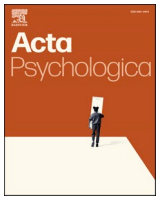
Acta Psychologica 262 (2026) 106055



Contents lists available at ScienceDirect

Acta Psychologica

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





The development and validation of a critical thinking disposition scale for high school students

Wencheng Liu [a,1], Huangdong Ma [b,1], Shiwen Ma [c], Ying Chen [d], Gaofeng Li [e,*]

[a] Faculty of Education, Shaanxi Normal University, Xi'an, China
[b] School of Teacher Development, Shaanxi Normal University, Xi'an, China
[c] College of Physical Science and Technology, Yili Normal University, Yining, China
[d] College of Mechanical Engineering, Chongqing University of Technology, China
[e] College of Life Sciences, Shaanxi Normal University, No. 620, West Chang'an Avenue, Xi'an, 710119, Chang'an District, China

ARTICLE INFO

Keywords:
Critical thinking disposition
Critical thinking
High school students
Scale development
Measurement invariance

ABSTRACT

This study develops and validates a Critical Thinking Disposition (CTD) scale specifically designed for Chinese high school students, addressing the need for assessment tools tailored to this particular demographic. Anchored in Facione's seven-dimensional framework and previous conceptual research, the study employed a rigorous three-phase process—item development, pilot testing, and final validation—using two convenience samples to examine the scale's psychometric properties. Exploratory and confirmatory factor analyses refined the instrument into a five-dimensional structure: Systematic-analyticity, Open-mindedness, Confidence in Reasoning, Reflective Skepticism, and Truth-seeking. The final model demonstrated optimal fit indices (χ^2/df = 2.173, CFI = 0.964, RMSEA = 0.043), confirming its structural stability. In addition, convergent and discriminant validity were supported by Average Variance Extracted (AVE), Composite Reliability (CR), and the Fornell–Larcker criterion. At the same time, internal consistency was evidenced by robust Cronbach's α (0.881) and McDonald's ω (0.908) values for the overall scale. Multi-group CFA further established measurement invariance across gender (ΔCFI <0.01, ΔRMSEA <0.015). Overall, the scale provides a reliable and theoretically grounded instrument for assessing CTD in secondary education contexts, enabling educators to evaluate the CTD of high school students comprehensively.

1. Introduction

Critical thinking (CT) is widely acknowledged as a fundamental skill in modern education, essential for academic success and preparing students for the challenges of the 21st century (Dwyer et al., 2014). By enhancing CT, students not only improve their academic performance but also develop other important skills, such as concentration, in-depth analysis, and critical reasoning (Bielik & Krüger, 2024; Hu & Bi, 2025). Numerous studies, when defining CT, frequently associate it with core skills like reasoning, analysis, and evaluation, highlighting its status as a key competency alongside communication and collaboration in educational contexts (Altun & Yildirim, 2023; Halpern & Dunn, 2021; Hwang et al., 2023; Santos-Meneses et al., 2023). Beyond academic achievement, CT is also integral to effective decision-making in personal and professional contexts (Yin et al., 2023). In today's information-rich environment, individuals must be active, critical thinkers to navigate the complexities of the digital age (Walter, 2024). As a result, fostering CT has become one of the central objectives in education systems worldwide.

Despite its recognized importance, one of the key challenges in promoting CT lies in the lack of a universally accepted definition. To address this, the American Philosophical Association (APA) assembled a Delphi panel of experts who defined CT as a cognitive process grounded in the rational application of skills (Facione, 1990). According to the panel, CT comprises both skills—such as the ability to analyze and evaluate—and dispositions, which refer to the attitudes and tendencies that influence whether and how individuals engage in CT.

Ever since the concept of CT was established with consensus, research has tended to focus more on skills than dispositions, often emphasizing reasoning and evaluation within these skills (Fabio et al.,

* Corresponding author.
E-mail addresses: wenchengsir@snnu.edu.cn (W. Liu), mahuangdong@snnu.edu.cn (H. Ma), ligaofeng@snnu.edu.cn (G. Li).
[1] Wencheng Liu and Huangdong Ma share the first authorship.

https://doi.org/10.1016/j.actpsy.2025.106055
Received 19 April 2025; Received in revised form 6 October 2025; Accepted 1 December 2025
Available online 3 December 2025
0001-6918/© 2025 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/).

2024; Sosu, 2013). These sub-skills are centered on an individual's ability to understand problems, analyze arguments, claims, or evidence, and propose reasonable solutions for the identified issues. This focus has also led to the development of numerous instruments designed to measure individual skill levels (Ku, 2009). Most studies measure skills including analysis, evaluation, inference, recognition of assumptions, and deduction (Li & Liu, 2021).

However, dispositions are equally crucial alongside CT skills, as they determine how individuals approach problems, seek information, and evaluate evidence (Manassero-Mas et al., 2022; Quinn et al., 2020). For instance, an individual may possess the necessary skills but fail to apply them due to a lack of motivation or other personal factors, resulting in ineffective or incomplete CT (Altun & Yildirim, 2023).

As CT education expands from higher education to other educational levels, adolescents have become the focal point for both researchers and educators (Chen et al., 2024; Hu & Bi, 2025; Manassero-Mas et al., 2022; Sarigoz, 2012). In China, the latest high school curriculum reform, implemented in 2017, incorporated CT as part of the new curriculum standards. However, no guidelines were provided regarding assessment tools, leading to a gap in the measurement of critical thinking dispositions (CTD) among Chinese high school students (Gao et al., 2025). While international tools, such as the California Critical Thinking Disposition Inventory (CCTDI) and the UF/EMI Critical Thinking Disposition Scale, exist, they have limitations when applied to Chinese high school students. The CCTDI, derived from Facione's seven-dimensional framework, was developed for adults and includes items with limited developmental relevance for adolescents. By contrast, the UF/EMI, though designed for secondary students (Ricketts & Rudd, 2005), compresses the Delphi-derived dispositions into three factors and can underrepresent core facets such as truth-seeking and systematicity.

Although specific items in the CCTDI may not be contextually suitable for Chinese high school students, its seven theoretical dimensions remain widely recognized and empirically supported in CTD research. Therefore, this study adopts Facione's seven-dimensional framework while also taking into account the broader conceptual insights of the Delphi consensus. Building on this foundation, we reconstruct item content to create a tool tailored to the Chinese educational context—linguistically, cognitively, and culturally appropriate—so as to ensure both psychometric robustness and developmental relevance.

2. Literature review

2.1. Critical thinking disposition

CT is defined by APA Delphi consensus as a purposeful, self-regulatory judgment that includes the cognitive processes of interpretation, analysis, evaluation, inference, and explanation, as well as a reflective consideration of evidence, concepts, methods, and contextual factors to reach reasoned decisions (Facione, 1990; Yanchar et al., 2008). The importance of CT lies in its essential role in fostering informed decision-making, problem-solving, and independent thought, particularly within increasingly complex and information-rich environments. According to the Delphi report, CT is comprised of both skills and dispositions, and both are indispensable for the full realization of critical thought (Altun & Yildirim, 2023; Facione, 1990).

It is essential to distinguish CT skills—the actual ability to analyze and evaluate—from CTD—the habitual inclinations or attitudes that motivate an individual to think critically (Altun & Yildirim, 2023). While the former reflects what a person can do, the latter reflects what a person is willing to do when faced with problems or dilemmas (Ennis, 2015; Facione et al., 2000). In this regard, CTD is not merely supplementary but constitutes the internal readiness and reflective mindset necessary for the practical application of CT.

The APA consensus identifies key dispositions of CT, including inquisitiveness, being well-informed, self-confidence in one's reasoning abilities, open-mindedness, willingness to reconsider and revise views when warranted, orderliness in dealing with complexity, and precision as permitted by context and subject matter (Facione, 1990). These affective tendencies underscore the multidimensional nature of CTD, highlighting the mental attitudes that enable complex judgment and sustained reasoning. Consequently, the development and assessment of CTD have become crucial to the broader goal of cultivating critical thinkers, particularly in educational and professional settings where reasoned judgment is essential.

Subsequently, Facione delineated seven distinct dimensions of CTD, which include Inquisitiveness, Open-mindedness, Systematicity, Analyticity, Truth-seeking, CT Self-confidence, and Maturity. Compared to the APA consensus, this list is more clearly defined, providing measurable traits that can be more easily assessed (Facione et al., 1994).

