contrast, Hungarian education, influenced by Western pedagogical theories, prioritizes student-centered learning and independent reasoning ([Fischer-Dárdai & Kaposi, 2021](#)). These differences provide a theoretically rich context for examining whether CT is culturally universal or context dependent. Additionally, few studies have rigorously tested CT measurement invariance across non-Western and Central European samples, making this comparison empirically novel.

CT gained interest in Hungary in the late 1990s with the practice of the Reading and Writing for Critical Thinking program (RWCT; [Klooster et al., 2001](#)). However, only materials for teaching in schools were prepared in this program ([Bárdossy et al., 2002](#)) without any practical assessments. Later, an assessment on CTS was conducted for 7th and 11th-grade students showing poor performance of it ([Molnár, 2002](#)). Other CT related researches were followed regarding critical reading ([Tóth, 2007](#)), instructors' perceptions on CT instructions ([Fábián, 2014](#); [Kaposi, 2020](#)), and domain-specific CTS assessment of history ([Érsek, 2020](#)). Besides, Hungary also participated in the Organisation for Economic Cooperation and Development- Centre for Educational Research and Innovation (OECD-CERI) project, in which, the CT learning disposition was assessed, demonstrating that Hungarian students had one of the highest self-evaluations of CT among the participating countries ([Vincent-Lancrin et al., 2019](#)). To sum up, currently, CTS have been introduced and conducted relatively widely in Hungary, only learning CTD was assessed by OECD, and these assessments were almost for primary and secondary school students, which indicates that there is a dearth of CTD assessments in Hungary, especially in higher education.

The concept of CT was introduced in China in the late 1980s while the first Chinese version of CTD assessment, the CCTDI, was translated and adapted in 2001, confirming its basic suitability for Chinese undergraduates and making a comparison between Chinese and American undergraduates that Chinese undergraduates' CTD was lower ([Luo & Yang, 2001](#)). However, in view of the lack of localization with cultural issues, another Chinese version of CCTDI (CCTDI-CV) was adapted showing the lower level of digging-truth and systematicity than other dimensions ([Peng et al., 2004](#)). Despite its preliminary target was for undergraduates in nursing, even currently, it was widely used in China for most fields. On its popularity, the CTD results regarding nursing undergraduates were blossomed. For example, [Yeh and Chen \(2003\)](#) compared the nursing undergraduates' CTD between China and America, and it showed that six out of seven domains of Chinese students' CTD were lower except inquisitiveness. The same tendency can also be found in the study of a comparison between Chinese and Australian nursing students ([Tiwari et al., 2003](#)), and between Chinese and American physical normal students ([McBride et al., 2002](#)). However, as reviewed by [Fan and See \(2022\)](#), no difference was found between Chinese and American undergraduates in CTD ([Dennett, 2014](#); [Dennett & DeDonno, 2021](#)). Meanwhile, a study on China, Japan, and Samoa produced mixed results ([Petrini & Kawashima, 2003](#)). Specifically, this study found a difference between Japanese and Chinese undergraduates in CTD, but no difference between Chinese and Samoan undergraduates. More details from the study showed that Chinese undergraduates scored higher than Japanese undergraduates in overall CTD, as well as in the CTD subdimensions of systematicity, analyticity, and self-confidence. In contrast, Chinese undergraduates scored lower than Japanese undergraduates in the CTD subdimensions of digging truth, open-mindedness, inquisitiveness, and maturity.

2. Research gaps, aims and questions

Despite advancements in CT research, CTD remains understudied relative to CTS, despite its foundational role in activating and

sustaining skill development. While CTD frameworks have evolved through abstract conceptualization, validated tool development (e. g., CCTDI), and multi-stakeholder refinement, cross-national assessments of CTD—particularly between culturally distinct regions like Hungary and China—are scarce. Existing studies often neglect measurement invariance (MI) testing, risking culturally biased interpretations of CTD structures and scores. Furthermore, Hungary lacks systematic CTD evaluation, while China's localized assessments suffer from inconsistent methodologies and limited cross-cultural validation. These gaps hinder a comprehensive understanding of how CTD manifests across diverse educational systems and cultural contexts.

This study aims to address the aforementioned gaps by (1) validating the structural fit of the CTD framework in Hungarian and Chinese undergraduate samples, (2) testing MI to ensure cross-cultural comparability of CTD assessments and comparing overall and latent CTD mean differences between the two countries, (3) evaluating the extent to which the assessment tool endorses students' CTD across contexts and identifying country-specific CTD characteristics. By prioritizing CTD—a neglected yet essential component of CT—this research seeks to establish a culturally robust foundation for future cross-national CT studies and pedagogical interventions.

To achieve these objectives, the following research questions (RQs) guide the analysis:

RQ 1: Does the CTD framework demonstrate acceptable model fit for Hungarian and Chinese undergraduate samples?

RQ 2: To what extent does the 2ES-CTDI scale demonstrate measurement invariance (configural, metric, scalar) across Hungarian and Chinese samples, and how do overall and latent mean scores of CTD differ between the two countries?

RQ 3: How does the 2ES-CTDI assessment tool endorse students' critical thinking dispositions (CTD) in Hungarian versus Chinese cultural contexts, and what distinct CTD characteristics emerge among undergraduates in the two countries?

3. Methods

3.1. Research design

Based on the research aims and questions, this study will be conducted (see Fig. 2) with the data collected from both Hungary (HU) and China (CN). After the administration of these datasets, the single-group confirmative factor analysis (CFA) of Hungarian and Chinese samples will be checked to ensure the goodness of model fit for the further analysis. If each CFA model cannot be achieved, the study should be stopped. Oppositely, if the model fit can be acceptable, the measurement invariance (MI) should be adopted to explore the equality of the reflection of students on the used scale for the two countries. If MI cannot be established, the comparisons can be made just regarding the test endorsement and students' characteristics of each country separately, including the test and item suitability and students' ability of each country. However, if MI can be achieved, not only the overall and latent mean of each component can be compared, but also the test endorsement and students' characteristics of each country can be explored.

3.2. Data collection

3.2.1. Ethical approvals

The data collection consent of this study was approved by the university research committee (Approval committee: The

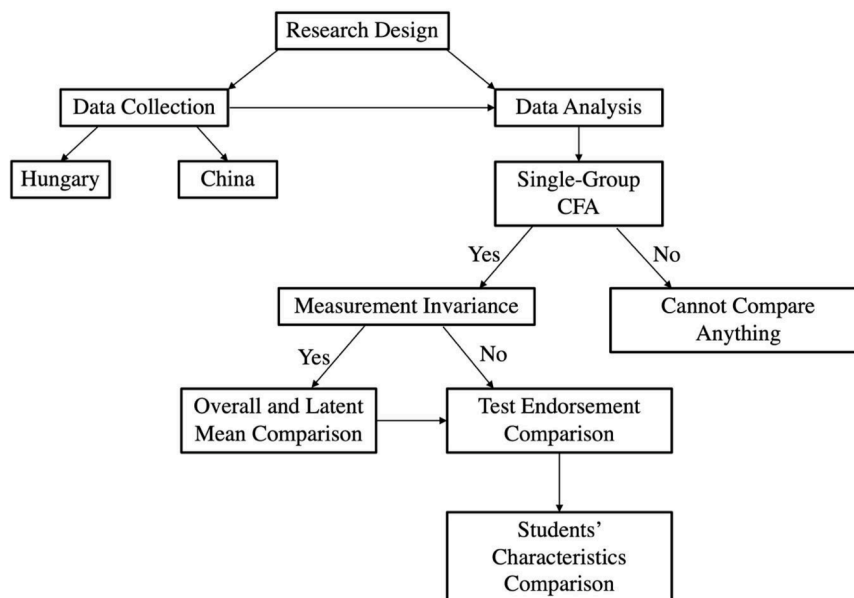


Fig. 2. The flow of research design of this study.

Institutional Review Board of the Doctoral School of Education, University of Szeged. Ethical approval number: 11/2021). All participants in this study agreed to participate voluntarily and anonymously, and the data collected from them were only used for research.

3.2.2. Participants

Table 1 demonstrates all the information of demographic characteristics of sample. The data collection in China was conducted between the end of May and the beginning of June 2022, and that in Hungary was carried out between the end of August and the beginning of September 2022. Participants were recruited via stratified convenience sampling from one university in Hungary (University of X), a comprehensive public institution with strong emphasis on interdisciplinary education and critical thinking development. Similarly, all Chinese participants were enrolled at four universities in Chongqing, China (University of A, B, C, and D). These universities represent a mix of research-intensive, teacher-training, and medical institutions in a major urban center of southwest China. Recruitment occurred through departmental announcements, academic advisors, and course instructors. These procedures were all manipulated via the online questionnaire. The total samples of this study were 2625 undergraduate students from Hungary and China, including 1090 Hungarian undergraduate students (Mean Age = 19.52, SD = 1.795) with 477 males (Mean Age = 19.61, SD = 1.833) and 613 females (Mean Age = 19.45, SD = 1.763), and 1535 Chinese undergraduate students (Mean Age = 20.89, SD = 1.425) with 397 males (Mean Age = 20.82, SD = 1.728) and 1138 females (Mean Age = 20.92, SD = 1.302). Students from both countries were specializing in various disciplines, such as social sciences and arts, natural sciences, medicine, and others. Inclusion criteria required participants to be enrolled as undergraduate students, and fluent in the survey language. Exclusion criteria included exchange students, prior participation in critical thinking training, or incomplete questionnaire responses.

