Research and Development of Computational Thinking Assessment for College Students Integrating Cognitive and Non-Cognitive Abilities

[1st] Wei Zhang[a], [2nd] * Panpan Li[b], [3rd] Xinyao Zeng[c], [4th] Lingling Song[d]

[a]zwccnu@ccnu.edu.cn, * Corresponding author: [b]panpanli@mails.ccnu.edu.cn, [c]zengxy_98@mails.ccnu.edu.cn, [d]linglingsong@mails.ccnu.edu.cn

Department of Artificial Intelligence Education, Central China Normal University, Wuhan, Hubei, China

Abstract. Currently, the tools for assessing the computational thinking(CT) abilities of college students lack comprehensive coverage, focusing only on the cognitive aspects of CT and neglecting exploration of non-cognitive abilities. In this study, through a literature review, a theoretical framework for college students' CT was constructed, encompassing both cognitive and non-cognitive abilities. The framework includes four core dimensions: conceptual knowledge, algorithmic thinking, problem exploration, and values, along with 11 secondary indicators. Based on this framework, a test tool suitable for assessing college students' CT was developed. The tool consists of 18 questions assessing non-cognitive abilities and 22 questions assessing cognitive abilities. Statistical analysis methods and multidimensional item response theory were employed to analyze the empirical results of the test. The results indicate that the tool has reasonable difficulty and discrimination; the α reliability coefficient method shows good internal consistency; factor analysis results demonstrate a good fit between the theoretical model and observed data, indicating high structural validity. Overall, the test tool exhibits high quality and performance, and the comprehensive dimensionality provides important guidance for the reform and development of CT in higher education.

Keywords: Computational thinking; Assessment; Multidimensional item response theory; Test tool development; Higher education

1. INTRODUCTION

Computational thinking(CT), regarded as an essential professional competency in the era of artificial intelligence [1], effectively assists students in engaging with and understanding the computational world. It not only cultivates students' ability to solve problems using the powerful computational capabilities of computers but also trains them to approach problem-solving from the perspective of computer scientists. Therefore, CT is considered a new literacy for the 21st century, alongside reading, writing, and arithmetic[2]. As a complex cognitive skill set, CT encompasses various cognitive skills for problem-solving [3]. Encouraging the development of CT in everyone has become a shared goal among numerous experts and scholars[4].

In the process of CT development, assessment plays a crucial role[5]. Reliable and effective assessment not only analyzes the extent to which students master CT but also provides a basis

ITEI 2023, November 24-26, Zhengzhou, People's Republic of China

DOI 10.4108/eai.24-11-2023.2343713

for evaluating the effectiveness of CT cultivation[6]. Currently, research and development in CT assessment predominantly focus on the K-12 education stage[7]. In contrast to the development of CT in the K-12 stage, the progress of CT in higher education is considered slow and less emphasized [8]. Higher education, as a crucial gateway for producing talents for society, requires highly skilled individuals with innovative capabilities and comprehensive qualities[9]. Hence, students need to possess CT literacy oriented towards problem-solving. A literature review reveals a limited focus on evaluating CT in higher education, with most studies leaning towards measuring cognitive abilities while neglecting students' non-cognitive abilities[10]. Non-cognitive abilities, as a distinct capability, play a crucial role in computation, learning, and problem-solving [11]. Scholars in science education have researched non-cognitive abilities in CT from various perspectives and levels. For example, Xue et al. [12] studied computer vision, finding that students had an advantage in attention retention but lacked sustained and stable executive control. Additionally, Tang et al. [13] systematically investigated non-cognitive abilities in CT from theoretical and experimental perspectives. These works indicate a close connection between CT and non-cognitive abilities. The non-intellectual factors such as knowledge, thoughts, culture, and emotions manifested in the CT process are crucial factors influencing the improvement and innovative development of CT abilities, as well as the formation and retention of learning outcomes [14].

Recognizing the importance of non-cognitive abilities in the development of CT, this study aims to develop a theoretical framework for CT by integrating multiple definitions and identifying its core elements. Using this framework, we will develop a CT assessment tool (CAPV-t) for cognitive and non-cognitive abilities in college students, and analyze and validate CAPV-t through statistical analysis methods and multidimensional item response theory. In summary, this research aims to develop a comprehensive and effective assessment tool to evaluate various dimensions of CT abilities in college students, analyze the relationships between multiple abilities, identify students' shortcomings in CT learning, and provide a standardized assessment tool for CT education.

To ensure that the assessment of CT covers the entire educational stage, particularly addressing the neglect of CT abilities in college students, this study will focus on developing an effective tool to assess both cognitive and non-cognitive CT abilities in college students. The tool is expected to demonstrate high reliability and validity. In the following sections, we will address the following questions:

Q1: What is the reliability and validity of the assessment tool we developed?

Q2: How is the quality of the assessment tool?