Halpern (1998) extended the conceptualization of CTD by identifying specific dispositions of a critical thinker, which include (a) willingness to engage in and persist at a complex task, (b) habitual use of plans and the suppression of impulsive activity, (c) flexibility or open-mindedness, (d) willingness to abandon nonproductive strategies in an attempt to self-correct, and (e) an awareness of the social realities that need to be overcome so that thoughts can translate into actions. This conceptualization highlights not only the cognitive and emotional elements of CT but also the contextual factors influencing effective decision-making (Halpern & Dunn, 2021).

Ennis (2015) describes the concepts of "seeking alternatives and being open to them" and "seriously considering other points of view," which encapsulate the trait of open-mindedness and emphasize remaining unbiased and receptive to different perspectives. The traits of "take into account the total situation" and "reflectively aware of own beliefs" reflect the dimensions of systematicity and reflectivity. Furthermore, Ennis (1996) critiqued Facione's list, particularly the terms "maturity" and "CT self-confidence", arguing that they are too vague and lack the necessary clarity for precise conceptualization.

The concept of Maturity by Facione presents the CT-mature person as one who approaches problems, inquiry, and decision-making with a sense that some problems are necessarily ill-structured, and some situations admit of more than one plausible option. The idea that "some situations admit of more than one plausible option" suggests an analytical tendency, while recognizing that "some problems are necessarily ill-structured" may require skepticism awareness (Chen et al., 2024). This interpretation suggests that maturity in CTD may not be fully defined, and that certain aspects of this disposition—particularly the tendency to question assumptions and delay closure in the face of ambiguity—may align more closely with a distinct dimension of Reflective Skepticism. The Delphi expert consensus also includes "willingness to reconsider and revise views where honest reflection suggests that change is warranted," which reflects a clearly defined reflective component of CTD.

Table 1 summarizes the key points of the research on CTD. As shown in Table 1, while different theorists use varying terminology to conceptualize CTD, there is significant overlap in the traits identified across frameworks, such as Open-mindedness, Reflection, and Truth-seeking (Manassero-Mas et al., 2022). For example, the concepts of "to seek and evaluate reasons", "seek as much precision as the situation requires" (Ennis, 2015), and "precision to the degree permitted by the subject and the circumstance" all reflect the disposition of Truth-seeking. Similarly, "attempt to self-correct" (Halpern, 1998), "use credible sources and observations" (Ennis, 2015), and "to be metacognitive" represent a shared notion of Reflective Skepticism. Additionally, these theoretical frameworks also embody characteristics of Systematicity (e.g., "habitual use of plans," "take into account the total situation," "orderliness in working with complexity," "to be planful and strategic") and Analyticity (e.g., "try to be well informed").

In conclusion, CT is a multifaceted construct underpinned by dispositions such as open-mindedness, reflection, and truth-seeking. These accounts provide a coherent basis for defining CTD and for designing assessments. Accordingly, we ground the present study in Facione's

Table 1
Summary of key points on CTD.

Authors	Conceptualization
APA consensus (Delphi) Facione (1990)	1. Inquisitiveness, 2. Well-informed, 3. Open-mindedness, 4. Self-confidence in one's own ability to reason, 5. Willingness to reconsider and revise views where honest reflection suggests that change is warranted, 6. Orderliness in working with complexity, precision to the degree permitted by the subject and the circumstance.
Facione et al. (1994)	1. Inquisitiveness, 2. Open-mindedness, 3. Systematicity, 4. Analyticity, 5. Truth-seeking, 6. CT Self-confidence, 7. Cognitive Maturity.
Perkins et al. (1993)	1. To be broad and adventurous, 2. Toward sustained intellectual activity, 3. To clarify and seek understanding, 4. To be planful and strategic, 5. To be intellectually careful, 6. To seek and evaluate reasons, 7. To be metacognitive.
Ennis (2015)	1. Try to be well informed, 2. Use credible sources and observations, 3. Take into account the total situation, 4. Be alert for alternatives, 5. Be open-minded, 6. Seek as much precision as the situation requires.
Halpern (1998)	1. willingness to engage in and persist at a complex task, 2. habitual use of plans and the suppression of impulsive activity, 3. flexibility or open-mindedness, 4. willingness to abandon nonproductive strategies in an attempt to self-correct, 5. an awareness of the social realities that need to be overcome so that thoughts can become actions.

Note. Due to space constraints, overlapping content is not presented in the table.

seven-dimension framework and integrate insights from the Delphi consensus to ensure comprehensive construct coverage. This combined foundation supports an instrument that is conceptually rigorous yet practical for use with high school students.

2.2. The assessment of critical thinking dispositions

The assessment of CTD is grounded in theoretical frameworks developed by the consensus of Delphi experts, Facione, Ennis, and Halpern. As detailed in section 2.1, these frameworks define CTD as a multifaceted construct, which is pivotal for developing and validating assessment tools.

Currently, Self-report scales are the most widely adopted method for assessing CTD. CTD—such as open-mindedness and truth-seeking—are internal psychological traits that are inherently introspective and implicit, making them difficult to measure through direct observation. Self-report tools enable respondents to reflect on and express their attitudes and inclinations, thus offering a practical approach to capturing these internal experiences (Kam, 2020). This method is also standard in other domains of psychological assessment. However, self-report instruments are susceptible to social desirability bias, wherein respondents may respond in a way they perceive as socially acceptable rather than truthful (Podsakoff et al., 2003). Nevertheless, techniques such as anonymous administration and reverse-coded items can effectively mitigate such biases and enhance the validity of the collected data.

Among CTD measurement tools, the CCTDI has been widely used, especially in higher education settings. Developed by Facione based on the Delphi expert consensus, the CCTDI conceptualizes CTD into seven dimensions: Inquisitiveness, Open-mindedness, Truth-seeking, Analyticity, Systematicity, Confidence in Reasoning, and Cognitive Maturity.

The CCTDI comprises 75 items measured on a six-point Likert scale. Its multidimensional structure closely aligns with Facione's theoretical framework, thereby supporting its construct validity. It has provided valuable insights into assessing individual CTD and facilitating cross-cultural comparisons (Bravo et al., 2020; Ku & Ho, 2010; Orhan, 2022; Yeh, 2002; Zoller et al., 2000). For instance, Boso et al. (2021) assessed undergraduate nursing students in Ghana and found generally positive CTD, yet significant weakness in Truth-seeking.

Despite its contributions, existing literature highlights several psychometric concerns with the CCTDI, including overlapping item structures, excessive cross-loading, and unstable factor structures (Manassero-Mas et al., 2022; Sosu, 2013). Most notably, Orhan (2022) conducted a meta-analysis that highlighted inconsistencies in the factor structure of the CCTDI across multiple studies, with the number and structure of factors being unstable, which challenges the instrument's dimensional clarity. These issues raise questions about the validity and reliability of the CCTDI, particularly its ability to distinguish accurately between the different dimensions of CTD. For example, overlapping items and cross-loading can blur the distinction between dimensions, leading to measurement ambiguity and reduced clarity in factor interpretation. For example, İskifoğlu and Ağazade (2013) conducted confirmatory factor analyses (CFA) on a Turkish version and found poor model fit. They proposed a four-factor alternative better suited for cross-cultural applications. These findings underscore the need to reconsider the CCTDI's applicability and its structural soundness.

In response to these limitations, researchers have developed alternative CTD scales for specific populations (Liu & Pásztor, 2022; Quinn et al., 2020; Wang et al., 2019). Yuan et al. (2014) designed a CTD instrument tailored to healthcare professionals, extracting three factors—Systematicity and Analyticity, Inquisitiveness and Conversance, and Maturity and Skepticism—with high internal consistency. Similarly, Sosu (2013) created an 11-item CTD scale featuring two dimensions—Critical Openness and Reflective Skepticism—which demonstrated good convergent validity.

In parallel, researchers have been made to expand CTD assessment to adolescent populations (Giancarlo et al., 2004; Hu & Bi, 2025; Kawashima & Shiomi, 2007). Manassero-Mas et al. (2022) adapted the CCTDI for students aged 11–15. They developed a six-factor version, merging Maturity and Analyticity into a single factor labeled Prudence-Analysis, which showed satisfactory reliability ranging from 0.713 to 0.87. Likewise, Chen et al. (2024) proposed a seven-dimension scale to assess CTD in Chinese students from Grades 4 to 12. While both studies extended the age range for CTD measurement, neither developed a tool specifically for high school students, thus limiting their practical utility for secondary education contexts (Çelik İskifoğlu et al., 2022; Liu & Pásztor, 2023; Orhan, 2022).