3.2.3. Instrument

The instrument adopted in this study was the Employ-Employee-Supported Critical Thinking Disposition Inventory (2ES-CTDI) which decreased the gap between the outcomes of schools and expectations of the workforce regarding undergraduate students' CTD. 2ES-CTDI is a seven-point Likert scale (from strongly disagree to strongly agree) with 19 items distributed in three domains: self-efficacy (five items), instant judgment (three items), and habitual truth-digging (11 items).

The original 2ES-CTDI was used in the Chinese context with a good reliability and validity (Liu & Pásztor, 2022a), and the validation in the Hungarian context was also done recently to confirm that this tool was also reliable and valid for assessing Hungarian undergraduate students' CTD (Liu et al., 2026). However, for broad and international use, it has been published in the English version (Liu & Pásztor, 2022a). The Hungarian version this study adopted was translated from the English version. The forward-and-back translation approach was carried out for its accuracy (O'Donald & Calia, 2025). Four experts in English-Hungarian translation participated in this procedure, in which two were set as the forward translation from English to Hungarian, and then, two did the back translation from Hungarian to English. Then, the consistency of the English version after this procedure was checked to ensure the adequacy of the translation so that Hungarian students can understand the contents and expressions well.

The reliability and validity of this scale in this study were shown in Table 2, in which, it can be concluded that 2ES-CTDI was also reliable (Cronbach's alpha > 0.70) and valid (CFI and TLI > 0.90, SRMR < 0.08, and RMSEA < 0.10) in this study for further analysis.

3.3. Data analysis

3.3.1. Model fit of each single group

CFA, a facet of structural equation modelling, was adopted to ensure the model fit of the expected construct. The suitability of the model can be investigated with the acceptance criteria of $\chi^2/df < 5$ (Marsh & Hocevar, 1985), CFI and TLI > 0.90, and SRMR < 0.08 (Hu & Bentler, 1999), and RMSEA < 0.08 is considered good fit, 0.08 < RMSEA < 0.10 shows adequate fit, and RMSEA > 0.10 depicts poor fit (Kenny et al., 2015). In addition, chi-square is sample size sensitive (Babak & Green, 2010), and the estimator utilized in this

Table 1
Demographic Characteristics of the Sample.

Characteristic	Hungarian Sample (N = 1090)	Chinese Sample (N = 1535)
Mean Age (Years)	19.52 (SD=1.795)	20.89 (SD=1.425)
Gender Distribution		
Males	477 (Mean Age=19.61, SD=1.833)	(Mean Age=20.82, SD=1.728)
Females	613 (Mean Age=19.45, SD=1.763)	1138 (Mean Age=20.92, SD=1.302)
Field of Study		
Social Sciences and Arts	Included	Included
Natural Sciences	Included	Included
Medicine	Included	Included
Others	Included	Included
Recruitment Period	End of May – Beginning of June 2022	End of August – Beginning of September 2022
Recruitment Method	Stratified convenience sampling from University of X	Recruited from four universities in Chongqing (University A, B, C, and D) via departmental announcements, academic advisors, and course instructors
Inclusion Criteria	Enrolled as undergraduate students, fluent in survey language	
Exclusion Criteria	Exchange students, prior participation in critical thinking training, incomplete questionnaire responses	

Table 2

Reliability and validity of 2ES-CTDI in this study for HU and CN.

	N	Reliability	χ^2	df	CFI	TLI	RMSEA	SRMR
HU	1090	.784	921.167***	149	.921	.909	0.069	0.044
CN	1535	.834	2029.220***	149	.928	.917	0.091	0.045

Notes: *** $p < .001$, CI = 95 %; χ^2 = chi square; df = degree of freedom; CFI = Comparative Fit Index (Bentler, 1990); TLI = Tucker–Lewis index (Tucker & Lewis, 1973); RMSEA = Root Mean Square Error of the Approximation (Steiger, 1990); SRMR = Standardized Root Mean Square Residual (Bentler, 1995).

investigation was weighted least square mean and variance adjusted (WLSMV; Muthén, 1984). The WLSMV is a trustworthy estimator that does not need normally distributed variables and is appropriate for modeling categorical or ordered data (Brown, 2015; Proitsi et al., 2011), therefore chi-square comparison will not function under WLSMV estimation. As a result, chi-square, and its ratio with df will not be included as criteria in this study.

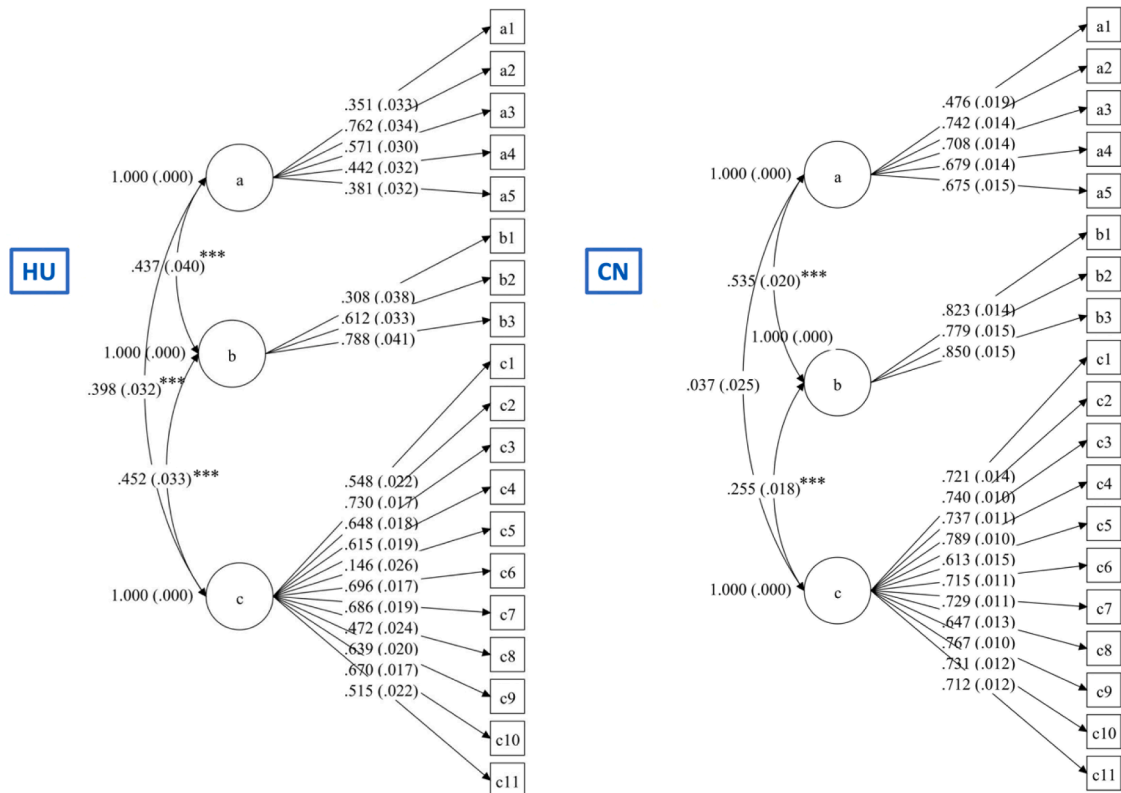
3.3.2. Measurement invariance and latent mean comparison

MI, the prerequisite of latent mean comparisons (Joreskog & Goldberger, 1975; Sörbom, 1974), is a method for determining if there is a substantial difference between groups, settings, and time series in the connections between their latent variables and scale-based indicators (Xu, 2018). MI tests must be performed in the following order: configural (baseline), metric (weak), scalar (strong), and strict invariance models (Chen, 2008; Sorrel et al., 2021). In regard to the estimator used for CFA, the analysis approach of the MI model comparison this study adopted was still the estimator of WLSMV.

Since the sample sizes in this study for both countries are large, more stable values will be adopted additionally to avoid the bias caused by the sample size. These values are the differences of CFI, RMSEA, and SRMR (Δ CFI, Δ RMSEA, and Δ SRMR). The cut points are Δ CFI ≤ 0.01 , Δ RMSEA ≤ 0.015 , and Δ SRMR ≤ 0.03 for the metric invariance model, and Δ CFI ≤ 0.01 , Δ RMSEA ≤ 0.015 , and Δ SRMR ≤ 0.01 for scalar and strict invariance models (Chen, 2007).

3.3.3. Test endorsement and construct characteristics identification

The test endorsement in this study consists of the exploration on the suitability of the difficulty of the scale coordinated with the ability of students, the item characteristics of the whole scale, and these issues in each dimension. To achieve this goal, the single- and

**Fig. 3.** The comparison of correlations of each CTD component between HU and CN.

multi-dimensional partial credit models (PCM; Masters, 1982) of Rasch analysis based on item response theory (IRT) will be employed.

The infit MNSQ provides an evaluation of the sensitivity of people's reaction patterns. While the MNSQ outfit represents the examination of the sensibility of the response with difficulty away from people. Although MNSQs are more dependent on sample size, it is effective for measurements in the 0.5 to 1.5 range (Linacre, 2002). Besides, some strict cut points can also be seen, such as 0.7 to 1.3 (Mok et al., 2006; Zhu & Cole, 1996) and 0.8 to 1.4 (Wolfe & Chiu, 1999). To be more accurate, this study will adopt the strictest one. Discriminations with a value > 0.4 are considered acceptable, whereas those with a value greater than 0.2 are considered borderline (Ebel, 1954).

The identification of constructs' characteristics of students will be analyzed by unidimensional and multi-dimensional PCM analyses for students' CTD characteristics in the whole construct and its sub-dimensions. In addition, the dispositional features were explored by the descriptive percentage of frequency of item options to show the preference on negative or positive sides of CTD, and a simple *t*-test will also be deployed to demonstrate the statistical difference of means of each item between two countries.