2. LITERATURE REVIEW

2.1. The concept of CT

CT was first proposed by Papert [15], and its concept was further clarified by Wing in a influential article in 2006. Wing defined CT as a cognitive activity that involves solving problems, designing systems, and understanding human behavior through the foundational concepts of computer science. While CT originated in the field of computer science, research has primarily focused on the domain of programming. Brennan and Resnick introduced a threedimensional framework for CT, encompassing both cognitive and non-cognitive abilities and expanding the understanding of CT through the lenses of computational concepts, practices, and perspectives[16]. Other researchers, such as Korkmaz and Bai [17], as well as Selby and Woollard [18], have provided additional definitions and breakdowns of CT from various perspectives.

CT is not only widely applied in the fields of computer science and engineering but also demonstrates its influence across other disciplines [19]. Researchers have explored diverse and scattered applications in higher education, including integrating CT into university-level computer science courses, adopting microlearning approaches to enhance CT among professional students, and incorporating it into undergraduate courses across various disciplines[20-22].

As a tool for undergraduate students to address professional problems, CT holds profound significance in changing thinking patterns and driving technological development and societal progress. Although there is no universally agreed-upon definition of CT, researchers and educators unanimously recognize its primary focus on developing students' cognitive abilities.

2.2. CT assessment

MetExtensive literature review reveals a predominant focus on integrating CT into undergraduate courses, with limited attention to universally applicable assessment tools for evaluating higher education CT levels and efficacy [23]. However, assessment is a crucial component of CT development. Internationally recognized tools are often programming-based, evaluating students through coding tasks or projects. For instance, the PECT CT development model [24] and Dr.Scratch for K-12 students [25] provide summative assessments based on specific programming tasks.

While programming-based assessments have limitations, there is a growing trend towards question-based assessments for their convenience and quick feedback. Notable examples include the CTt for K-12 students in Spain [26] and the Bebras Tasks challenge for K-12 students in Lithuania [27]. Additionally, scale-based tools like the CTS scale for university students in Turkey [28] have been developed, covering dimensions like creativity, algorithmic thinking, critical thinking, collaboration, and problem-solving.

Classic assessment tools primarily focus on K-12 education, leaving a gap in effective tools for evaluating CT in higher education. The existing literature on higher education CT assessment can be broadly categorized into qualitative and quantitative assessments. While most assessments are qualitative and susceptible to subjective influences, a few quantitative tools face challenges in robust development and validation methods.

Our subsequent research will concentrate on developing a comprehensive test tool to assess both cognitive and non-cognitive aspects of CT in university students. We will employ statistical analysis methods and multidimensional item response theory to analyze and validate the reliability and validity of this assessment tool.

3. METHOD

3.1. Development of assessment tools

1) Assessment framework

After an extensive literature review, it is evident that various experts, scholars, and academic organizations have different interpretations of CT. However, the majority agree that CT is a cognitive process that occurs when formulating solutions to problems. In the process of applying CT to problem-solving, not only foundational knowledge of computer concepts is utilized, but also abstract thinking is employed to transform real-world problems into ones that can be addressed by computers. Furthermore, specific algorithms are used to solve these problems. Therefore, at the cognitive level of CT, we identify three core dimensions: conceptual knowledge, algorithmic thinking, and problem exploration.

CT, as a mature skill, not only involves cognitive abilities but also requires non-cognitive abilities. Non-cognitive abilities refer to attitudes, self-confidence, teamwork, and other non-intellectual factors that students exhibit during the learning or application of CT skills. These non-cognitive factors are an integral part of cultivating CT abilities. Hence, we encompass students' attitudes, teamwork, inquiry, and creativity under the umbrella term "value perspectives" as the core dimensions of non-cognitive abilities in CT.

- Conceptual knowledge is the basis for learners to carry out CT practice activities. It can be broadly classified into two categories: definitional knowledge, which involves the fundamental definitions of specific concepts, and operational knowledge, encompassing the basic operational concepts potentially utilized in programming.

- Algorithmic thinking denotes a learner's capacity to formulate computer algorithms for solving specific problems. The practical application involves grasping a series of algorithmic instructions and participating in algorithm design. Therefore, students' algorithmic thinking abilities can be evaluated through two dimensions: understanding algorithms and designing algorithms.

- Problem exploration refers to the learner's ability to analyze and solve problems using computational methods. The process involves abstraction, decomposition, and transfer.

- Values refer to the constantly evolving understanding of how learners interact with others and perceive the world during the CT learning process. This aspect can be assessed across four dimensions: attitude, questioning, teamwork, and creativity.

CT is the process of problem-solving. In the elucidation of the core dimensions, this study seeks to concretize the elements by finely dividing them into secondary dimensions, thereby obtaining more precise measurement criteria. In summary, this research will utilize conceptual knowledge, algorithmic thinking, problem exploration, and values as primary core dimensions. The secondary dimensions will include definitional knowledge, operational knowledge, algorithm understanding, algorithm design, abstraction, decomposition, transfer, attitude, teamwork, questioning, and creativity, serving as the basis for developing the testing tool. Specific details and connotations are described in Table 1.