Furthermore, the UF/EMI Critical Thinking Disposition Instrument was specifically designed for high school and youth leadership contexts (Ricketts & Rudd, 2005). The UF/EMI condenses the theoretical foundation of the Delphi report into three core dimensions: Engagement (active willingness to apply CT), Cognitive Maturity (awareness of complexity and openness to new ideas), and Innovativeness (intellectual curiosity). Adaptation studies in Turkey have confirmed the cross-cultural reliability and factorial validity of the scale, reporting Cronbach's alpha values of 0.88 (Engagement), 0.70 (Cognitive Maturity), and 0.73 (Innovativeness) (Demircioğlu & Kilmen, 2014; Ertaş Kılıç & Şen, 2014).

Table 2 summarizes the key tools for measuring CTD. Most instruments were developed for college or adult populations and converge on core dimensions such as Open-mindedness, Truth-seeking, Analyticity, and Systematicity (Cui et al., 2021). Dimensions like Confidence in Reasoning and Reflective Skepticism have received partial empirical support. Moreover, the development of purpose-built tools for specific groups (e.g., university or medical students) has further constrained the accumulation of direct empirical evidence for application in high-school populations. Against this backdrop, the UF/EMI is developmentally appropriate for high-school students and has documented cross-cultural reliability and validity. However, its three-factor condensation does not fully represent several Delphi-identified dispositions—particularly Truth-seeking and Systematicity—potentially trading theoretical breadth for practical efficiency. When assessment goals require both adolescent relevance and broader dispositional coverage, UF/EMI's

Table 2
The taxonomies of CTD.

Author	Year	Level	Dimensions	Key characteristics
Facione	1994	college students	1. Inquisitiveness; 2. Open-mindedness; 3. Systematicity; 4. Analyticity; 5. Truth-seeking; 6. Confidence in Reasoning; 7. Cognitive Maturity	Foundational and widely used with broad dispositional coverage; however, its factor structure/ construct validity has shown inconsistencies, and several items show limited developmental relevance for high-school populations.
Ricketts & Rudd	2005	high school & youth leadership	1. Engagement; 2. Cognitive Maturity; 3. Innovativeness	Designed for high school and youth leadership with cross-cultural adaptations; however, the three-factor condensation may underrepresent Delphi-derived facets (e.g., truth-seeking, systematicity) relative to comprehensive frameworks.
Sosu	2013	college students	1. Critical openness; 2. Reflective Skepticism	Features a clear two-factor structure with supporting evidence in university samples.
Yuan et al.	2014	healthcare professionals and students	1. Systematicity and Analyticity; 2. Inquisitiveness and Conversance; 3. Maturity and Skepticism	Tailored to healthcare professionals with domain-relevant content; generalization to broad high-school populations is limited by professional sampling.
Wang et al.	2019	medical college students	1. Open-mindedness; 2. Systematicity/ Analyticity; 3. Truth-seeking	Provides updated evidence in Chinese medical education; however, domain-specific content and sampling constrain use beyond medical students.
Manassero-Mas et al	2022	primary and secondary students (Age 11–15)	1. Self-Confidence; 2. Truth-seeking; 3. Open-mindedness; 4. Curiosity; 5. System; 6. Prudence-Analysis	CCTDI-based adaptation for students aged 11–15 with a six-factor solution; however, it has not been specifically validated for high-school settings.
Chen et al	2024	primary and secondary students (Grades 4–12)	1. questioning spirit; 2. independent thinking; 3. open-mindedness; 4. curiosity; 5. truth-seeking spirit; 6. evidence-based reasoning; 7. self-reflection	Covers Grades 4–12 in China with a dimension set differing from the CCTDI; however, high-school-specific validation has not yet been demonstrated.

Note. Statements summarize each instrument's scope and literature-noted issues to facilitate comparison with the present study; they are not exhaustive evaluations.

strengths in brevity and fit must be weighed against the underrepresentation of these core facets. In contrast, the present instrument seeks to balance adolescent relevance with broader dispositional coverage, addressing this gap.

Therefore, this study adopts Facione's seven-dimensional framework as the theoretical foundation, incorporating insights from the Delphi consensus and empirical validations to ensure comprehensive structural coverage. To enhance cultural relevance, Inquisitiveness is integrated into Truth-seeking in the Chinese context, where the pursuit of truth is inherently driven by curiosity. On this basis, six dimensions are hypothesized for the new CTD scale—Open-mindedness, Truth-seeking, Analyticity, Systematicity, Confidence in Reasoning, and Reflective Skepticism—and the instrument's reliability and validity are evaluated to provide robust psychometric evidence for its use.

3. Methods

3.1. Research design

This study employed a three-step approach—item construction, pilot testing, and final testing—to develop and validate the CTD scale. During the item construction, four specific stages were followed: item generation, item review, item revision, and item validation. A total of 30 initial items were generated, drawing upon both existing scales identified through literature review and newly developed items designed specifically for the research context. Following the construction of the item, a pilot test was conducted with a sample of 215 Chinese high school students to examine the preliminary structure. Based on the results of exploratory factor analyses (EFA), the scale was revised. Specifically, nine items were removed due to low factor loadings and because some factors contained fewer than three items. The revised scale was then administered to a new sample of 684 students to evaluate its psychometric properties, including validity and reliability. Data from both samples were screened to exclude cases with inattentive or careless responses.

Both the pilot and final versions of the questionnaire adopted a 5-point Likert scale (“1 = not at all like me”, “5 = very much like me”). This scale format was selected to minimize respondents' cognitive load and to provide a neutral midpoint, enabling participants to express ambivalence without being forced into a directional choice. To control for acquiescence bias, reverse-coded items were evenly distributed throughout the questionnaire (see Appendix A). Items were presented in a fixed order in the paper-based format.

The study received ethical approval from the Research Management Department of a university in Shaanxi Province, China (Approval No. ARB 2024-05-01). Both the pilot and final tests were conducted in classroom settings under the supervision of teachers. Prior to administration, teachers explained the purpose of the study and assured participants of the confidentiality and anonymity of their responses. All participants provided informed consent for inclusion before participating in the study. Completion of the questionnaire took approximately 15 to 20 min. All responses were collected and stored in Microsoft Excel, with data access restricted to members of the research team to ensure data security and confidentiality.

3.2. Item development

This study developed the CTD scale based on a conceptual structure of CTD identified in previous literature (Manassero-Mas et al., 2022; Sosu, 2013). After defining the relevant dimensions from theoretical analysis and a comprehensive literature review, we analyzed the Chinese version of the CCTDI, the scale developed by Sosu (2013), and

items from Manassero-Mas et al. (2022) to construct the initial item pool.

The development process consisted of four stages. During the item generation stage, two members of the research team—both doctoral candidates in education with over ten years of teaching experience in high schools (the first and second authors)—developed an initial pool of 30 items across six dimensions. Of these, 14 items were independently developed. At the same time, the remaining 16 were adapted or drawn from established instruments, including the Chinese version of the CCTDI, Sosu's (2013) scale, and Manassero-Mas et al. (2022) (see Appendix A). For example, the item "I am able to understand things by myself" under the Self-confidence dimension was considered lacking in context and was modified to "When solving problems, I enjoy trying to understand things by myself." In the Self-confidence dimension, the item "I notice that I reason better and better" was viewed as capturing a general perception of improvement rather than confidence in reasoning during problem-solving, and was therefore revised to "I feel confident when I reason through problems." Other items were self-developed, guided by the theoretical conceptualization of each dimension. During item construction, care was taken to ensure that each sentence contained a distinct keyword and did not conflict semantically with the background or terminology of other items, thereby minimizing confusion for respondents. For instance, "Before making a decision, I tend to review my decision-making process multiple times" reflects self-doubt and reflective thinking and was designed to assess Reflective Skepticism. All item sources are documented in Appendix A. Nine items were negatively worded (e.g., "In class discussions, I am often unwilling to consider the views proposed by others" under Open-mindedness) and were reverse-coded so that higher scores reflected lower dispositions.

To ensure a clear distinction between CTD and cognitive skills during item construction, we followed the definitions outlined in the literature. CTD refers to the habitual inclinations or attitudes that motivate individuals to think critically, while cognitive skills focus on the ability to analyze and evaluate information. Based on this distinction, we designed items to reflect dispositions—the tendencies and attitudes that guide how individuals approach CT tasks—using language that emphasizes habitual behavior or motivational intent (e.g., "enjoy," "prefer," "tend to"), rather than language associated with cognitive abilities (e.g., "can," "able to").