To examine potential measurement bias across cultural groups, the Differential Item Functioning (DIF) analysis was conducted using logistic regression models (Magis et al., 2012). Uniform DIF was tested by comparing nested models with and without group membership, while non-uniform DIF was evaluated by including an interaction term between group and ability. Items were flagged for DIF if the group effect reached statistical significance ($p < .05$) and the effect size (McFadden's pseudo R^2) exceeded 0.02. Items with effect sizes > 0.10 were classified as exhibiting severe DIF, warranting special attention in cross-cultural interpretation. The Chinese sample was specified as the reference group (coded 2), and the Hungarian sample as the focal group (coded 1). The resulting odds ratios (OR) indicate the relative likelihood of Hungarian students endorsing a higher response category compared to Chinese students of equal ability level. An OR > 1 indicates the item favors Hungarian respondents, while an OR < 1 indicates it favors Chinese respondents.

4. Results

4.1. Model fit of each group

The model fit of each group (Cronbach's alpha > 0.70 , CFI and TLI > 0.90 , SRMR < 0.08 , and RMSEA < 0.10) indicates an acceptable model construction for both HU and CN, which sets a robust basis for the exploration of the MI between them. In addition, Fig. 3 showed the correlations among each sub-domain of CTD and factor loadings of each item to the latent variables of both countries generated from CFA. Based on CFA models of two countries, the significantly positive correlations among each component of CTD (a: self-efficacy; b: instant judgment; c: habitual truth-digging; $***p < .001$) for Hungarian undergraduates could be seen (all *p*-values < 0.001), whereas those for Chinese, it seemed that there was no correlative path between a and c ($p > .05$), whereas b positively correlated with c and a positively correlated with b (both *p*-values < 0.001), indicating that for Hungarians, all factors of CTD influence each other actively in this positive cycle, while for Chinese, the opposite interaction exists in the relationship between self-efficacy and habitual truth-digging, and self-efficacy may interact with habitual truth-digging via instant judgment, which may arouse the interest to explore the mediating role of instant judgment in further study. Besides, the factor loadings of item C5 for Hungarian samples seemed lower, which leads the deeper analysis of this item later.

4.2. Measurement invariance across each country

On the basis of the established single-group CFA models of both countries, the check on MI between HU and CN starts. Table 3 showed the comparisons by nation among configural and metric models. It can be concluded that the configural and metric MI models could be achieved with acceptable model fits (CFI and TLI > 0.90 , SRMR < 0.08 , and RMSEA < 0.10). Despite the ignorance of the significance of chi-squares comparisons because of its sensitivity of sample size, the MI requirements cannot be fully achieved. The part of the differences between configural and metric models are within the criteria (Δ RMSEA ≤ 0.015 and Δ SRMR ≤ 0.03), but the distance between the CFIs of both models is out of the required range (Δ CFI ≤ 0.01). These results demonstrated that undergraduates from HU and CN may reflect the construct of the scale differently despite the existence of the acceptable reliability and validity, and some distinctions behind may exist and need to be explored, such as the cultural issues, the degree of the CTD instructional development, the maturity of the CTD assessment, and so on. Thus, the comparison of the overall CTD and their latent means between both countries cannot be conducted without the at-least partial scalar invariance model.

In addition to the MI analysis, a partial MI analysis was conducted to identify specific items that may contribute to the lack of full MI. The partial scalar model allowed for certain item intercepts to vary freely between the two groups while constraining others to be equal.

Table 3
Comparisons by nation among configural and metric models.

Model	$\chi^2 (\Delta\chi^2)$	df	CFI	TLI	RMSEA	SRMR	Δ CFI	Δ RMSEA	Δ SRMR
a. Configural	2944.72***	298	.932	.922	.082	.044			
b. Metric	3437.31***	314	.910	.900	.087	.053			
a VS b	419.28***	16					.022	.005	.009

Notes: *** $p < .001$, CI = 95 %.

Partial scalar analysis revealed that while the full scalar MI model could not be achieved, several items demonstrated strong measurement invariance. Specifically, items related to self-efficacy (e.g., A1, A2, A3) and instant judgment (e.g., B1, B2) showed consistent factor loadings and intercepts across the two groups, indicating that these items function well in both cultural contexts. However, certain items within the habitual truth-digging domain (e.g., C5, C11) exhibited significant differences in intercepts, suggesting that these items may not perform equally well in both samples.

The partial MI model fit statistics indicated an improvement over the configural and metric models, with a decrease in chi-square values and an increase in CFI and TLI values. Specifically, the partial scalar MI model yielded a chi-square value of 1563.420 ($df = 318$, $p < .001$), a CFI of 0.898, and a TLI of 0.890. Although these values do not meet the stringent criteria for full MI, they suggest that partial MI can be achieved by allowing certain item intercepts to vary.

4.3. Test endorsement and construct characteristics within each country

To figure out the endorsement of the inventory, Fig. 4 showed the comparison of relationships between students' ability and expected score, in which, with the increase of students' ability level, the expected scores increased, indicating the scale can reflect the students' CTD level exactly with higher ability level students obtaining higher expected scores.

Besides, Fig. 5 depicted the degree of information provided by the scale with students' ability, in which, the most information the scale can provide was for the ability level around -2 students for both countries, indicating this scale is more suitable for students with lower-average level and this inventory may be a little easier for students.

Digging into the endorsement of items (see details in Appendix, Table S1), all the infit and outfit MNSQs for both countries were within the strictly productive range of 0.8 to 1.4 (Wolfe & Chiu, 1999), showing the sensibility of the participants' reflection and items' difficulty was not fluctuated largely, and the items were dependable. Although the discriminations of items for Chinese samples were acceptable, that of one of the items (C5) in Hungarian samples was not good enough with the value of 0.18, indicating this item cannot differentiate the students' CTD level. If the assessment is a test, this low discrimination cannot be allowed. However, the assessment in this study was an inventory used to collect the disposition. The same tendency of students can also provide valuable information on students' CTD characteristics. For item C5 "I prefer raising doubts while others are in a calm mood", most students chose "slightly agree", proving that they may be consistent with this statement for questioning while others were calm.

Besides, to combine the characteristics of items and participants together, in Fig. 6, the item-person map can show both the overview and the detailed information about the scale and items. The left side of the item-person map listed the distribution of participants, and the right side distributed the items. The ability of students and the difficulty of items decreased from the top to the bottom of the map.

Therefore, according to Fig. 6, the distributions of Hungarian and Chinese students approximate a normal curve, indicating that the 2ES-CTDI scale effectively differentiates students' CTD levels. From the perspective of the distribution of items, most of the items in both Hungarian and Chinese samples located at the bottom, indicating that items for undergraduates of two countries were easy to solve, and it will need more harder items for the higher-level CTD students.

Furthermore, the difficult items for Hungarians were items 5, 3, 13, and 1, and those for Chinese were items 1, 2, 3, 4, 5, most of

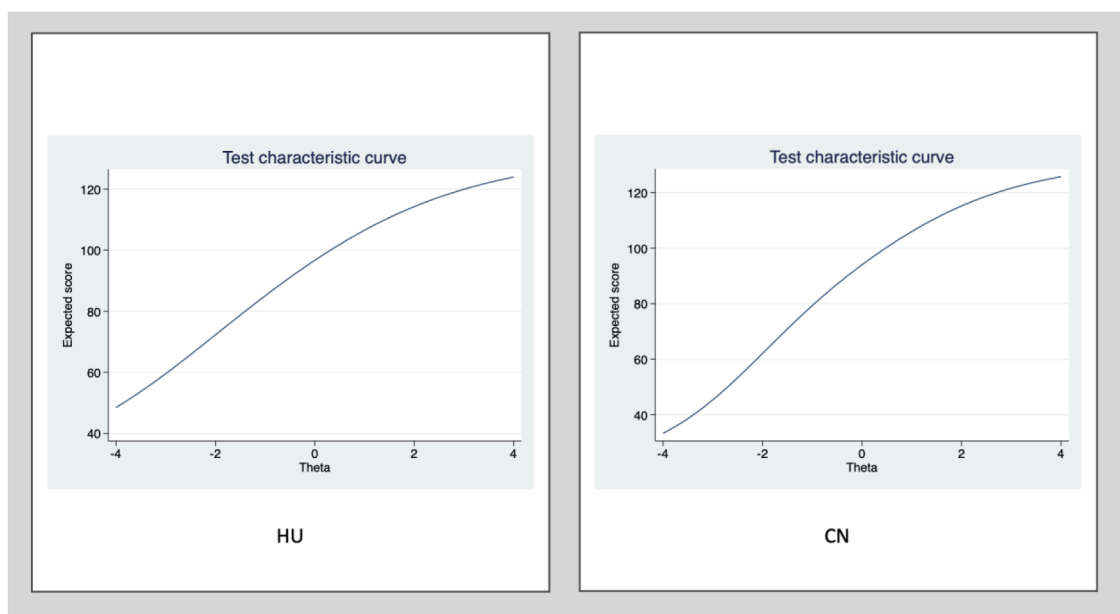


Fig. 4. The comparison of relationships between students' ability and expected score.