Table 1: The Theoretical Framework for CT Cognitive and Non-Cognitive Abilities

Ability division	First level dimension	Secondary Dimension	Connotation Description
Cognitive ability	Conceptual knowledge(C)	Definitional knowledge (C1)	Basic computer definitions, such as events, operators, data types, variables, constants, etc.
		Operational knowledge (C2)	Basic operational concepts of programming in computers, such as sequence, judgment, loops, parallelism, conditionals, etc.
	Algorithmic thinking (A)	Algorithm understanding (A1)	Be able to understand the basic algorithmic steps.
		Algorithm design (A2)	Be able to correctly design a series of algorithmic steps for effective problem-solving.
	Problem exploration (P)	Abstraction (P1)	Extracting common, essential characteristics from things.
		Decomposition (P2)	Break down the problem into smaller problems that can be easily solved.
		Transfer (P3)	Apply the problem-solving approach to other problem scenarios.
Non-cognitive ability	Values (V)	Attitude (V1)	Interest, confidence and self-efficacy in computer science and programming.
		Questioning (V2)	Challenge and doubt existing knowledge and solutions, and pursue a deeper and more comprehensive understanding.
		Teamwork (V3)	Actively integrate into the team and work with the team to accomplish tasks together.
		Creativity (V4)	Generate new ideas and solve problems independently and innovatively.

2) Assessment tool

Based on the theoretical framework proposed earlier, the development of the testing tool involves categorizing dimensions into cognitive and non-cognitive abilities. Cognitive abilities refer to the mastery of specific skills, typically measurable through test questions. Non-cognitive abilities are associated with a student's personality, encompassing different attitudes towards tasks and proactive behavior, which are challenging to measure through test questions. Therefore, the testing tool is designed in two parts: the first part assesses students' CT value perspectives, examining non-cognitive abilities through a scale. The second part evaluates students' conceptual knowledge, algorithmic thinking, and problem exploration, assessing cognitive abilities through a questionnaire.

To assess various dimensions of CT in college students, this study, based on existing test questions and scales such as Bebras Tasks CT Challenge, CTt, The Fairy Assessment, and CTS, combined with college students' life experiences and characteristics, adapted and initially formulated 26 test questions and 14 scale questions. The scale section was uniformly designed as scoring multiple-choice questions, each with four options. The score range was set from 0 to 3 points, with each option corresponding to a score based on the degree to which it reflected the ability dimension. For the questionnaire section, question types were set as multiple-choice and fill-in-the-blank, with a secondary scoring system of 0 and 1. After completing the item development, two experts in artificial intelligence and two experts in educational technology were invited to evaluate the tool separately. Items with unclear assessment dimensions were removed based on expert judgments, and several scale questions assessing students' non-cognitive abilities were added. Subsequently, a small-scale test was conducted based on expert suggestions, and the results were analyzed through student feedback and submissions. Items with unclear meanings and high difficulty were eliminated, and some items were revised. Ultimately, 18 scale questions and 22 test questions remained. Each item involved one or more abilities, and all questions were designed using a multidimensional model within each item. During the determination of dimensions, experts repeatedly deliberated and calibrated the basic abilities addressed by each test question. The Q-matrix for the basic abilities assessed by each item is shown in Appendix 1.

3.2. Participants

The study focused on undergraduate students from a comprehensive university in China. We employed a random sampling method and selected 400 students from different grades and majors as the study sample. The sample included both male and female students, with ages ranging from 18 to 25 years. To ensure the representativeness of the study, we considered a balanced representation across various academic disciplines and aimed to cover different colleges within the university. Out of the 387 students who volunteered to participate, 366 completed a 45-minute test online. Prior to the test, students were required to provide basic personal information, including gender, age, major, school, and whether they had prior programming experience. After excluding 34 incomplete reports, we ultimately obtained 366 valid datasets, resulting in a data validity rate of 94%. The distribution of the sample is presented in Table 2.

Table 2: Participant demographics

	Boy	**Girl**	**total**
freshman	53	28	81
sophomore	43	46	89
senior	56	55	111
junior	48	37	85
Total	200	166	366

3.3. Data Analysis

1) Verification of CAPV-t

The validation of the tool involves reliability analysis, validity analysis, and parameter estimation analysis, which can be divided into the validation of the scale section and the validation of the test paper section. For the scale section, SPSS software was used for analysis, while for the test paper section, as it assesses multiple dimensions, MIRT was employed to analyze the relationships among different dimensions. The analysis used the mirt package and psych package in R.