Second, in the item review stage, two additional researchers—the third author, a doctoral candidate in education with over ten years of university-level teaching experience, and the corresponding author, a professor in education—reviewed each item independently. Items were retained only if both reviewers approved them; in cases of disagreements, all four researchers collaboratively discussed and modified the items. Additionally, a Chinese language teacher with 13 years of teaching experience at the high school level reviewed the wording of all items and offered suggestions for improved clarity and readability.

Third, in the item revision stage, five items were revised based on the review process. For example, in the Truth-seeking dimension, the item "Even if there is evidence that contradicts my ideas, I still insist on my ideas" was seen as reflecting closed-mindedness rather than a genuine truth-seeking disposition. It was revised to "I still persist in finding other evidence to verify my ideas," which reflects a more open and inquisitive attitude. In Reflective Skepticism, the phrase "I often think about my actions to see whether I could improve them" was considered vague; the revised version, "After making a decision, I usually want to check to see if the entire process can be improved," more clearly focuses on post-decision reflection dispositions. Another item in the same dimension, "I often re-evaluate my experiences so that I can learn from them," was revised to "When making decisions, I often re-evaluate my experiences because I want to learn from them," to emphasize decision-related reflection dispositions. In the Systematicity dimension, the original item, "I typically organize disparate pieces of evidence and opinion to form a new opinion," included two focal terms— "evidence" and "opinion"—which may distract students. It was revised to "I like to organize evidence before forming an opinion," to maintain conceptual clarity. Additionally, "I like to learn how everything works" was revised to "When I discover something new, I like to learn how it works" to enhance contextual relevance in the Truth-seeking dimension.

Fourth, in the item validation stage, six 11th-grade students and six graduate students majoring in education were invited to complete the draft questionnaire. They were asked to evaluate whether the items were relevant to CTD and whether they were easy to understand and respond to. Participants were encouraged to identify items that were potentially vague, ambiguous, or complex in nature. Feedback indicated that all items were transparent and interpretable. As a result, 30 finalized items were retained for empirical testing.

3.3. Instruments

The development of the scale resulted in a measurement tool for pilot testing, comprising six dimensions with five items each, totaling 30 items. The dimensions identified were Open-mindedness, Confidence in Reasoning, Truth-seeking, Reflective Skepticism, Analyticity, and Systematicity. An EFA was conducted on the pilot test data; the Analyticity dimension exhibited strong loadings on the Systematicity dimension, prompting their integration into a single dimension.

In the final testing, the revised scale was applied to a new sample of students to evaluate its psychometric properties further. The final scale, therefore, consisted of 21 items across five dimensions: Systematic-analyticity, Open-mindedness, Confidence in Reasoning, Reflective Skepticism, and Truth-seeking. The complete questionnaire is available in Appendix A.

3.4. Participants

This study involved two distinct convenience samples for the pilot test and the final test. To ensure the practicality of data collection and to capture the most representative segment of the student population, we adopted a convenience sampling method by selecting students from middle-level classes in 10th and 11th grades. This decision was based on the consideration that most high school students fall into the mid-performance range, making them more suitable for exploring general patterns of motivation in using CT skills.

In the pilot test, 215 high school students from 10th and 11th grades, aged 15 to 18, were initially recruited from three different schools. After excluding 32 responses due to inattentive answering patterns—such as identical responses across all items or repetitive ascending/descending patterns (e.g., 1–2–3-4-5 or 5–4–3-2-1)—a final sample of 183 valid cases was retained.

For the final test, 684 high school students aged 15 to 18 were recruited from nine schools located in northern, central-western, and southeastern coastal regions of China. These participants were recruited through the researchers' acquaintances—high school teachers based in various regions across mainland China—to enhance the geographical diversity of the sample. Similar to the pilot test, 59 responses were excluded due to inattentive response behaviors, resulting in a final sample of 625 valid participants. To encourage participation, each student received a bookmark as a token of appreciation.

The sample size was deemed sufficient for conducting both EFA and CFA, which ensured the reliability and validity of the instrument across diverse geographical contexts (MacCallum et al., 1999; Memon et al., 2020). Detailed demographic information about the sample is provided in Table 3.

3.5. Data analysis

In this study, the pilot test instrument was evaluated using EFA to identify potential variables related to CTD. The suitability of the data was assessed with the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's test. A KMO value greater than 0.6, combined with a significant

Table 3
The sample's information for the questionnaire.

Variables	category	Pilot test number	Percentage	Final test number	Percentage
Gender	Girls	99	54	333	53.3
	Boys	84	46	292	46.7
Age	15	4	2.2	14	2.2
	16	59	32.2	215	34.4
	17	78	42.6	303	48.5
	18	42	23	93	14.9
Level	Grades 10	89	48.6	343	54.9
	Grades 11	94	51.4	282	45.1

Bartlett's test ($p < 0.05$), indicated that the data were appropriate for factor analysis.

Factor extraction was conducted using the maximum likelihood method, retaining factors with eigenvalues greater than 1, followed by Promax oblique rotation for improved interpretability. Items with loadings below 0.32 or cross-loadings where the difference between the highest and second-highest loadings was less than 0.1 were removed (Worthington & Whittaker, 2006). Factors with fewer than three items were discarded. This iterative refinement helped produce a factor structure that was both stable and conceptually coherent, thereby supporting the instrument's construct validity.

The final instrument's reliability and validity were evaluated through various measures. Content validity was supported by references to the Chinese CCTDI scale, Sosu (2013), and Manassero-Mas et al. (2022). Convergent validity was assessed using the Average Variance Extracted (AVE) and Composite Reliability (CR). An AVE greater than 0.50 and a CR exceeding 0.70 indicate adequate convergent validity. Discriminant validity was evaluated using two methods: cross-loading analysis—where an item with a loading difference of at least 0.10 on its designated and non-designated factors signified discriminant validity—and the Fornell-Larcker criterion, which requires the square root of each construct's AVE to exceed its correlations with other constructs (Fornell & Larcker, 1981; Lin et al., 2023). Internal consistency was measured using Cronbach's alpha and McDonald's omega coefficients.

Furthermore, CFA was conducted using AMOS 26.0 to evaluate the construct validity of the model. A well-fitting model was determined based on multiple fit indices, including a chi-square to degrees of freedom ratio (χ^2/df) less than 5.0, a Root Mean Square Error of Approximation (RMSEA) below 0.08, a Comparative Fit Index (CFI) greater than 0.90, a Goodness-of-Fit Index (GFI) above 0.90, a Normed Fit Index (NFI) exceeding 0.90, a Root Mean Square Residual (RMR) less than 0.05, and a Tucker-Lewis Index (TLI) above 0.90.

Project quality was further examined through item analysis, which included calculating corrected item-total correlations, mean scores, and standard deviations. A corrected item-total correlation of 0.3 or higher was interpreted as indicating acceptable internal consistency. Discriminative power was evaluated by comparing the top and bottom 27 % groups on each item using *t*-tests; items showing significant differences between the two groups were considered to demonstrate good discriminative validity. Measurement invariance across gender groups was tested via multi-group CFA (MG-CFA). Invariance was considered to hold if changes in fit indices—specifically, ΔCFI <0.01 and ΔRMSEA <0.015—remained within accepted thresholds.

Finally, a preliminary normality assessment revealed that all items satisfied the univariate normality criteria. Kurtosis values ranged from −0.85 (Item 23) to 1.28 (Item 16), and skewness from −0.92 (Item 27) to 0.68 (Item 14) in the pilot data. In the final data set, kurtosis ranged from −0.45 (Item 17) to 1.77 (Item 5), while skewness ranged from −1.16 (Item 5) to 0.23 (Item 7). All these values fell within the commonly accepted range of ±2, indicating that the data distributions were sufficiently close to normal for maximum likelihood estimation.

4. Results

4.1. The pilot test

A pilot test of the CTD scale was conducted to evaluate its factor structure and psychometric properties. The initial version of the scale consisted of 30 items distributed across six subdimensions, with five items in each subdimension. EFA was conducted to refine the instrument. The KMO value was 0.76, and Bartlett's Test of Sphericity was significant ($p < 0.05$), supporting the suitability of the data for factor analysis.

Nine initial factors were extracted, accounting for 59.09 % of the total variance. Items were removed based on both statistical thresholds and conceptual coherence: Item 10 and Item 2 were excluded due to factor loadings below 0.32, and Items 6, 14, 19, 23, 25, 27, and 30 were eliminated because they loaded on factors with fewer than three items. These refinements yielded a coherent factor solution comprising five interpretable dimensions. The factor loadings from the initial round of EFA are provided in Appendix B.