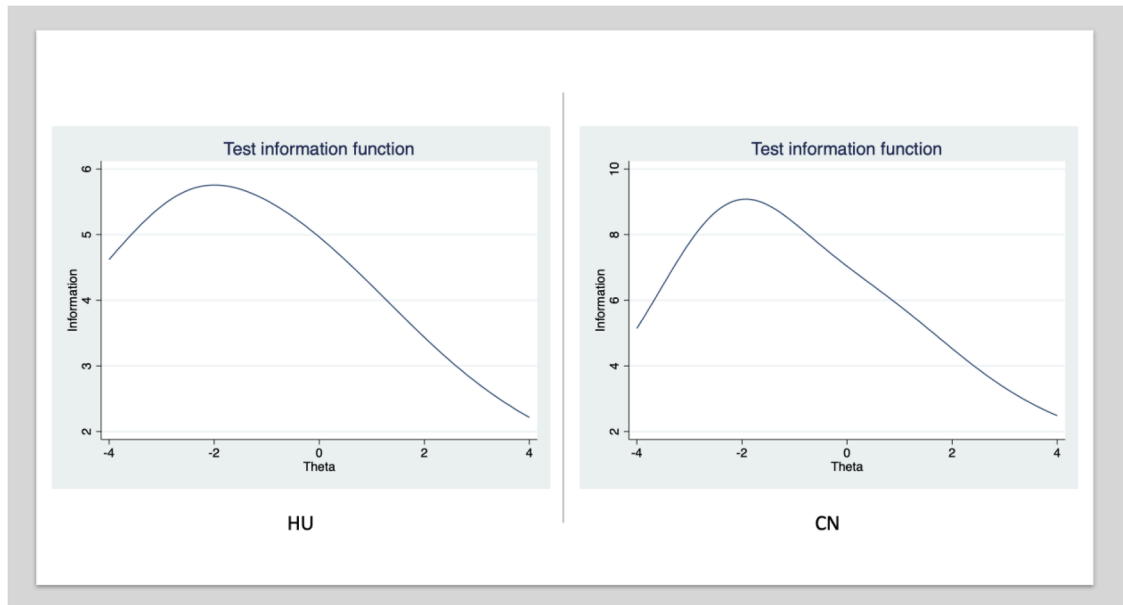


Fig. 5. The degree of information provided by the scale with students' ability.

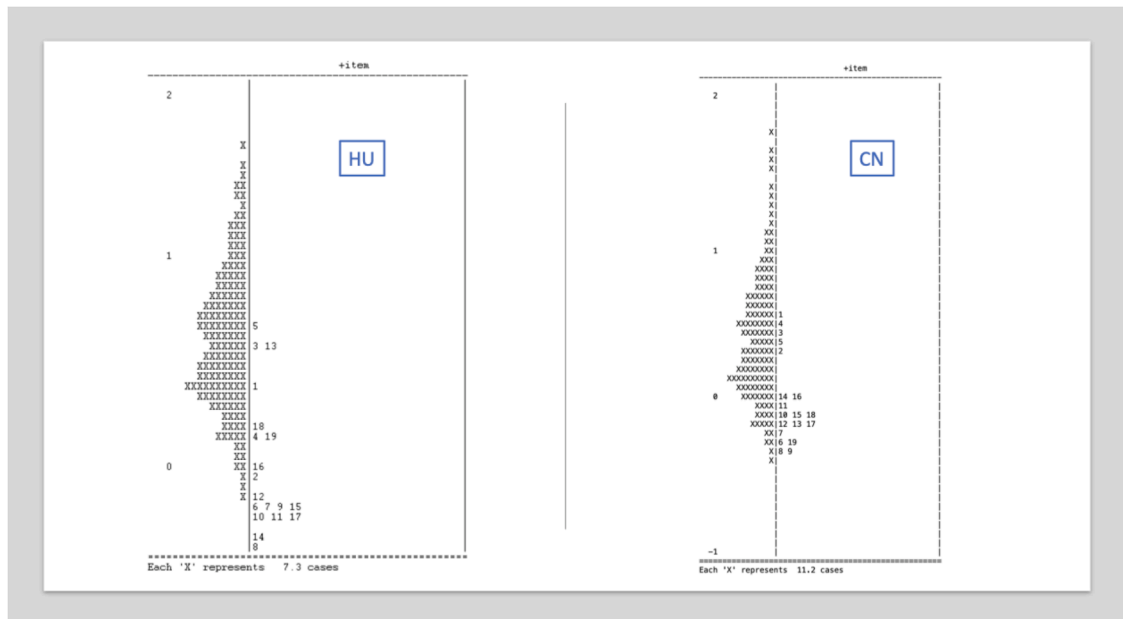


Fig. 6. The item-person maps of HU and CN.

which for both countries were items of self-efficacy, indicating the scores of self-efficacies in Hungarian and Chinese students were low and they lacked CTD self-efficacy. However, the items of instant judgment and habitual truth-digging for Hungarian distributed widely at the lower part, indicating they are easier for Hungarians, and Hungarians' instant judgment and habitual truth-digging levels may be similar but better than self-efficacy. While for Chinese, items (6, 7, and 8) of instant judgment were listed at almost the lowest part than those of habitual truth-digging, showing the lower disposition exists in Chinese undergraduates to make fast judgment adequately than digging truth habitually.

This conclusion can also be found in the item function curves in Fig. 7 that items B1, B2, B3 (items 6, 7, and 8) can provide more information for lower-level students for Hungarians and Chinese, demonstrating that they were the easiest items. Besides, the most degree of information was provided by item B3 (item 8) for lower ability students, also depicting item 8 was the easiest for students of both countries.

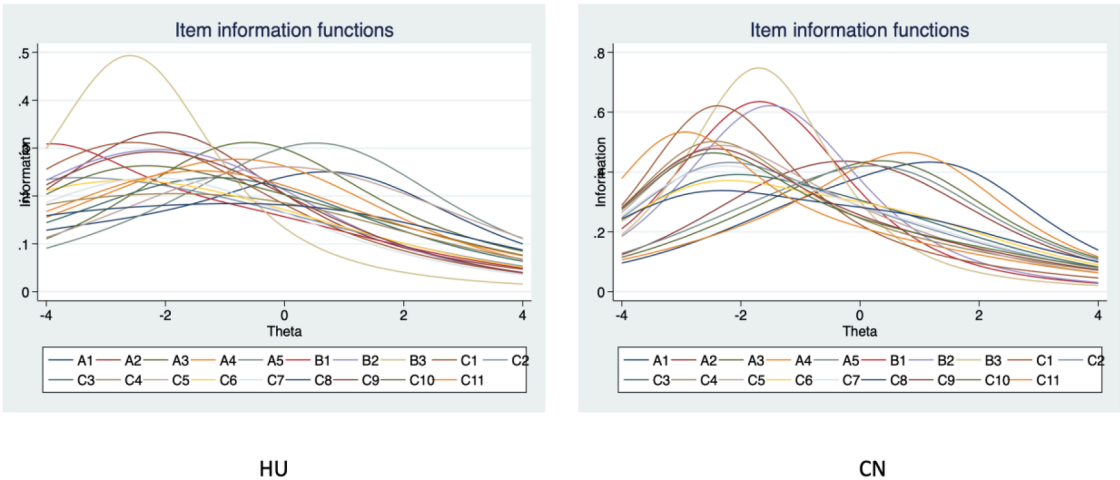


Fig. 7. The degree of information provided by items with students' ability.

In addition, Fig. 8 demonstrated the comparison of the level of CTD components in both countries, in which, the Hungarian samples distributed on the left, Chinese samples located on the right, and dimension 1, 2, and 3 represented self-efficacy, instant judgment, and habitual truth-digging, respectively.

Fig. 8 showed that for Hungarian undergraduate students, most of the persons in the dimension 2 distributed at the upper part of the scale, and most of them in dimension 3 sat at the middle part, while those locations in dimension 1 were almost at the bottom, indicating that the disposition level of instant judgment of Hungarians was, to some extent, higher than self-efficacy and habitual truth-digging, and habitual truth-digging was a little better than self-efficacy, which was similar as the previous results. In addition, for Chinese samples, the persons of self-efficacy and instant judgment distributed almost same but self-efficacy was a little upper than instant judgment, and those in habitual truth-digging located at the top part, indicating that the dispositional habitual truth-digging level was higher than other two, and self-efficacy may be a bit stronger than instant judgment for Chinese undergraduate students.

Furthermore, to figure out the dispositional characteristics of students from both countries, the percentage of the frequency of positive and negative selections and mean difference of each item were calculated (see details in Appendix, Table S2). The options of 1

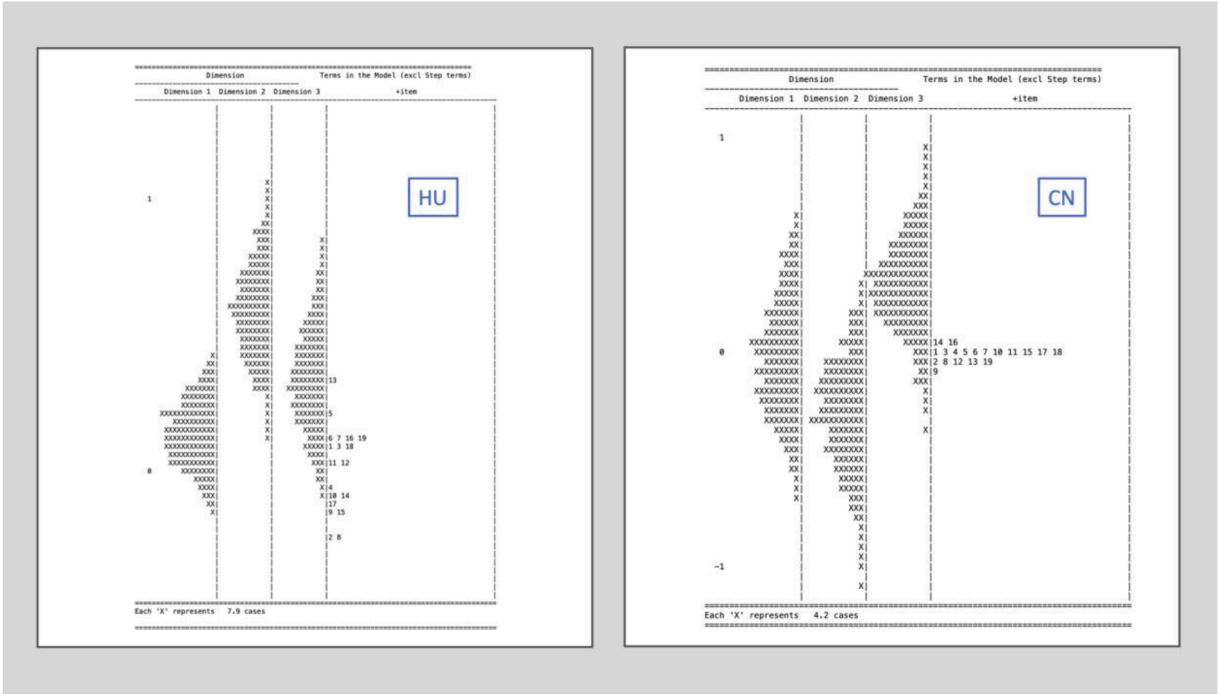


Fig. 8. The multi-dimensional Rasch analyses of Hungarian and Chinese samples.