2) Reliability analysis

For the scale section, intrinsic reliability is generally considered, which examines whether there is high internal consistency among items. The commonly used method for analysis is the alpha reliability coefficient. For the test paper section, the psych package was utilized for reliability analysis. Internal consistency of the test paper was assessed through the analysis of Cronbach's α coefficient. A Cronbach's α coefficient greater than 0.7 indicates high reliability of the test paper.

3) Validity analysis

The validation of CAPV-t involves two aspects: content validity and structural validity. Content validity is assessed using the Delphi method and analysis of pre-test results. Structural validity encompasses both the scale and the test paper. Exploratory factor analysis is employed for the structural validity of the scale, with the suitability for factor analysis determined through the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's test [29]. If deemed suitable, an analysis of factor loadings for each item across dimensions is conducted[30].For the test paper, structural validity is evaluated by analyzing the discrimination parameters of test questions. In R, the mirt package is utilized for assessing the structural validity of the test paper[31].Initially, unidimensional item response theory (UIRT) and 5-dimensional, 6-dimensional, and 7-dimensional MIRT models are fitted to the data. The fit is evaluated using Akaike information criterion (AIC), Bayesian information criterion (BIC), Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and Root-Mean-Square Error of Approximation (RMSEA). If a multidimensional model demonstrates superior fit, further exploration for the optimal dimensions is conducted.Subsequently, exploratory factor analysis is employed to identify the test dimensions. The QMCEM [32] algorithm is used to fit the data, and a discrimination matrix with a critical value of 0.25 is applied to filter out discriminations above this threshold, approximating the attribution of items to each dimension.

4) Parameter estimation

When conducting a quality analysis of the tool, it is essential to first choose an appropriate model to ensure the attainment of desired outcomes [33]. For the validation of questions 19-36, considering these items are binary scoring single-choice questions and recognizing the possibility of student guessing in multiple-choice questions, we consider guessing as a significant factor to more accurately reflect students' actual proficiency. Therefore, we employed the three-parameter multidimensional logistic model (3PL) proposed by Reckase [34]for modeling, encompassing three key parameters: difficulty parameter (d), discrimination parameter (a), and guessing parameter (c). It's noteworthy that the value of parameter c does not vary with changes in the ability level of examinees; even among examinees with the lowest and highest abilities, the probability of guessing the correct answer remains the same. To ensure the validity of results, the theoretical range for parameter c is 0-1, but in practice, it generally does not exceed 0.35.

For questions 37-40, which are binary scoring fill-in-the-blank questions, we opted for the two-parameter multidimensional logistic model (2PL). In comparison to the 3PL model, the 2PL model reduces the guessing coefficient in parameter settings. This choice takes into account the nature of fill-in-the-blank questions and more precisely reflects student performance during the modeling process. The application of this modeling strategy contributes to better adapting to the characteristics of different question types and enhances the accuracy of understanding students' test-taking behavior.

4. RESULTS

4.1. Reliability and validity of CAPV-t (answer to Q1)

1) Reliability of CAPV-t

Table 3 presents the overall reliability of the Non-Cognitive Ability Beliefs Inventory and the reliability of each sub-dimension. According to the results in the table, the reliabilities for Attitude, Questioning, Teamwork, and Creativity are all greater than 0.85. The overall reliability of the Inventory exceeds 0.7, indicating a high level of reliability for assessing students' non-cognitive abilities using this inventory.

Table 3: Cronbach's α consistency coefficient test results of the scale

	Cronbach's Alpha	**Number**
Attitude (V1)	0.909	4
Questioning (V2)	0.950	4
Teamwork (V3)	0.860	5
Creativity (V4)	0.919	5
Values	0.703	18

For the cognitive abilities section, reliability analysis of the questionnaire was conducted using the psych package in R, and the results indicate high reliability for the overall questionnaire and its various sub-dimensions. The Cronbach's α value for the entire sample is 0.8, with α values of 0.77 for conceptual knowledge, 0.81 for algorithmic thinking, and 0.71 for problem exploration. The overall consistency coefficients for the questionnaire and its sub-dimensions are all above 0.7, indicating a high level of reliability for the questionnaire. Therefore, the questionnaire meets the requirements for test development.

2) Validity of CAPV-t

Initially, a Kaiser-Meyer-Olkin (KMO) test was conducted on the non-cognitive abilities inventory, yielding a KMO value of 0.863 with a significance level below 0.05, indicating suitability for factor analysis. Subsequently, principal component analysis with maximum variance orthogonal rotation identified four factors, explaining a cumulative variance of 76.645%, exceeding 60%. This suggests that the four extracted factors effectively represent the entire dataset. The rotated component matrix, presented in Table 4, indicates high factor loadings for items T1, T2, T3, and T13 on the first factor, labeled "Attitude"; items T4 to T7 on the second factor, labeled "Questioning"; items T8 to T12 on the third factor, labeled "Teamwork"; and items T13 to T18 on the fourth factor, labeled "Creativity." The component matrix aligns perfectly with the original dimensions of the inventory (Appendix 1), affirming the inventory's robust validity.