Notably, Item 24, initially developed to assess Analyticity, loaded more strongly on the Systematicity factor. Based on both empirical evidence and conceptual alignment, it was reassigned to a newly defined factor labeled Systematic-analyticity. The final structure comprised 21 items, categorized into five dimensions: Systematic-analyticity (4 items), Open-mindedness (3 items), Confidence in Reasoning (4 items), Reflective Skepticism (5 items), and Truth-seeking (5 items).

Reliability analysis revealed that, although some subdimensions demonstrated relatively low internal consistency (e.g., Analyticity: Cronbach's α = 0.406; McDonald's ω = 0.528), the overall reliability of the scale was high (Cronbach's α = 0.822; McDonald's ω = 0.855). No issues were reported regarding item comprehension or response process during the pilot test, indicating that the participants well understood the items.

4.2. The final test

4.2.1. Validity analysis: CFA, convergent and discriminant validity

To evaluate the construct validity of the refined 21-item CTD scale, CFA was conducted, followed by assessments of convergent and discriminant validity. These analyses aimed to verify the internal structure of the scale, the coherence of its dimensions, and the distinctiveness among factors.

CFA was performed using responses from 625 participants. The five-factor model demonstrated a good fit to the data, with the following indices: χ^2/df = 2.173, CFI = 0.964, GFI = 0.949, NFI = 0.935, RMR = 0.030, RMSEA = 0.043, and TLI = 0.956 (Table 4). Items 19 and 20 were removed due to standardized factor loadings coefficients of 0.476 and 0.351, respectively, which were below the acceptable threshold of 0.5. As a result, the final model retained 19 items, forming a robust and theoretically coherent five-factor structure for assessing CTD (see Fig. 1).

For the refined 19-item CTD scale, all items demonstrated standardized factor loadings exceeding the 0.50 threshold, indicating good construct validity. The scale showed satisfactory convergent validity, with AVE values ranging from 0.508 to 0.595, all exceeding the 0.50

Table 4
Statistical parameters of the CTD scale.

Statistics	Criteria of fit indices	CTD
χ^2/df	< 3 is excellent;	2.173
GFI	> 0.90	0.949
CFI	> 0.90	0.964
TLI	> 0.90	0.956
NFI	> 0.90	0.935
RMR	< 0.05	0.030
RMSEA	< 0.05 is excellent	0.043

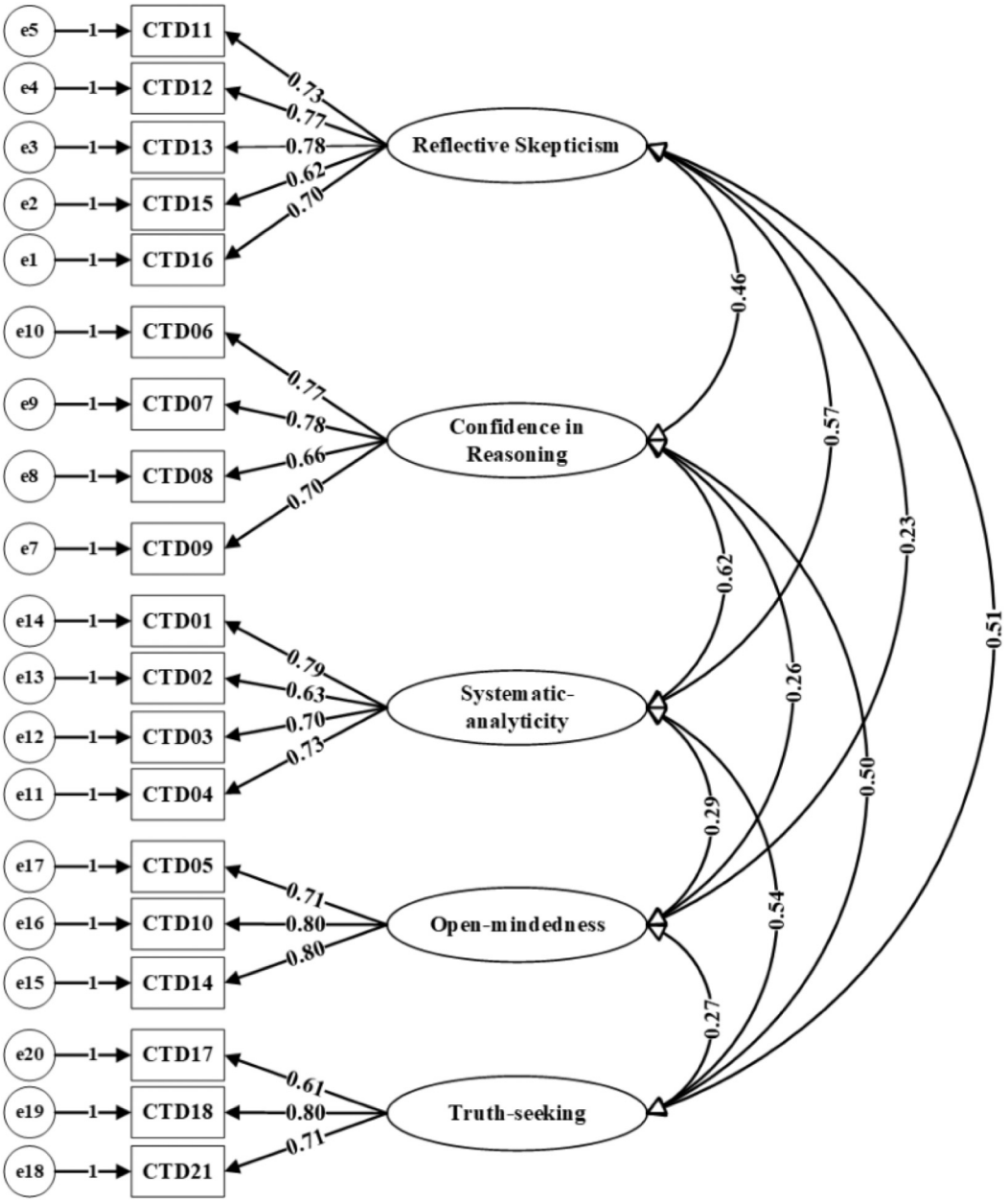


Fig. 1. The confirmatory factor analysis of the CTD.

threshold. Additionally, CR coefficients for the subscales ranged from 0.753 to 0.844, surpassing the 0.70 benchmark, further supporting convergent validity (Table 5).

Discriminant validity was assessed through cross-loading analysis. Each item loaded more strongly on its intended factor than on any other factor, with differences of at least 0.10. All items had primary loadings above 0.40, while cross-loadings on other factors remained below 0.30 (Appendix B). These results support the discriminant validity of the scale. Moreover, the square roots of the AVE values for each factor exceeded the inter-factor correlations (ranging from 0.228 to 0.618), fulfilling the Fornell–Larcker criterion (Fornell & Larcker, 1981) and providing further evidence of discriminant validity among the five subscales (Table 5).

Table 5
Latent variable correlations among the five subscales (n = 625).

	AVE	CR	Systematic-analyticity	Open-mindedness	Confidence in Reasoning	Reflective Skepticism	Truth-seeking
Systematic-analyticity	0.509	0.805	0.713				
Open-mindedness	0.595	0.815	0.289***	0.772			
Confidence in Reasoning	0.528	0.817	0.618***	0.258***	0.727		
Reflective Skepticism	0.522	0.844	0.574***	0.228***	0.460***	0.722	
Truth-seeking	0.508	0.753	0.544***	0.273***	0.496***	0.511***	0.713

Note.
*** The correlation is significant at the 0.001 level (bilateral).

4.2.2. Reliability analysis

After the removal of Items 19 and 20—due to their failure to meet the criteria in CFA—reliability analyses were conducted on the remaining 19 items. The overall Cronbach's alpha coefficient was 0.881, and McDonald's ω was 0.908, both indicating strong internal consistency of the CTD scale. For the five subscales, Cronbach's alpha values ranged from 0.744 to 0.843, and McDonald's ω coefficients ranged from 0.761 to 0.876 (Table 6). These results suggest that each subscale demonstrates acceptable to good reliability. The improvement in internal consistency compared to the pilot study likely resulted from the increased sample size and the deletion of statistically underperforming items during the exploratory phase.

4.2.3. Item analyses

Item analysis was conducted to examine the internal consistency and discriminative power of the CTD scale. As shown in Table 7, all items demonstrated corrected item-total correlations above the commonly accepted threshold of 0.30, suggesting satisfactory internal consistency across the scale. The highest corrected item-total correlation was observed for Item 4 with a value of 0.60, while the lowest was for Item 10, with a value of 0.34. The item means ranged from 3.12 to 4.16, and standard deviations ranged from 0.839 to 0.992, indicating a moderate dispersion of responses across items.