to 7 represent strongly disagree, disagree, slightly disagree, neutral, slightly agree, agree, and strongly agree. Besides, for items A1 to B3, they were reversed items, thus, their scores were reversed for the calculation of means. In addition, to figure out the negative and positive sides clearly, options 1 to 3 were categorized as negative and 5 to 7 as positive, and option 4 was omitted because of its no meaning.

Digging into each item, for those in self-efficacy, the large quantity of Hungarian students chose the negative side except item A5, while the large portion of Chinese students showed the positive tendency to it except item A2, indicating that Hungarian undergraduates were self-confident and persistent to think critically.

For items in instant judgment, the majority of students of both countries tended to choose the negative side proving that most of them were facilitated with exactly instant judgment of CTD and could be open-minded and fair for everyone.

In regard to habitual truth-digging, most of students in both countries selected the positive side, indicating the majority of them had the disposition to think logically, reasonably, and systematically. However, for item C5 in Hungarian samples, the proportions of choosing negative and positive sides were almost same, depicting that most of Hungarian undergraduates would not care about others' mood when proposing doubt, which was quite different from Chinese undergraduates who tended to question others when they were calm rather than when they were in anger.

DIF analysis revealed substantial cross-cultural measurement variance across several items (Table S3 in Appendix). Severe DIF (effect size > 0.10) was identified in multiple items, with particularly pronounced effects in the Habitual Truth-Digging dimension. Most notably: Item C10, OR = 17.05, strongly favoring Hungarian respondents ($p < .001$); Item B1, OR = 0.012, strongly favoring Chinese respondents ($p < .001$); and Item A1, OR = 0.14, moderately favoring Chinese respondents ($p = .0003$).

The Item Characteristics Curves (ICC, Figure S1 in Appendix) visually demonstrate these disparities, with Hungarian respondents showing substantially higher probabilities of endorsing highest ratings on C10 across all ability levels (Fig. 9), while Chinese respondents display the reverse pattern for B1 (Fig. 9). Non-uniform DIF was also observed in items B2, B3, and C8, indicating that group differences vary across ability continua.

Item C5 exhibited a complex pattern of severe DIF. It demonstrated extreme uniform DIF favoring Chinese respondents (OR = 0.052, $p < .001$), indicating they were substantially more likely to endorse this item. Furthermore, a significant non-uniform DIF effect was detected ($p < .001$), suggesting that the magnitude of this cross-cultural difference varied across the ability continuum. This item also showed low discrimination within the Hungarian sample in the Rasch analysis, aligning with its unstable measurement properties across groups.

Patterns varied significantly by dimension (Table 4). Self-Efficacy was the most stable (40 % DIF), Instant Judgment showed a complex mixed pattern (100 % DIF), and Habitual truth digging was the most culturally sensitive, with 82 % DIF strongly favoring Hungarian respondents. The results indicate that while DIF is widespread across the scale, its nature and severity are dimension-specific.

These DIF findings align with and elaborate upon the measurement invariance results, suggesting that scalar non-invariance arises from specific item-level cultural biases rather than general construct incomparability.

5. Discussion

The primary aims of this study are to investigate the CTD framework model fit and the MI models between HU and CN (RQs 1 and 2). The follow-up purposes are: if both of them could be achieved, the overall and latent CTD mean comparisons could be carried out (RQ 2); if these models cannot be obtained, the test endorsement and students' characteristics between two countries will be explored (RQ 3). Thus, based on these research objects and the obtained results, the expected research questions will be discussed below:

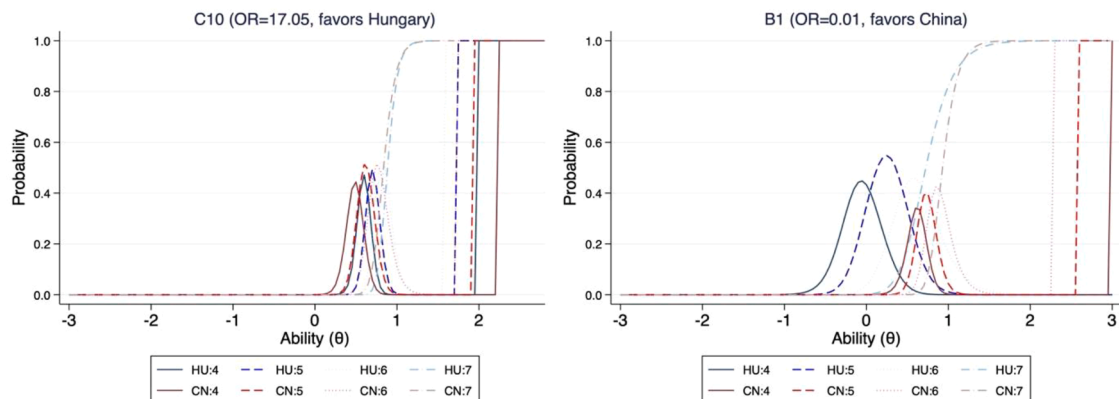


Fig. 9. ICC curves for Items C10 and B1.

Table 4
DIF Patterns by Dimension.

Dimension	Total Items	Items with DIF	Severe DIF Items	Primary Direction	Key Characteristics
Self-Efficacy	5	2 (40 %)	2 (40 %)	Favors China	A1 and A4 show uniform DIF favoring Chinese respondents Moderate to small effect sizes Relatively stable cross-culturally
Instant Judgment	3	3 (100 %)	1 (33 %)	Mixed pattern	B1 shows extreme DIF favoring China (OR=0.012) B2 and B3 show non-uniform DIF Dimension shows complex cross-cultural functioning
Habitual Truth Digging	11	9 (82 %)	8 (73 %)	Favors Hungary	8 items favor Hungarian respondents C10 shows extreme DIF (OR=17.05) Only C5 favors Chinese respondents Most culturally sensitive dimension
Overall	19	14 (74 %)	11 (58 %)	Mixed	Widespread DIF across the scale Habitual Truth Digging most affected Self-Efficacy most invariant

5.1. Discussion on model fit (RQ 1)

The model fits of each country were acceptable enough to show the stability of the structure of 2ES-CTDI, which is the basis for any further analyses, especially for MI exploration. These results were similar to the previous studies regarding the this instrument used in Chinese and Hungarian samples (Liu & Pásztor, 2022a; Liu et al., 2026).

The results of the preliminary comparison between Chinese and Hungarians demonstrated that no correlation between instant judgment and habitual truth-digging for Chinese students while positive significance of that for Hungarians. For Chinese students, it may be influenced by the Chinese Golden Mean thoughts that students did not make rash decisions and did not hesitate, which demonstrated a balanced inclination to think critically. However, for Hungarians, CT has been introduced since the very early period as many western countries did, which strengthened students' CT in almost all aspects, including nearly all facets of CTD. Thus, they have both stronger truth-digging tendencies and the inclination to make fast decisions. This phenomenon can also be explained by Hofstede's cultural dimensions theory (Hofstede & Bond, 1984), which includes power distance, uncertainty avoidance, individualism versus collectivism, and masculinity versus femininity, offer a structured framework for examining cultural influences on cognitive dispositions. For example, high power distance cultures may emphasize respect for authority and conformity, potentially influencing students' inclination towards critical thinking and their openness to questioning established norms. In contrast, low power distance cultures may encourage more independent thinking and challenge to authority, fostering a stronger disposition towards critical thinking.

5.2. Discussion on MI and means comparisons (RQ 2)

Although the configural and metric MI models could be achieved, the difference between them was not fulfilled the requirement, which causes the terminal of the overall and latent means comparison between countries. Despite a recent study illustrated that MI was not the basis for latent means comparisons (Robitzsch & Lüdtke, 2023), it violated the common sense, and the conclusion will need more evidence to be supported in the future. Thus, this study still adopted the thought that MI was the prerequisite of latent means comparison.

The findings from the partial scalar analysis provide important clues about the cross-cultural functioning of the CTD inventory. The consistency of self-efficacy and instant judgment items across the two groups suggests that these constructs are relatively stable and can be reliably measured in both Hungarian and Chinese undergraduate students. This stability may be attributed to the universal nature of these cognitive dispositions, which are less influenced by cultural differences.