Table 4: Matrix of components after rotation

Item		T1	T2	T3	T4	T5	T6	T7	T8	T9	T10	T11	T12	T13	T14	T15	T16	T17	T18
Components	1	0.867	.872	.870										.866					
	2				.894	.879	.923	.933											
	3								.789	.868	.721	.719	.813						
	4														.812	.842	.855	.878	.866

Validity analysis was conducted for the cognitive abilities section by comparing the fit results of Unidimensional Item Response Theory (UIRT) with 5D, 6D, and 7D Multidimensional Item Response Theory (MIRT) models. The fit results, shown in Table 5, indicate that smaller values of AIC and BIC represent better model fit, while larger values of CFI and TLI suggest improved model fit (typically above 0.9), and smaller RMSEA values indicate better model fit (usually below 0.08). The table reveals that multidimensional models outperform the UIRT model, with the 7D model providing the best fit to the data.

Table 5: Comparison of multiple model fit indicators

Dimensionality	AIC	BIC	CFI	TLI	RMSEA
UIRT	9756.568	9928.283	0.493	0.436	0.113
5-Factor Model	9234.831	9796.811	0.719	0.671	0.081
6-Factor Model	9180.202	9808.526	0.861	0.811	0.078
7-Factor Model	9148.172	9738.938	0.937	0.917	0.046

Exploring the dimensions of the questionnaire items, Table 6 presents parameter estimates derived from the 7D 3PL and 2PL models using the QMCEM algorithm in MIRT. The parameter "a" represents the item's discrimination parameter, indicating the extent to which an item distinguishes students on the latent trait, serving as a primary reference for judging item dimensionality. Each item loads on one or more dimensions, with larger absolute values indicating a higher probability of the item assessing that ability dimension. When loading values are below 0.25, the item is considered not to assess that dimension significantly. For instance, item 21 loads on C2 and A1, suggesting that it primarily assesses operational definition and algorithm understanding dimensions. The discrimination matrix indicates that the exploratory analysis aligns with the predefined item dimensions, affirming good structural validity for the questionnaire.

Table 6: Project parameters based on 2PL and 3PL models

Item	a							d	c	MDIFF	MDISC	RMSEA
	C1	C2	A1	A2	P1	P2	P3					
T19		0.87						0.82	0.24	-0.94	0.87	0.046
T20							0.52	0.16	0.26	-0.31	0.52	0.051
T21		0.98	0.33					-0.75	0.24	0.72	1.03	0.046
T22			0.83				0.35	1.01	0.22	-1.11	0.91	0.032
T23		0.58	0.56			0.44	1.34	-1.04	0.29	0.64	1.62	0.058
T24			0.91					0.56	0.15	-0.61	0.91	0.049
T25	0.75	0.83	1.03					3.68	0.22	-2.42	1.52	0.070
T26				1.02				1.07	0.24	-1.04	1.03	0.046
T27				0.81				1.06	0.22	-1.30	0.81	0.068
T28					0.98			0.72	0.26	-0.73	0.99	0.045
T29				0.73				0.41	0.26	-0.56	0.73	0.062

T30				1.53	0.41			-1.22	0.22	0.77	1.59	0.059
T31				0.96	1.78			-1.21	0.26	0.60	2.02	0.043
T32				0.34	0.39			0.30	0.24	-0.58	0.52	0.050
T33		0.87	1.46					0.59	0.24	-0.35	1.70	0.061
T34				1.42	1.30			-0.37	0.24	0.19	1.93	0.035
T35			0.28			0.36		-1.21	0.25	2.65	0.46	0.049
T36		1.29		0.74				-1.21	0.23	0.81	1.49	0.045
T37	0.49	0.82	1.32					-0.47	-	0.59	0.80	0.027
T38			1.25			1.79	1.53	-0.46	-	0.17	2.67	0.023
T39				0.74			1.03	-3.09	-	2.45	1.26	0.045
T40				0.82			1.15	-3.62	-	2.57	1.41	0.036

4.2. Quality of CAPV-t (answer to Q2)

1) Project - model fitting analysis

The mirt package's itemfit function was employed to analyze the fit of the item model. The fit indices utilized the Root Mean Square Error of Approximation (RMSEA), ranging from 0 to 1. A lower RMSEA value indicates better fit between items and the model. According to general standards, an RMSEA below 0.1 suggests good fit, below 0.05 indicates very good fit, and below 0.01 implies excellent fit. The RMSEA values for the 22 items are provided in Table 6, ranging from 0.23 to 0.7, indicating satisfactory fit for the items.