To evaluate the discriminative validity of the items, *t*-tests were conducted by comparing the top 27 % and bottom 27 % of participants based on total scale scores. All items showed statistically significant differences between the two groups ($p < 0.001$), demonstrating that the items effectively distinguish between individuals with higher and lower levels on the construct being measured (Table 7).

These results support the appropriateness of the items for inclusion in the scale, providing additional evidence of the instrument's measurement quality.

4.2.4. Measurement invariance testing across gender

A series of MG-CFA was conducted to evaluate the measurement invariance of the scale across gender. Four hierarchical models were evaluated: a configural model (with no parameter constraints), a metric invariance model (factor loadings constrained), a scalar invariance model (factor loadings and item intercepts constrained), and a residual invariance model (factor loadings, intercepts, and residual variances constrained).

Fit indices for each model are summarized in Table 8, all of which demonstrated adequate fit according to conventional criteria (CFI $\geq$ 0.90, RMSEA $\leq$0.06; Hu & Bentler, 1999). Specifically, the hypothesized factor structure was well represented across gender groups, with χ^2/df ratios consistently below 2. These results support the adequacy of the hypothesized factor structure across gender groups.

Model comparisons based on ΔCFI and ΔRMSEA are reported in Table 9. Across all nested comparisons, ΔCFI values were well below the recommended cutoff of 0.01 (Cheung & Rensvold, 2002), and ΔRMSEA values did not exceed the 0.015 criterion, indicating that additional constraints did not result in a significant loss of fit. These findings provide strong support for configural, metric, scalar, and residual invariance, affirming the measurement equivalence of the scale across gender.

5. Discussion

This study aimed to develop and validate a measurement tool for CTD that is tailored to Chinese high school students. It addresses limitations in existing instruments related to cultural adaptation, developmental relevance, and structural stability. Grounded in a robust theoretical framework, the study employed a multi-stage empirical design. The conceptual foundation was informed by Facione's seven-dimensional CTD model, which was further integrated with insights from Delphi expert consensus, Sosu (2013), and Manassero-Mas et al. (2022), resulting in a five-dimensional conceptual structure. Key concepts were semantically adapted to suit the linguistic and cultural context of the Chinese education system.

The scale development process proceeded through item generation, item review, iterative revision, and empirical validation, yielding an initial pool of 30 items. EFA and CFA ultimately supported a five-factor structure with 19 items, encompassing the following dimensions: Systematic-analyticity, Open-mindedness, Confidence in Reasoning, Reflective Skepticism, and Truth-seeking.

The dimensional structure of the CTD scale reflects both international theoretical consensus and culturally nuanced adaptations aligned with the cognitive dispositions of Chinese high school students. Notably, Reflective Skepticism emerged as a distinct dimension, in line with Sosu (2013), but diverging from previous frameworks such as the CCTDI and Manassero-Mas et al. (2022), where it was subsumed under broader constructs like Cognitive Maturity or Prudence-Analyticity. This divergence may reflect culturally specific conceptions in Chinese educational discourse, wherein adolescents interpret skepticism and reflection as an independent disposition indicative of critical engagement rather than as a subset of Cognitive Maturity dispositions.

Furthermore, Analyticity and Systematicity showed substantial empirical overlap in the EFA and were integrated into Systematic-analyticity, resonating with structured, procedural problem-solving styles reported in Chinese learners (Wang et al., 2019; Yuan et al., 2014). Taken together, the five-factor solution is theoretically coherent and internationally legible, while its culturally interpretable refinements provide a clear basis for testing applicability in non-Chinese settings (İskifoğlu & Ağazade, 2013; Orhan, 2022). At the same time, culturally sensitive dimensions—such as Analyticity–Systematicity merger—may not fully replicate outside China and therefore require empirical verification before broader adoption.

For use beyond China, cross-cultural applicability is best examined

Table 7
Descriptive statistics for CTD scale items.

Item	The corrected item-total correlation	Mean	SD	t	p
1	0.58	3.50	0.913	16.604	0.000
2	0.53	3.57	0.928	15.014	0.000
3	0.50	3.35	0.953	14.410	0.000
4	0.60	3.74	0.954	17.761	0.000
5	0.30	4.16	0.857	7.224	0.000
6	0.58	3.48	0.932	17.738	0.000
7	0.51	3.12	0.839	13.669	0.000
8	0.49	3.67	0.856	14.437	0.000
9	0.49	3.16	0.927	15.058	0.000
10	0.34	3.97	0.972	9.046	0.000
11	0.56	3.95	0.899	15.420	0.000
12	0.57	4.03	0.892	14.902	0.000
13	0.54	3.85	0.934	14.782	0.000
14	0.37	4.14	0.941	9.919	0.000
15	0.51	3.65	0.908	13.720	0.000
16	0.49	3.73	0.951	12.948	0.000
17	0.41	3.76	0.992	11.856	0.000
18	0.53	3.76	0.889	14.530	0.000
21	0.51	3.88	0.913	14.500	0.000

Table 6
Reliability analysis results of CTD scale.

	Total scale	Systematic-analyticity	Open-mindedness	Confidence in Reasoning	Reflective Skepticism	Truth-seeking
Omega	0.908	0.834	0.817	0.822	0.876	0.761
Cronbach's alpha	0.881	0.800	0.813	0.813	0.843	0.744

Table 8
Fit indices of measurement invariance models across gender.

Model	χ^2	df	χ^2/df	CFI	GFI	NFI	RMR	RMSEA	TLI
Configural	476.27	284	1.68	0.957	0.989	0.904	0.036	0.047	0.950
Metric	486.97	298	1.63	0.959	0.989	0.902	0.037	0.045	0.953
Scalar	501.45	312	1.61	0.959	0.989	0.899	0.038	0.044	0.955
Residual	548.08	331	1.66	0.953	0.988	0.890	0.039	0.046	0.952

Note. All models were estimated using MG-CFA across gender groups (Female and Male).

Table 9
Model comparison across nested invariance models.

Model Comparison	ΔCFI	ΔRMSEA
Configural vs. Metric	−0.001	0.002
Metric vs. Scalar	0.000	0.001
Scalar vs. Residual	0.006	−0.002

through a staged adaptation: bilingual translation/back-translation with cognitive interviewing to secure conceptual—not merely literal—equivalence, followed by multi-group CFA to evaluate configural, metric, and scalar invariance; where full scalar invariance is not observed, partial invariance can still support cautious mean comparisons. Our item-construction approach—anchoring each statement with a distinct keyword and screening for semantic conflicts—functions as a transferable design principle that minimizes respondent confusion. In practical terms, structural use across groups is warranted once configural and metric invariance are established, whereas cross-group mean comparisons should be avoided unless scalar (or partial scalar) invariance is demonstrated.

The 19-item CTD scale demonstrated strong psychometric properties, with a McDonald's ω coefficient of 0.908 and subdimension reliability ranging from 0.761 to 0.876, indicating high internal consistency. These values are comparable to those reported in international validations of the CCTDI, such as $\alpha = 0.87$ in Iran (Bakhtiari-Dovvombaygi et al., 2024) and $\alpha = 0.77$ in Quinn et al. (2020). These previously reported values serve as a reference point, placing the reliability of our scale within the range of commonly accepted psychometric standards in the field. This supports the conclusion that the CTD scale demonstrates acceptable reliability for educational assessment.

CFA supported the five-factor model, with model fit indices indicating strong adequacy (CFI = 0.964, RMSEA = 0.043). All dimensions demonstrated satisfactory convergent validity, with AVE values exceeding 0.50 and CR values exceeding 0.70. These findings are consistent with the recommendations by Hu and Bi (2025) on evaluating internal consistency and conceptual coherence using AVE and CR jointly.

A notable finding was the high correlation between Confidence in Reasoning and Systematic-analyticity ($r = 0.618$). This suggests that students who habitually work in a planful, orderly, and analytical manner also tend to feel more assured about their reasoning. Consistent with this pattern, Manassero-Mas et al. (2022) report a positive System–Self-Confidence correlation of approximately 0.40 and report that inter-factor correlations remain below 0.65, supporting related-yet-distinct constructs—mirroring our finding. At the processing level, converging evidence indicates that higher CTD aligns with more top-down, evaluative processing—congruent with a systematic-analytic style. Event-related potential (ERP) studies show that individuals with high CTD exhibit stronger P2 indices during explicit (top-down) creative processing, whereas those with low CTD show larger N1 amplitudes to low-creative stimuli than to high-creative stimuli in explicit tasks; implicit cognition shows no reliable differences. This indicates a systematic link between CTD and information-processing pathways (explicit/implicit; early/late components) and provides physiological evidence for the association between affective dispositions and processing styles (Zhang et al., 2023). This also supports the plausibility of the strong correlation between Confidence in Reasoning and Systematic-analyticity observed in this study.