In contrast, the variability observed in certain habitual truth-digging items highlights the potential impact of cultural factors on students' approaches to seeking and evaluating information. The differences in item intercepts may reflect cultural norms, educational practices, or societal expectations that influence how students engage in critical thinking processes. For instance, the item "I prefer raising doubts while others are in a calm mood" (C5) showed significant differences between the two groups, indicating that cultural context may affect students' willingness to challenge others' viewpoints.

This study cannot make an exact and adequate explanation of these distinctions based on previous studies due to the complexities of identifying the influencing factors between countries, which is not the focus of this study. However, the reasons why Hungarians and Chinese reflected the scale differently need to be unfolded with a plausible explanation stated here. Fan and See (2022) provided a structured review of existing evidence comparing Chinese students' CTD with other international cohorts and highlighted both similarities and differences, attributing some of the variations to cultural factors such as educational philosophy and societal expectations. Wu (2019) and Wu and Molnár (2021) conducted cross-national studies on Chinese and Hungarian primary students' inductive reasoning and problem-solving skills, respectively, showing that the differences between these two countries were, besides the cultural distinctions between the East and West, probably caused by learning and teaching strategies, methods, and focuses. Although their participants were primary students, they provided some clues for the interpretations of the results in this study that these, including cultural distinctions, may the factors have affected students' cognition on CTD. Besides, the differential item functioning (DIF) is also a

potential factor. Although the focus of this study is not on the item-level adaptation of the instrument, the possible clue for further study can be shown on this issue. Even so, the construct's manifestation difference is also the effect of different cultures.

5.3. Discussion on endorsement and dispositional characteristics (RQ 3)

The scale endorsement was good enough for both Hungarians and Chinese, which is consistent with the previous validity results adopting 2ES-CTDI in Hungarians and Chinese (Liu & Pásztor, 2022a; Liu et al., 2026). It can be explained that despite the existence of outliers (such as item C5 for Hungarian version), it will not affect the scale's function and the difficulty and discrimination of the whole scale and items are acceptable for future use. To make it clear, item C5 was originally designed to assess students' willingness to challenge ideas or situations even when it might be socially uncomfortable or when others are not expecting dissent. In the Hungarian educational context, which emphasizes independent reasoning and student-centered learning (Fischer-Dárdai & Kaposi, 2021), this disposition is relevant. Although the item showed low discrimination in our sample, it might still capture important nuances of CTD that are culturally specific. Besides, Hungarian classrooms often encourage debate and critical engagement, even if it means questioning authority or prevailing norms. Item C5 could reflect this cultural tendency to speak up regardless of the social context, which is a valuable aspect of CTD. In addition, the item aligns with the broader theoretical framework of CTD, specifically the dimension of "habitual truth-digging," which involves a persistent and systematic pursuit of truth. The willingness to propose suspensions or challenge ideas in uncomfortable situations reflects a deep-seated habit of seeking truth, regardless of the social context. Retaining this item allows this study to capture a wider range of behaviors associated with habitual truth-digging, even if those behaviors are less common or more context-dependent.

Self-efficacy was weak for both, instant judgment was higher than other two for Hungarians, and habitual truth-digging was stronger than other two for Chinese. The similar trends can be found in previous studies despite the use of different inventories. For example, the study (Huang et al., 2021) for Chinese medical undergraduates showed the CTD mean scores in inquisitiveness and analyticity were higher than other domains (no statistical difference results, just descriptive map). Although truth-digging score in that study was low, the habitual truth-digging in this study includes truth-digging, inquisitiveness, and analyticity, and if integrated, the CTD means in this domain will also be higher than other two subdimensions, which is also, to some extent, echoed in the study of Ma (2021). The plausible explanation on these results regarding Chinese undergraduates is that they may dig truth habitually with curiosity and enthusiasm but without persistence and self-confidence to reach the destination (Zhang et al., 2009). Besides, Fan and See (2022) found that Chinese students often exhibit lower levels of self-efficacy in CTD compared to their Western counterparts, potentially due to cultural values emphasizing humility and collective harmony over individual assertion. In contrast, Hungarian students, influenced by Western pedagogical theories prioritizing independent reasoning (Fischer-Dárdai & Kaposi, 2021), may demonstrate stronger self-efficacy and instant judgment in CTD.

In addition, although a study explored the educational culture in CT in Hungary showing that individuals will not be able to practice critical thinking in class due to the classroom learning environment's dominance of collectivist elements, and students will not have many opportunities to see critical thinking behaviors in action (Fábián, 2015), and a validity assessment on CTD conducted for Hungarians (Liu et al., 2026) showing the suitability of CTD assessment in Hungary, no studies explored the CTD components and the comparisons in Hungarian undergraduates. Thus, the results in this study were the basic overview on Hungarian undergraduates' CTD level and characteristics.

Hungarian undergraduates appeared to exhibit higher levels of self-confidence and persistence in critical thinking compared to their Chinese counterparts. However, both groups of students demonstrated open-mindedness and fairness in their interactions. Additionally, the majority of students from both countries showed a disposition towards seeking truth in a logical, reasonable, and systematic manner. Although some studies (e.g., Bixler et al., 2015; Tiwari et al., 2003; Yeh & Chen, 2003) made the comparisons between Chinese and other countries without MI said that the Chinese undergraduates' CTD may be weaker or lower than Western students, this study just did the in-line comparisons between two countries to show the feathers of each country instead of the cross-national comparisons to show difference between both since the MI models cannot be achieved. Even so, the fruitful conclusions can also show a better map in terms of undergraduates' CTD levels of each country.

The DIF analysis provides granular insight into which specific aspects of CTD manifest differently across cultures. The extreme DIF observed for item C10 (OR=17.05) suggests that habitual truth-digging may be more culturally valued or linguistically salient in Hungarian educational contexts, which emphasize independent inquiry (Fischer-Dárdai & Kaposi, 2021). Conversely, item B1's strong favoring of Chinese respondents (OR=0.012) may reflect culturally instilled values of rapid, harmonious decision-making aligned with Confucian educational traditions (Fan & See, 2022). Importantly, DIF was not uniformly distributed across dimensions. The habitual truth-digging dimension contained the highest concentration of severe DIF items, suggesting this construct may be particularly sensitive to cultural interpretation. In contrast, self-efficacy items showed more moderate DIF, indicating greater cross-cultural consistency in self-assessment of critical thinking capabilities. These item-level analyses complement the MI findings by identifying specific measurement points requiring cultural adaptation while preserving the overall structural validity of the 2ES-CTDI framework.

6. Educational implications

Cross-national study on the CTD of undergraduate students from Hungary and China might provide important insights into the educational systems, pedagogical approaches, and cultural impacts on students' CT. (a) Having identified the differences may be caused by cultural differences and instructional methods. CTD is influenced by cultural variables, which must be understood. Learning preferences, communication preferences, and collectivism versus individualism differences can all have an impact on how students

develop their CTD. When developing curricula and pedagogical strategies, educators should take these aspects into account. (b) Educators can tailor the curriculum to address the specific strengths and weaknesses of students from each culture in light of the study's findings. For instance, if Chinese students show less inclination toward CTD self-efficacy and instant judgment, curriculum designers might introduce more tasks into the coursework for the improvement of these two factors. Specifically, teachers could assign weekly debate sessions where students must quickly formulate and articulate their viewpoints, gradually building their self-efficacy and instant judgment inclinations. (c) Cross-national studies can be included into teacher preparation programs to better equip educators to teach CTD while taking into account the particular requirements and traits of students from various cultural backgrounds. For example, during teacher training, workshops could be conducted where participants engage in simulated cross-cultural classroom scenarios, practicing how to adapt their teaching styles to accommodate diverse student needs. (d) The findings can lead to a reevaluation of assessment methods. If the two countries' students are found to excel in specific types of CTD but struggle with others, assessments can be designed to provide a more comprehensive picture of CTD. Teachers could implement a portfolio assessment system that includes a variety of tasks, such as essays, group projects, and oral presentations, to capture a broader range of CTD. (e) Educational institutions in both countries can engage in cross-cultural collaboration to share best practices and learn from each other. Exchange programs, joint research projects, and collaborative initiatives can be established to promote cross-cultural understanding and enhance critical thinking development. For instance, a joint seminar could be organized where students from both countries present their research findings on a shared topic, fostering cross-cultural dialogue and critical thinking. (f) The study's findings may highlight the importance of promoting diversity and inclusion in educational settings. Encouraging diverse perspectives and experiences can enrich CT outcomes. At the classroom level, teachers can implement "diversity days" where students are encouraged to share their cultural backgrounds, traditions, and viewpoints, creating an inclusive environment that stimulates critical thinking through exposure to different perspectives.

7. Limitations

Despite positive results obtained in this cross-national comparison study, drawbacks can also be found. (a) Generalizability constraints. The gender imbalance between groups (e.g., higher proportion of Chinese female participants) may introduce unmeasured confounding. Besides, the Hungarian sample's homogeneity (all from University of A) may limit generalizability but ensures pedagogical consistency. In contrast, the Chinese sample's diversity (four universities) captures variations in educational approaches, which could explain the wider dispersion in critical thinking scores. (b) Item-level challenges. A notable limitation concerns Item C5, which exhibited lower discrimination in the Hungarian sample. Despite revisions to reduce translation ambiguity, residual cultural differences in interpreting may persist. This could attenuate the reliability of the subscale in Hungarian respondents, particularly for those with lower educational attainment. (c) Plausible explanations. The differences identified between Hungarian and Chinese undergraduates can be essential, but the deeper reasons behind this observed presentation cannot be unfolded in this study. We acknowledge that these explanations remain somewhat speculative and require further empirical validation. The cultural mechanisms proposed are based on theoretical frameworks and limited cross-cultural CTD studies, necessitating caution in interpreting these findings as definitive. (d) Inconsistency of analysis. Even though some research showed no MI was needed for group comparison, this conclusion was discussed hotly on the table without consistency. Accordingly, the scale used in this study cannot reach the required (partial) scalar model for the comparison of overall and latent means between two countries, especially for identifying the specific items functioning differently across cultures. If other scales can fulfill the requirement, the comparison results will be more significant, but item-level differences may exist and recommend caution when interpreting specific item scores across cultures. In addition, although MI was regarded as the prerequisite of group comparisons, *t*-test and ANOVA results might provide plausibly overall and latent mean results between countries. (e) The presence of severe DIF in several items, particularly within the habitual truth-digging dimension, indicates that cross-cultural comparisons of raw scores may be influenced by measurement bias. While we have identified and reported these DIF effects, future applications should consider item refinement or calibration procedures to enhance cross-cultural measurement equivalence.