2) Parameter quality analysis of test paper

Table 6 presents the parameter estimation results for the 22 items. Here, "a" represents the item discrimination, "d" denotes the item difficulty, and "c" signifies the item guessing. In the multidimensional item response model, we utilized the multidimensional difficulty coefficient (MDIFF) instead of "d" to indicate item difficulty. Larger MDIFF values imply greater difficulty, falling within the standard range of [-3, 3]. The difficulty of the items ranged from -2.42 to 2.65, meeting the difficulty standards for item design. There is a gradual increase in item difficulty with the item number, aligning with the vertical organization of the test and reflecting students' thinking habits during answering.

The multidimensional discrimination index (MDISC) reflects the overall discrimination of items. Larger MDISC values indicate higher discrimination in multidimensional space, although this doesn't guarantee high discrimination on every dimension. MDISC values above 1.5 are considered excellent discrimination, 1.0-1.5 are good, 0.5-1.0 are moderate, and below 0.5 are considered poor and should be considered for removal. Applying this criterion, seven items exhibit excellent overall discrimination, constituting approximately 32% of the entire test. Five items are categorized as good, nine as moderate, and only item T35 has an overall discrimination below 0.5, suggesting its removal.

3) Student ability estimation based on CAPV-t

The distribution of scores on the values and skills levels for the participants is illustrated in Figure 1. Overall, scores in the skills domain are concentrated between 0.5 and 0.65, indicating the highest total scores in the decomposition dimension and the lowest total scores in the transfer dimension. In contrast, scores in the values domain range from 1.4 to 2.2, with the highest total scores observed in the questioning dimension and the lowest total scores in the creativity dimension.

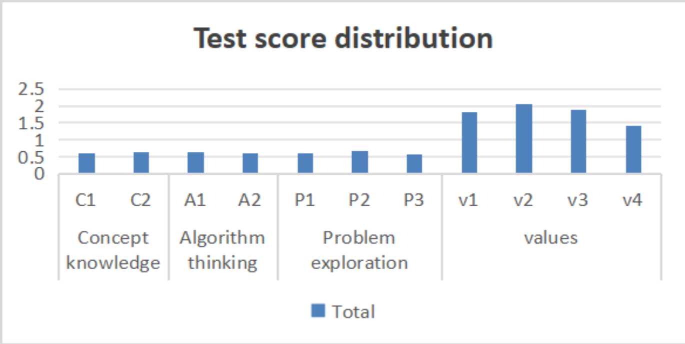

Figure 1. CT score distribution of subjects in each dimension

5. DISCUSSION

5.1. CAPV-t based on cognitive and non-cognitive abilities is innovative and complete

1) The innovation and integrity of the CT framework

Clarifying the dimensions of CT is crucial for teaching and assessment. Current research summaries indicate a traditional inclination to categorize CT abilities as cognitive skills, such as abstraction, algorithmic thinking, and decomposition. This emphasis underscores students' capabilities in information processing and extraction, resembling an assessment of students' "IQ" levels. Conversely, non-cognitive abilities are often overlooked in research but are equally significant in educational studies. Recognized as a core competency of the 21st century, non-cognitive abilities are considered implicit human capital. Both academia and society are gradually acknowledging the necessity for schools to impart skills beyond textbooks, particularly those related to students' attitudes, beliefs, and values.

The essence of CT lies not only in identifying and abstracting problem scenarios but also in the analysis and handling of intricate relationships between problem scenarios and data information. This viewpoint breaks away from traditional CT training methods, representing a direction for educational reform and development. Therefore, this study integrates literature research, defining conceptual knowledge, algorithmic thinking, and problem exploration as cognitive components of CT. Simultaneously, it identifies values as non-cognitive components, forming a theoretical framework for CT. Under the standards of core dimensions, secondary dimensions are finely delineated, providing clear definitions to explicate core dimensions. The CT framework, comprising both cognitive and non-cognitive abilities, transcends traditional definitions, enriching the connotation of CT.

2) Innovation and integrity of CAPV-t

CAPV-t is a testing tool developed based on the aforementioned theoretical framework, aiming to comprehensively assess students' cognitive and non-cognitive abilities. Cognitive abilities are often manifested through horizontal tests, so we assess students' performance in the cognitive dimensions of conceptual knowledge, algorithmic thinking, and problem exploration by compiling questionnaires. In contrast, non-cognitive abilities cannot be judged numerically. Therefore, we choose to employ a scale to describe students' performance in attitude, skepticism, teamwork, and creativity, aiming to evaluate their CT values. This study innovatively designs the CT assessment tool to include both a scale and a questionnaire, breaking away from the conventional restriction of having only one type of testing method. This approach allows for a diverse and comprehensive evaluation of students' CT abilities.