MG-CFA provided empirical support for configural, metric, and scalar invariance across gender groups, with ΔCFI and ΔRMSEA values falling within accepted thresholds (Bravo et al., 2020; Memon et al., 2020; Wicherts & Dolan, 2010), affirming the scale's suitability for gender-based comparative analyses. Additionally, item-total correlation analyses revealed that all corrected correlations exceeded 0.30, and each item significantly differentiated between high and low scorers ($p < 0.001$), supporting strong item discrimination and internal reliability.

Although the scale demonstrated measurement invariance across gender groups, these findings may not generalize to other sociodemographic subgroups (e.g., geographic regions, school types, or cultural backgrounds). Given the considerable educational diversity across China, future studies should examine whether the scale retains measurement invariance across these populations, thereby supporting its broader applicability and fair use in varied educational settings.

Beyond psychometric rigor, the CTD scale offers considerable practical value. It provides educators with a standardized instrument to assess individual differences in thinking dispositions, enabling the development of targeted instructional strategies. For researchers, the scale facilitates exploration of CTD's relationships with constructs such as scientific creativity, metacognition, and learning motivation (Taşdelen Baş et al., 2022; Zhang et al., 2023). Previous studies have identified positive associations between CTD and scientific creativity, with creative self-efficacy serving as a mediating variable (Qiang et al., 2020). Future research could adopt joint modeling using tools such as Runco's Ideational Behavior Scale, the Metacognitive Awareness Inventory, and the Motivated Strategies for Learning Questionnaire to construct a nomological network of CTD and expand theory-driven inquiry into learning processes.

This study has several limitations. First, the sample was obtained through convenience sampling and included only 10th- and 11th-grade students from diverse regions of China (northern, central–western, and southeastern coastal). The absence of 12th-grade participants—who are preparing for the Gaokao—limits generalizability. Accordingly, the scale should be used cautiously with 12th graders (avoid high-stakes decisions until grade-specific evidence is available), as students nearing a high-stakes exam may display different motivational and affective profiles that could shift both dispositional levels and inter-factor correlations. At present, the extent of generalization to 12th graders is unknown and should be established empirically. Future research will employ stratified sampling that includes 12th graders, examine measurement invariance across grades (10/11/12), and compare latent means to determine whether the scale functions equivalently for students approaching the Gaokao. Moreover, although the pilot EFA included $N = 183$ cases—meeting commonly cited minimums—it remains a substantive limitation for factor exploration. The exploratory structure should be treated as provisional and interpreted with caution; replication with larger, independently sampled cohorts and confirmatory cross-validation (e.g., hold-out CFA and multi-group invariance tests) will be conducted to verify the factor solution.

Second, while the five-point Likert format is efficient and widely adopted, it may be susceptible to social desirability bias and central tendency bias. Future iterations may incorporate behavioral measures or teacher-based assessments to triangulate self-reported data and

strengthen the robustness of findings.

Third, the cross-sectional design limits insights into the temporal development and responsiveness of CTD to interventions. Longitudinal research is recommended to examine developmental trajectories and assess changes over time.

Lastly, although preliminary recommendations for implementation have been offered, further efforts are needed to develop user manuals, scoring rubrics, and professional training modules. These tools would support educators and facilitate the effective implementation of the scale across diverse educational settings.

6. Conclusions

This study developed and validated a CTD scale specifically for Chinese high school students. Starting from an initial pool of 30 items across six dimensions, a final five-factor model with 19 items was established through rigorous item development, pilot testing, and psychometric validation. The five retained dimensions—Systematic-analyticity, Open-mindedness, Confidence in Reasoning, Reflective Skepticism, and Truth-seeking—demonstrated good reliability, construct validity, and measurement invariance across gender groups. The scale offers a reliable and theoretically grounded tool for assessing CTD in secondary education contexts. It has the potential to inform both research and instructional practice aimed at cultivating CT among adolescent learners.

CRediT authorship contribution statement

Wencheng Liu: Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Huangdong Ma:** Writing – review & editing, Investigation, Data curation. **Shiwen Ma:** Writing – review & editing, Data curation. **Ying Chen:** Writing – review & editing, Methodology, Formal analysis. **Gaofeng Li:** Writing – review & editing, Supervision, Project administration, Methodology, Funding acquisition, Conceptualization.

Funding

This work was supported by the Planning Subject for the 14th Five-year Plan of National Education Sciences under Grant number ECA240440.

Declaration of competing interest

No potential conflict of interest was reported by the authors.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.actpsy.2025.106055.

Data availability

Data will be made available on request.

References

Altun, E., & Yildirim, N. (2023). What does critical thinking mean? Examination of pre-service teachers' cognitive structures and definitions for critical thinking. *Thinking Skills and Creativity, 49*, Article 101367. https://doi.org/10.1016/j.tsc.2023.101367

Bakhtiari-Dovvombaygi, H., Pourhasan, K., Rahmaty, Z., Zare-Kaseb, A., Abbaszadeh, A., Rashtbarzadeh, A., & Borhani, F. (2024). Evaluation of cross-cultural adaptation and validation of the Persian version of the critical thinking disposition scale: Methodological study. *BMC Nursing, 23*(1), 463. https://doi.org/10.1186/s12912-024-02129-y

Bielik, T., & Krüger, D. (2024). Perceived relevance of critical thinking aspects for biology graduate students. *Journal of Biological Education, 58*(1), 166–181. https://doi.org/10.1080/00219266.2022.2026806

Boso, C. M., van der Merwe, A. S., & Gross, J. (2021). Critical thinking disposition of nursing students: A quantitative investigation. *Nurse Education in Practice, 55*, Article 103167. https://doi.org/10.1016/j.nepr.2021.103167

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

Çelik İskifoğlu, T., Çerkez, Y., & İskifoğlu, G. (2022). Thinking culture and critical thinking dispositions of high school students in Turkish republic of northern Cyprus. *Frontiers in Psychology, 13*, Article 1017747. https://doi.org/10.3389/fpsyg.2022.1017747

Chen, J., Wang, X., & Zheng, X. (2024). The investigation of critical thinking disposition among Chinese primary and middle school students. *Thinking Skills and Creativity, 51*, Article 101444. https://doi.org/10.1016/j.tsc.2023.101444

Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5

Cui, L., Zhu, Y., Qu, J., Tie, L., Wang, Z., & Qu, B. (2021). Psychometric properties of the critical thinking disposition assessment test amongst medical students in China: A cross-sectional study. *BMC Medical Education, 21*(1), 10. https://doi.org/10.1186/s12909-020-02437-2

Demircioğlu, E., & Kilmen, S. (2014). An adaptation study of critical thinking disposition scale. *The Journal of Academic Social Science Studies, 8*, 203–208. https://doi.org/10.9761/JASSS2434

Dwyer, C. P., Hogan, M. J., & Stewart, I. (2014). An integrated critical thinking framework for the 21st century. *Thinking Skills and Creativity, 12*, 43–52. https://doi.org/10.1016/j.tsc.2013.12.004

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. (2015). Critical thinking: A streamlined conception. In M. Davies, & R. Barnett (Eds.), *The Palgrave handbook of critical thinking in higher education* (pp. 31–47). Palgrave MacMillan. https://doi.org/10.1057/9781137378057_2.

Ertaş Kılıç, H., & Şen, A. (2014). Turkish adaptation study of UF/EMI critical thinking disposition instrument. *Eğitim ve Bilim, 39*, 1–12. https://doi.org/10.15390/EB.2014.3632

Fabio, R. A., Plebe, A., & Suriano, R. (2024). AI-based chatbot interactions and critical thinking skills: An exploratory study. *Current Psychology. Advance online publication.*. https://doi.org/10.1007/s12144-024-06795-8

Facione, N. C., Facione, P. A., & Sánchez, C. A. (1994). Critical thinking disposition as a measure of competent clinical judgment: The development of the California critical thinking disposition inventory. *Journal of Nursing Education, 33*(8), 345–350. https://doi.org/10.3928/0148-4834-19941001-05

Facione, P., Facione, N., & Giancarlo, C. A. (2000). The disposition toward critical thinking: Its character, measurement, and relationship to critical thinking skill. *Informal Logic, 20*(1), 61–84. https://doi.org/10.22329/il.v20i1.2254

Facione, P. A. (1990). *Critical thinking: A statement of expert consensus for purposes of educational assessment and instruction (the delphi report, ERIC document no. ED315423)*. California Academic Press. https://files.eric.ed.gov/fulltext/ED315423.pdf.