8. Conclusion

This study compared CTD between Hungarian and Chinese undergraduates using CFA, item response theory, MI, descriptive statistics, and DIF. Results showed Hungarian students exhibited significant correlations between self-efficacy and both instant judgment and habitual truth-digging, whereas Chinese students demonstrated no correlation between self-efficacy and habitual truth-digging. Although the invariance model confirmed comparable factor structures, latent mean comparisons remain provisional until scalar invariance is established, as cultural response biases may distort item functioning—particularly for instant judgment and habitual truth-digging. Both groups showed similar openness; however, these cross-cultural differences require validation through culturally adapted measures to distinguish true score variations from measurement artifacts. Notably, weak self-efficacy in both samples highlights the need for cross-culturally validated CTD instruments.

9. Future directions

Based on the limitations identified, some potential future directions need to be stated. (a) Methodological refinements for enhanced generalizability. To mitigate sampling biases observed in the current study, future research may adopt stratified multi-site sampling across both countries. Additionally, longitudinal designs tracking students across multiple academic years or even throughout their

entire undergraduate programs would clarify whether score dispersion in the Chinese sample reflects developmental trajectories or institutional variability. (b) Item-level optimization through cultural adaptation. The cultural misalignment of Item C5 underscores the need for cognitive interviewing protocols to identify culturally loaded phrases. Furthermore, applying alignment optimization techniques may enable partial scalar invariance analysis despite non-invariant items, preserving statistical power for cross-group comparisons. (c) Theoretical expansion of cross-cultural mechanisms. The observed differences in CTD between Hungarian and Chinese undergraduates warrant deeper theoretical exploration. Future studies should employ mediation analysis to examine whether cultural dimensions such as Hofstede's power distance or uncertainty avoidance mediate CTD scores. Moreover, extending the research to other cultural contexts beyond Hungary and China, such as including samples from East Asian, Western European, or North American countries, would allow for a more comprehensive understanding of how cultural factors universally or differentially impact CTD. This cross-cultural expansion would help identify both universal and culture-specific mechanisms underlying critical thinking development. (d) Advanced measurement invariance strategies. applying Bayesian structural equation modeling could estimate approximate invariance, providing more nuanced insights than traditional frequentist approaches. If scalar invariance remains unattainable, researchers could adopt latent mean difference testing with measurement invariance constraints to validate between-group comparisons. Transparent reporting of all invariance tests (configural, metric, scalar) would align with best practices. Future studies should employ mixed methods (e.g., cognitive interviews) to unpack why items like C10 and B1 function so differently. Longitudinal designs could examine whether DIF diminishes with intercultural education. Extension to other cultural contexts would clarify whether observed patterns reflect broader East-West differences or specific Hungarian-Chinese contrasts.

Ethical approval

Participants and the university research committee both gave their consent by ethical standards. Approval Committee: The Institutional Review Board (IRB) of the Doctoral School of Education, University of Szeged. Approval number: 11/2021.

CRediT authorship contribution statement

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

Declaration of competing interest

We have no conflicts of interest to disclose.

Acknowledgement

This is to acknowledge that this work is supported by the Research Start-up Project *The Assessment and Measurement of Undergraduates' Critical Thinking Disposition* for Introduced High-Level Talent of Chongqing Technology and Business University (Grant Number: 2555003), Hungarian National Research, Development and Innovation Fund grant (OTKAK135727), Hungarian Academy of Sciences Research Programme for Public Education Development grant (KOZOKT2021–16), Humanities and Social Sciences Cluster of the Centre of Excellence for Interdisciplinary Research, Development and Innovation of the University of Szeged. GM and AP are members of the Digital Learning Technologies Incubation Research Group. AP was additionally supported by the János Bolyai Research Scholarship of the Hungarian Academy of Sciences (BO/00093/23/2).

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at [doi:10.1016/j.tsc.2025.102121](https://doi.org/10.1016/j.tsc.2025.102121).

Data availability

The data that support the findings of this study are available from the corresponding author upon reasonable request.

References

- Bandura, A., & Jourden, F. J. (1991). Self-regulatory mechanisms governing the impact of social comparison on complex decision making. *Journal of personality and social psychology*, 60(6), 941. <https://doi.org/10.1037/0022-3514.60.6.941>
- Bárdossy, I., Dudás, M., Pethőné, N. C., & Priskinné, R. E. (2002). *A kritikai gondolkodás fejlesztése. AZ interaktív és reflektív tanulás lehetőségei [Developing critical thinking. facilitating interactive and reflective learning]*. Hungary: University of Pécs.

- 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. (1995). *EQS structural equations program manual*, 6. Encino, CA: Multivariate software.
- Bixler, G. M., Brown, A., Way, D., Ledford, C., & Mahan, J. D. (2015). Collaborative concept mapping and critical thinking in fourth-year medical students. *Clinical Pediatrics*, 54(9), 833–839. <https://doi.org/10.1177/000922815590223>
- 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>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research*. Guilford publications. <http://www.kharazmi-statistics.ir/Uploads/Public/book/Methodology%20in%20the%20Social%20Sciences.pdf>
- Chen, F. (2008). What happens if we compare chopsticks with forks? The impact of making inappropriate comparisons in cross-cultural research. *Journal of personality and social psychology*, 95, 1005–1018. <https://doi.org/10.1037/a0013193>
- Chen, F. F. (2007). Sensitivity of goodness of fit indexes to lack of measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal*, 14(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Dennett, S. K. (2014). *A study to compare the critical thinking dispositions between Chinese and American college students, United States – Florida*. Ph.D., Florida Atlantic University (Publication Number 3584910) <https://www.proquest.com/dissertations-theses/study-compare-critical-thinking-dispositions/docview/1547946177/se-2?accountid=17096>
- Dennett, S. K., & DeDonno, M. A. (2021). A comparison between Chinese and American male and female college students' critical thinking dispositions. *International Journal of Chinese Education*, 10(3). <https://doi.org/10.1177/22125868211046966>
- Dewey, J. (1910). *How we think*. Lexington, MA: D.C. Heath and Company. <https://doi.org/10.1037/10903-000>
- Ebel, R. L. (1954). Procedures for the analysis of classroom tests. *Educational and Psychological Measurement*, 14(2), 352–364. <https://doi.org/10.1177/001316445401400215>
- Ennis, R. H. (1996). *Critical thinking*. Prentice-Hall. <https://doi.org/10.5840/inquiryctnews20112613>
- 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>
- Érsek, A. (2020). A történelmi tartalomba ágyazott kritikai gondolkodás kognitív elemeinek mérése, fejlesztési lehetőségei a 11-12. évfolyamos gimnáziumi tanulók körében elektronikus tanulási környezetben [Assessment and development possibilities of the cognitive elements of critical thinking embedded in historical teaching materials in the case of 11th and 12th graders in an e-learning environment]. Eger, Hungary: Dissertation, Eszterházy Károly Catholic University.
- Fábián, G. (2014). *Kritikai gondolkodás az osztályteremben [Critical thinking in the classroom]*. Gondolat Kiadó.
- Fábián, G. (2015). Non-critical thinking: What if not thinking? *Procedia - Social and Behavioral Sciences*, 186, 699–703. <https://doi.org/10.1016/j.sbspro.2015.04.088>
- Facione, P. (1990). Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (The Delphi Report). <https://philarchive.org/archive/FACCTA>
- Facione, P. A., & Facione, N. C. (1992). *California critical thinking disposition inventory*. Millbrae, CA: California Academic Press.
- Facione, P. A., Facione, N. C., & Giancarlo, C. A. (1997). *Setting expectations for student learning: New directions for higher education*. Millbrae: California Academic Press.
- Fan, K., & See, B. H. (2022). How do Chinese students' critical thinking compare with other students?: A structured review of the existing evidence. *Thinking Skills and Creativity*, 46, Article 101145. <https://doi.org/10.1016/j.tsc.2022.101145>
- Fischer-Dárdai, A., & Kaposi, J. (2021). Trends in changing history teaching in Hungary (1990–2020). *Hungarian Educational Research Journal*, 28. <https://doi.org/10.1556/063.2021.00058>
- 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>
- Hofstede, G., & Bond, M. H. (1984). Hofstede's culture dimensions: An independent validation using Rokeach's Value Survey. *Journal of Cross-Cultural Psychology*, 15(4), 417–433. <https://doi.org/10.1177/0022002184015004003>
- Hu, L. T., & Bentler, P. M. (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. <https://doi.org/10.1080/10705519909540118>
- Huang, L., Fan, A. P.-C., Su, N., Thai, J., Kosik, R. O., & Zhao, X. (2021). Chinese medical students' disposition for critical thinking: A mixed methods exploration. *Bmc Medical Education*, 21(1), 385. <https://doi.org/10.1186/s12909-021-02801-w>
- Joreskog, K. G., & Goldberger, A. S. (1975). Estimation of a model with multiple indicators and multiple causes of a single latent variable. *Journal of the American Statistical Association*, 70(351), 631–639. <https://doi.org/10.2307/2285946>
- Kaposi, J. (2020). Current issues concerning education for democracy in Hungary. *Hungarian Educational Research Journal*, 10(3), 280–295. <https://doi.org/10.1556/063.2020.00025>
- Kenny, D. A., Kaniskan, B., & McCoach, D. B. (2015). The performance of RMSEA in models with small degrees of freedom. *Sociological Methods & Research*, 44(3), 486–507. <https://doi.org/10.1177/0049124114543236>
- Klooster, D. J., Steele, J. L., & Bloem, P. L. (2001). *Ideas without boundaries: International education reform through reading and writing for critical thinking*. International Reading Association.
- Leslie, B. B., & Moilanen, J. (2010). Advancing self-efficacy and intuitive decision-making calls to action. In *Adult Education Research Conference*. <https://newprairiepress.org/aerc/2010/roundtables/8>
- Lewis, A., & Smith, D. (1993). Defining higher order thinking. *Theory into practice*, 32(3), 131–137. <https://www.jstor.org/stable/1476693?seq=1>
- Linacre, J. (2002). What do infit and outfit, mean-square and standardized mean? *Rasch Measurement Transactions*, 16.
- Liu, Y., & Pásztor, A. (2022a). Design and validate the employer-employee-supported Critical Thinking Disposition Inventory (2ES-CTDI) for undergraduates. *Thinking Skills and Creativity*, 46, Article 101169. <https://doi.org/10.1016/j.tsc.2022.101169>
- Liu, Y., & Pásztor, A. (2022b). 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>
- Liu, Y., & Pásztor, A. (2023). Moderated mediating effects of gender among the components of critical thinking disposition in undergraduate students. *Heliyon*, 9(4), Article e14664. <https://doi.org/10.1016/j.heliyon.2023.e14664>
- Liu, Y., Pásztor, A., & Molnár, G. (2026). Relationships among self-efficacy, instant judgment, and habitual truth-digging in undergraduates' critical thinking disposition. *European Journal of Education*. not published, in revision.
- Luo, Q., & Yang, X. (2001). Revision for CCTDI (Chinese version). *Psychological Development and Education*, 17(3), 47–51. <http://www.devpsy.com.cn/CN/Y2001/V17/I3/47#1>
- Ma, H. (2021). Critical thinking dispositions of undergraduate college students in a comprehensive university of Mainland China. *International Journal of Higher Education*, 10(6), 224–237. <https://doi.org/10.5430/ijhe.v10n6p224>
- Magis, D., Raiche, G., Béland, S., & Gerard, P. (2012). A generalized logistic regression procedure to detect differential item functioning among multiple groups. *International Journal of Testing*, 11, 365–386. <https://doi.org/10.1080/15305058.2011.602810>
- Marsh, H. W., & Hocevar, D. (1985). Application of confirmatory factor analysis to the study of self-concept: First-and higher order factor models and their invariance across groups. *Psychological Bulletin*, 97(3), 562–582. <https://doi.org/10.1037/0033-2909.97.3.562>
- Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika*, 47(2), 149–174. <https://doi.org/10.1007/BF02296272>
- McBride, R. E., Xiang, P., Wittenburg, D., & Shen, J. (2002). An analysis of preservice teachers' Dispositions toward critical thinking: A cross-cultural perspective. *Asia-Pacific Journal of Teacher Education*, 30(2), 131–140. <https://doi.org/10.1080/13598660220135649>
- Mok, M., Cheong, C., Moore, P., & Kennedy, K. (2006). The development and validation of the self-directed learning scales (SLS). *Journal of applied measurement*, 7, 418–449. <https://pubmed.ncbi.nlm.nih.gov/17068381/>
- Molnár, L. (2002). A kritikai gondolkodás [Critical thinking]. In B. Csapó (Ed.), *Az iskolai műveltség [Literacy at school]* (pp. 217–237). Osiris Press.