5.2. CAPV-t has strong psychometric characteristics

1) Reliability and validity analysis of CAPV-t

The study conducted reliability and validity analyses of CAPV-t using SPSS software and the mirt toolkit. The experimental results indicate that both the scale and the questionnaire achieved satisfactory levels of reliability. Factor analysis of the scale revealed consistency in the number of dimensions with the preset dimensions, and the component loadings on the factors matched the expectations, ensuring good structural validity. The validity analysis of the questionnaire suggested that fitting the data with a 7-dimensional model was most appropriate, with superior fit indices compared to other models. The discriminant matrix obtained through QMCEM method demonstrated alignment between the dimensions of the questionnaire and the preset Q matrix.

2) CAPV-t quality analysis

Based on the analysis of 366 sets of data, the quality of CAPV-t generally meets the standards of psychometrics. The research results indicate that the difficulty of CAPV-t ranges from -2.42 to 2.65, with the majority of items falling into the moderate difficulty level. Items that are either too difficult or too easy constitute only 18% of the test, aligning with the requirements of educational measurement. The evaluation of items using the MDISC index reveals that 54% of the items are rated as good, with only item 35 showing relatively poor discriminability. Participant feedback attributes this to unclear wording in the question. Given that the test does not assess students' reverse thinking, it has been decided to remove this particular item. In summary, the CAPV-t demonstrates good item discriminability and moderate difficulty, making it suitable for assessing college students' CT abilities.

6. CONCLUSIONS

This study has established a comprehensive and reasonable theoretical framework for CT, encompassing both cognitive and non-cognitive abilities of students. Based on this theoretical framework, a well-designed assessment tool has been created, demonstrating high-quality item characteristics and meeting the standards of psychometrics. This assessment tool enables effective inference of participants' abilities, assisting educators in evaluating students' proficiency or deficiencies in CT. It also serves as a foundation for future optimizations of the assessment tool and informs instructional practices. In summary, the developed CAPV-t in this study is a novel and validated assessment tool with high quality, specifically designed to evaluate cognitive and non-cognitive aspects of CT in college students.

Acknowledgment: The authors would like to acknowledge the General Program of the National Natural Science Foundation of China for its financial support for this project. Grant Sponsor:the General Program of the National Natural Science Foundation of China.Grant Number:61977031.

References

[1] Wing, J. M. (2006). Computational thinking. Communications of the ACM, 49(3), 33–35. https://doi.org/10.1145/1118178.1118215

[2] Wing, J. M. (2008). Computational thinking and thinking about computing. Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, 366(1881), 3717–3725. https://doi.org/10.1098/rsta.2008.0118

[3] Tsarava, K., Moeller, K., Román-González, M., Golle, J., Leifheit, L., Butz, M. V., & Ninaus, M. (2022). A cognitive definition of computational thinking in primary education. Computers & Education, 179, 104425.

[4] Israel-Fishelson, R., & Hershkovitz, A. (2022). Studying interrelations of computational thinking and creativity: A scoping review (2011–2020). Computers & Education, 176, 104353. https://doi.org/10.1016/j.compedu.2021.104353

[5] Guggemos, J., Seufert, S., & Román-González, M. (2023). Computational thinking assessment–towards more vivid interpretations. Technology, Knowledge and learning, 28(2), 539-568.

[6] Yağcı, M. (2019). A valid and reliable tool for examining computational thinking skills. Education and Information Technologies, 24(1), 929-951.

[7] Cutumisu, M., Adams, C., & Lu, C. (2019). A sco** review of empirical research on recent computational thinking assessments. Journal of Science Education and Technology, 28(6), 651-676.

[8] Lyon, J. A., & J. Magana, A. (2020). Computational thinking in higher education: A review of the literature. Computer Applications in Engineering Education, 28(5), 1174–1189. https://doi.org/10.1002/cae.22295

[9] Xue, Z. (2023). Practice of Higher Education Reform Based on the Demand for Innovative Talents. The Educational Review, USA, 7(9), 1259-1264.

[10] Yang, T., Cheon, J., & Solis, A. (2023, March). Exploring Undergraduate Students' Computational Thinking Comprehension: A Content Analysis Study. In Society for Information Technology & Teacher Education International Conference (pp. 118-126). Association for the Advancement of Computing in Education (AACE).

[11] Saltiel, F. (2020). Gritting it out: The importance of non-cognitive skills in academic mismatch. Economics of Education Review, 78, 102033. https://doi.org/10.1016/j.econedurev.2020.102033

[12] Xue, P., Li, B., Wang, N., & Zhu, T. (2019). Emotion Recognition from Human Gait Features Based on DCT Transform. Human Centered Computing, 511–517. https://doi.org/10.1007/978-3-030-37429-7_51

[13] Tang, X., Yin, Y., Lin, Q., Hadad, R., & Zhai, X. (2020). Assessing computational thinking: A systematic review of empirical studies. Computers & Education, 148, 103798. https://doi.org/10.1016/j.compedu.2019.103798

[14] Román-González, M., Pérez-González, J.-C., Moreno-León, J., & Robles, G. (2018). Extending the nomological network of computational thinking with non-cognitive factors. Computers in Human Behavior, 80, 441–459. https://doi.org/10.1016/j.chb.2017.09.030

[15] Papert, S. A. (1993). Mindstorms. Basic Books. https://book.douban.com/subject/2270614/

[16] Brennan, K., & Resnick, M. (2012). New frameworks for studying and assessing the development of computational thinking.