Fornell, C., & Larcker, D. F. (1981). Structural equation models with unobservable variables and measurement error: Algebra and statistics. *Journal of Marketing Research, 18*(3), 382–388. https://doi.org/10.1177/002224378101800313

Gao, Q., Cao, Y., Xie, H., & Li, X. (2025). Investigating the nature of science in reformed Chinese biology curriculum standards. *Science & Education, 34*(1), 315–343. https://doi.org/10.1007/s11191-023-00466-z

Giancarlo, C. A., Blohm, S. W., & Urdan, T. (2004). Assessing secondary students' disposition toward critical thinking: Development of the California measure of mental motivation. *Educational and Psychological Measurement, 64*(2), 347–364. https://doi.org/10.1177/0013164403258464

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains. Dispositions, skills, structure training, and metacognitive monitoring. *American Psychologist, 53*(4), 449–455. https://doi.org/10.1037/0003-066x.53.4.449

Halpern, D. F., & Dunn, D. S. (2021). Critical thinking: A model of intelligence for solving real-world problems. *Journal of Intelligence, 9*(2), 22. https://doi.org/10.3390/jintelligence9020022

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

Hu, X., & Bi, H. (2025). Exploring and validating the componential model of students' scientific critical thinking in science education. *Thinking Skills and Creativity, 55*, Article 101695. https://doi.org/10.1016/j.tsc.2024.101695

Hwang, J., Hand, B., & French, B. F. (2023). Critical thinking skills and science achievement: A latent profile analysis. *Thinking Skills and Creativity, 49*, Article 101349. https://doi.org/10.1016/j.tsc.2023.101349

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*(2), 187–196. https://doi.org/10.2224/sbp.2013.41.2.187

Kam, C. C. S. (2020). Expanded format shows better response consistency than Likert-scale format in the measurement of optimism. *Personality and Individual Differences, 152*, Article 109606. https://doi.org/10.1016/j.paid.2019.109606

Kawashima, N., & Shiomi, K. (2007). Factors of the thinking disposition of Japanese high school students. *Social Behavior and Personality: An International Journal, 35*(2), 187–194. https://doi.org/10.2224/sbp.2007.35.2.187

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

Ku, K. Y. L., & Ho, I. T. (2010). Dispositional factors predicting Chinese students' critical thinking performance. *Personality and Individual Differences, 48*(1), 54–58. https://doi.org/10.1016/j.paid.2009.08.015

Li, X., & Liu, J. (2021). Mapping the taxonomy of critical thinking ability in EFL. *Thinking Skills and Creativity, 41*, Article 100880. https://doi.org/10.1016/j.tsc.2021.100880

Lin, M., Liu, L. Y. J., & Pham, T. N. (2023). Towards developing a critical learning skills framework for master's students: Evidence from a UK university. *Thinking Skills and Creativity, 48*, Article 101267. https://doi.org/10.1016/j.tsc.2023.101267

Liu, Y., & Pásztor, A. (2022). 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. (2023). Survey on the influential demographic factors of Chinese undergraduate students' critical thinking disposition: Evidence from plausible values. *Thinking Skills and Creativity, 50*, Article 101397. https://doi.org/10.1016/j.tsc.2023.101397

MacCallum, R. C., Widaman, K. F., Zhang, S., & Hong, S. (1999). Sample size in factor analysis. *Psychological Methods, 4*(1), 84–99. https://doi.org/10.1037/1082-989X.4.1.84

Manassero-Mas, M. A., Moreno-Salvo, A., & Vázquez-Alonso, Á. (2022). Development of an instrument to assess young people's attitudes toward critical thinking. *Thinking Skills and Creativity, 45*, Article 101100. https://doi.org/10.1016/j.tsc.2022.101100

Memon, M. A., Ting, H., Cheah, J.-H., Thurasamy, R., Chuah, F., & Cham, T. H. (2020). Sample size for survey research: Review and recommendations. *Journal of Applied Structural Equation Modeling, 4*(2), i–xx. https://doi.org/10.47263/JASEM.4(2)01

Orhan, A. (2022). California critical thinking disposition inventory: Reliability generalization meta-analysis. *Journal of Psychoeducational Assessment, 40*(2), 202–220. https://doi.org/10.1177/07342829211048962

Perkins, D. N., Jay, E., & Tishman, S. (1993). Beyond abilities: A dispositional theory of thinking. *Merrill-Palmer Quarterly, 39*(1), 1–21. https://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), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Qiang, R., Han, Q., Guo, Y., Bai, J., & Karwowski, M. (2020). Critical thinking disposition and scientific creativity: The mediating role of creative self-efficacy. *The Journal of Creative Behavior, 54*(1), 90–99. https://doi.org/10.1002/jocb.347

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

Ricketts, J. C., & Rudd, R. D. (2005). Critical thinking skills of selected youth leaders: The efficacy of critical thinking dispositions, leadership, and academic performance. *Journal of Agricultural Education, 46*(1), 32–43. https://doi.org/10.5032/jae.2005.01032

Santos-Meneses, L. F., Pashchenko, T., & Mikhailova, A. (2023). Critical thinking in the context of adult learning through PBL and e-learning: A course framework. *Thinking Skills and Creativity, 49*, Article 101358. https://doi.org/10.1016/j.tsc.2023.101358

Sarigoz, O. (2012). Assessment of the high school students' critical thinking skills. *Procedia-Social and Behavioral Sciences, 46*, 5315–5319. https://doi.org/10.1016/j.sbspro.2012.06.430

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

Taşdelen Baş, M., Özpulat, F., Molu, B., & Dönmez, H. (2022). The effect of decorative arts course on nursing students' creativity and critical thinking dispositions. *Nurse Education Today, 119*, Article 105584. https://doi.org/10.1016/j.nedt.2022.105584

Walter, Y. (2024). Embracing the future of artificial intelligence in the classroom: The relevance of AI literacy, prompt engineering, and critical thinking in modern education. *International Journal of Educational Technology in Higher Education, 21*(1), 15. https://doi.org/10.1186/s41239-024-00448-3

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*(1), 200. https://doi.org/10.1186/s12909-019-1593-z

Wicherts, J. M., & Dolan, C. V. (2010). Measurement invariance in confirmatory factor analysis: An illustration using IQ test performance of minorities. *Educational Measurement: Issues and Practice, 29*(3), 39–47. https://doi.org/10.1111/j.1745-3992.2010.00182.x

Worthington, R. L., & Whittaker, T. A. (2006). Scale development research: A content analysis and recommendations for best practices. *The Counseling Psychologist, 34*, 806–838. https://doi.org/10.1177/0011000006288127

Yanchar, S. C., Slife, B. D., & Warne, R. (2008). Critical thinking as disciplinary practice. *Review of General Psychology, 12*(3), 265–281. https://doi.org/10.1037/1089-2680.12.3.265

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

Yin, X., Saad, M. R. B. M., & Halim, H. B. A. (2023). A systematic review of critical thinking instructional pedagogies in EFL writing: What do we know from a decade of research. *Thinking Skills and Creativity, 49*, Article 101363. https://doi.org/10.1016/j.tsc.2023.101363

Yuan, S.-P., Liao, H.-C., Wang, Y.-h., & Chou, M.-J. (2014). Development of a scale to measure the critical thinking disposition of medical care professionals. *Social Behavior and Personality: An International Journal, 42*(2), 303–312. https://doi.org/10.2224/sbp.2014.42.2.303

Zhang, Y., Bian, Y., Wu, H., Tang, W., & Li, Q. (2023). Intuition or rationality: Impact of critical thinking dispositions on the cognitive processing of creative information. *Thinking Skills and Creativity, 48*, Article 101278. https://doi.org/10.1016/j.tsc.2023.101278

Zoller, U., Ben-Chaim, D., Ron, S., Pentimalli, R., Scolastica, S., Chiara, M. S., & Borsese, A. (2000). The disposition toward critical thinking of high school and university science students: An interintra Israeli-Italian study. *International Journal of Science Education, 22*(6), 571–582. https://doi.org/10.1080/095006900289679