- Muthén, B. (1984). A general structural equation model with dichotomous, ordered categorical, and continuous latent variable indicators. *Psychometrika*, 49(1), 115–132. <https://doi.org/10.1007/BF02294210>
- O'Donald, F., & Calia, C. (2025). The process of translation and cross-cultural adaptation of functional assessment tools for dementia: A systematized review. *Health Sci Rep*, 8(1), Article e70289. <https://doi.org/10.1002/hsr2.70289>
- 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 xinxiadou ceshi yanjiu [Validity and Reliability of the Chinese Critical Thinking Disposition Inventory]. *Chinese Journal of Nursing*, 39(9), 644–647. <https://core.ac.uk/download/pdf/78762281.pdf>
- Perkins, D., Jay, E., & Tishman, S. (1993). Beyond abilities: A dispositional theory of thinking. *Merrill-Palmer Quarterly*, 39, 1–21. https://www.researchgate.net/publication/232462299_Beyond_Abilities_A_Dispositional_Theory_of_Thinking
- Petrini, M. A., & Kawashima, A. (2003). Comparison of critical thinking skills of nurses in Japan, China and Samoa. *Bulletin of the Graduate Schools Yamaguchi Prefectural University*, 4, A11–A28. A29–A32 <https://cir.nii.ac.jp/crid/1571417126782761984>
- Proitsi, P., Hamilton, G., Tsolaki, M., Lupton, M., Daniilidou, M., Hollingworth, P., Archer, N., Foy, C., Stylios, F., McGuinness, B., Todd, S., Lawlor, B., Gill, M., Brayne, C., Rubinsztein, D. C., Owen, M., Williams, J., Craig, D., Passmore, P., ... Powell, J. F. (2011). A multiple indicators multiple causes (MIMIC) model of behavioural and psychological symptoms in dementia (BPSD). *Neurobiol Aging*, 32(3), 434–442. <https://doi.org/10.1016/j.neurobiolaging.2009.03.005>
- 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>
- Robitzsch, A., & Lüdtke, O. (2023). Why full, partial, or approximate measurement invariance are not a prerequisite for meaningful and valid group comparisons. *Structural Equation Modeling: A Multidisciplinary Journal*, 1–12. <https://doi.org/10.1080/10705511.2023.2191292>
- Sörbom, D. (1974). A general method for studying differences in factor means and factor structure between groups. *British Journal of Mathematical and Statistical Psychology*, 27(2), 229–239. <https://doi.org/10.1111/j.2044-8317.1974.tb00543.x>
- Sorrel, M. A., García, L. F., Aluja, A., Rolland, J. P., Rossier, J., Roskam, I., & Abad, F. J. (2021). Cross-cultural measurement invariance in the personality Inventory for DSM-5. *Psychiatry Research*, 304, Article 114134. <https://doi.org/10.1016/j.psychres.2021.114134>
- 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
- Tiwari, A., Avery, A., & Lai, P. (2003). Critical thinking disposition of Hong Kong Chinese and Australian nursing students. *Journal of Advanced Nursing*, 44(3), 298–307. <https://doi.org/10.1046/j.1365-2648.2003.02805.x>
- Tóth, L. (2007). *Kritikai olvasás – kritikai gondolkodás [Critical reading – critical thinking]*. Pedellus.
- Tucker, L. R., & Lewis, C. (1973). A reliability coefficient for maximum likelihood factor analysis. *Psychometrika*, 38(1), 1–10.
- 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>
- Vincent-Lancrin, S., González-Sancho, C., Bouckaert, M., Luca, F. d., Fernández-Barrera, M., Jacotin, G., Urgel, J., & Vidal, Q. (2019). *Fostering students' creativity and critical thinking*. <https://doi.org/10.1787/62212c37-en>
- Wolfe, E. W., & Chiu, C. W. (1999). Measuring change across multiple occasions using the Rasch Rating Scale Model. *Journal of outcome measurement*, 3(4), 360–381. <https://pubmed.ncbi.nlm.nih.gov/10572387/>
- Wu, H. (2019). *Identifying the influencing factors of problem solving: A cross-national study*. Szeged, Hungary: University of Szeged [Doctoral Dissertation <https://doktori.bibl.u-szeged.hu/id/eprint/10333>]
- Wu, H., & Molnár, G. (2021). Logfile analyses of successful and unsuccessful strategy use in complex problem-solving: A cross-national comparison study. *European Journal of Psychology of Education*, 36(4), 1009–1032. <https://doi.org/10.1007/s10212-020-00516-y>
- Xie, Y., Smith, J., & Davies, M. (2025). The evolution of critical thinking in Chinese education context: Policy and curriculum perspectives. *International Studies in Sociology of Education*, 1–24. <https://doi.org/10.1080/09620214.2025.2462967>
- Xu, Y. (2018). Assessing measurement invariance and latent mean differences with bifactor multidimensional data in structural equation modeling. ARIZONA STATE UNIVERSITY. <https://core.ac.uk/download/pdf/161995233.pdf>
- Yeh, M. L., & Chen, H. H. (2003). Comparison affective dispositions toward critical thinking across chinese and american baccalaureate nursing students [Article]. *Journal of Nursing Research*, 11(1), 39–46. <https://doi.org/10.1097/01.JNR.0000347617.29413.96>
- 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: The Catholic University of Education.
- Zhang, Y., Li, L., Wu, P., & Chen, Y. (2009). Investigation and analysis of critical thinking ability in medical students. *Journal of Shanghai Jiaotong University (Medical Science)*, 869–872.
- Zhu, W., & Cole, E. L. (1996). Many-faceted rasch calibration of a gross motor instrument. *Research Quarterly for Exercise and Sport*, 67(1), 24–34. <https://doi.org/10.1080/02701367.1996.10607922>