[17] Korkmaz, Ö., & Bai, X. (2019). Adapting Computational Thinking Scale (CTS) for Chinese High School Students and Their Thinking Scale Skills Level. Participatory Educational Research, 6(1), Article 1. https://doi.org/10.17275/per.19.2.6.1

[18] Selby, C., & Woollard, J. (2013). Computational thinking: The developing definition [Monograph]. University of Southampton (E-prints). https://eprints.soton.ac.uk/356481/

[19] Yadav, A., Ocak, C., & Oliver, A. (2022). Computational Thinking and Metacognition. TechTrends, 66(3), 405–411. https://doi.org/10.1007/s11528-022-00695-z

[20] Shell, D. F., Soh, L.-K., Flanigan, A. E., Peteranetz, M. S., & Ingraham, E. (2017). Improving Students' Learning and Achievement in CS Classrooms through Computational Creativity Exercises that Integrate Computational and Creative Thinking. Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education, 543–548. https://doi.org/10.1145/3017680.3017718

[21] Pollock, L., Mouza, C., Guidry, K. R., & Pusecker, K. (2019). Infusing Computational Thinking Across Disciplines: Reflections & Lessons Learned. Proceedings of the 50th ACM Technical Symposium on Computer Science Education, 435–441. https://doi.org/10.1145/3287324.3287469

[22] Leela, S., Chookeaw, S., & Nilsook, P. (2020). An Effective Microlearning Approach Using Living Book to Promote Vocational Students' Computational Thinking. Proceedings of the 3rd International Conference on Digital Technology in Education, 25–29. https://doi.org/10.1145/3369199.3369200

[23] Sun, L., Hu, L., Zhou, D., & Yang, W. (2022). Evaluation and developmental suggestions on undergraduates' computational thinking: A theoretical framework guided by Marzano's new taxonomy. Interactive Learning Environments, 1–23. https://doi.org/10.1080/10494820.2022.2042311

[24] Seiter, L., & Foreman, B. (2013). Modeling the learning progressions of computational thinking of primary grade students. Proceedings of the ninth annual international ACM conference on International computing education research, 59–66. https://doi.org/10.1145/2493394.2493403

[25] Moreno-León, J., Robles, G., & Román-González, M. (2015). Dr. Scratch: Automatic Analysis of Scratch Projects to Assess and Foster Computational Thinking. RED-Revista de Educación a Distancia.

[26] Román-González, M., Pérez-González, J.-C., & Jiménez-Fernández, C. (2017). Which cognitive abilities underlie computational thinking? Criterion validity of the Computational Thinking Test. Computers in Human Behavior, 72, 678–691. https://doi.org/10.1016/j.chb.2016.08.047

[27] Dagienė, V., & Futschek, G. (2008). Bebras International Contest on Informatics and Computer Literacy: Criteria for Good Tasks. In R. T. Mittermeir & M. M. Sysło (Eds.), Informatics Education—Supporting Computational Thinking (Vol. 5090, pp. 19–30). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-69924-8_2

[28] Korkmaz, Ö., Çakir, R., & Özden, M. Y. (2017). A validity and reliability study of the computational thinking scales (CTS). Computers in Human Behavior, 72, 558–569. https://doi.org/10.1016/j.chb.2017.01.005

[29] Williams, B., Onsman, A., & Brown, T. (2010). Exploratory Factor Analysis: A Five-Step Guide for Novices. Australasian Journal of Paramedicine, 8, 1–13. https://doi.org/10.33151/ajp.8.3.93

[30] Shrestha, N. (2021). Factor Analysis as a Tool for Survey Analysis. American Journal of Applied Mathematics and Statistics, 9(1), 4–11. https://doi.org/10.12691/ajams-9-1-2

[31] Chalmers, R. P. (2012). mirt: A Multidimensional Item Response Theory Package for the R Environment. Journal of Statistical Software, 48, 1–29. https://doi.org/10.18637/jss.v048.i06

[32] Jank, W. (2005). Quasi-Monte Carlo sampling to improve the efficiency of Monte Carlo EM. Computational Statistics & Data Analysis, 48(4), 685–701. https://doi.org/10.1016/j.csda.2004.03.019
[33] McDonald, R. P. (2000). A Basis for Multidimensional Item Response Theory. Applied Psychological Measurement, 24(2), 99–114. https://doi.org/10.1177/01466210022031552
[34] Reckase, M. D. (2009). Multidimensional Item Response Theory Models. Multidimensional Item Response Theory, 79–112. https://doi.org/10.1007/978-0-387-89976-3_4