



Developing College students' computational thinking multidimensional test based on Life Story situations

Chunhua Kang¹ · Na Liu¹ · Yinrui Zhu¹ · Feiming Li¹ · Pingfei Zeng¹

Received: 26 August 2021 / Accepted: 20 June 2022 / Published online: 30 August 2022
© The Author(s) 2022

Abstract

At present, research on computational thinking in universities is gaining interest, and more attention is being paid to the cultivation and teaching of computational thinking. However, there is a lack of computational thinking assessment tools for college students, which makes it difficult to understand the current status and development of their computational thinking. In this study, computational thinking is regarded as the ability to solve practical problems. By analyzing the relevant literature, we identified five dimensions of computational thinking – decomposition, generalization, abstraction, algorithm and evaluation – and described their operational definitions. Referring to the Bebras and the problem situations in Google computational thinking education, we set up a life-based situation that college students are familiar with. Based on the life story situation, we developed a multidimensional assessment for college students' computational thinking. This assessment tool contains 14 items, all of which are multiple-choice questions, and the structure and quality of the tool are verified by multidimensional item response theory. The results show that the assessment tool has good internal validity and can discriminate different disciplines of college students. The college students' computational thinking test developed in this study can be used as an effective tool to assess college students' computational thinking.

Keywords College students · Computational thinking assessment · Tool development · Multidimensional item response theory · Item quality analysis

✉ Pingfei Zeng
zpf@zjnu.cn

¹ Key Laboratory of Intelligent Education Technology and Application of Zhejiang Province, Zhejiang Normal University, 321004 Jinhua, Zhejiang, China

1 Introduction

The definition of “computational thinking” has gradually become popular in education since 2006. Alongside logical thinking and empirical thinking, it has been listed as one of the three basic scientific thinking for humans to understand and change the world (Li, 2012). Its application is not limited to computers. With the advent of big data, computational thinking has attracted much attention as a comprehensive and efficient way of thinking for solving problems, designing systems and understanding human behaviour (Luo et al., 2019). As institutions that send young talents into society, universities should consider training and improving the problem-solving ability of contemporary college students an important goal. The Computer Science and Telecommunications Board (CSTB) of the National Research Council believes that computational thinking is the core ability of students in the 21st century. It is as important as basic skills such as reading, writing and arithmetic, and it is crucial to students’ core literacy, which could determine the innovative competitiveness of countries in the future (National Research Council, 2010; Qualls & Sherrell, 2010).

With China’s increasing attention to the cultivation and development of college students’ computational thinking, assessments of this kind of thinking have become a growing problem restricting the development of computational thinking education (Chen & Ma, 2020). Lyon and Magana overviewed the study of computational thinking in international universities and found that computational thinking is vaguely defined in universities and lacks assessment tools (Lyon & Magana, 2020). Thus, there is an urgent need to understand how to develop computational thinking assessment tools for college students based on the concept and connotation of computational thinking.

1.1 The definition of computational thinking

Academics do not agree on the definition and framework of computational thinking, but they believe that computational thinking is related to problem-solving. Grover & Pea (2018) defined computational thinking as a thinking process by which a problem is formulated that can be solved effectively by a human or machine. Cutumisu et al., (2019) defined computational thinking as a problem-solving process involving a series of cognitive or metacognitive activities and the use of strategies or algorithms to solve problems.

Generally, definitions of computational thinking include two aspects. The first falls under the context of the discipline, in which computational thinking is the knowledge of computer science (Brennan & Resnick, 2012). The British Computing Course Working Group defines computational thinking as the process of applying computational tools and techniques to understand artificial information systems and natural information systems (Csizmadia et al., 2015). Google Exploring Computational Thinking (ECT) thinks that computational thinking is a mental process and tangible result related to solving computational problems, including abstraction, algorithm design, automation, data analysis, data collection, data representation, decomposition, parallelization, pattern generalization, pattern recognition and simulation (ECT, 2017). Overemphasizing the concept of computer or programming may cause the

public to misunderstand computational thinking, while an individual's computer knowledge or programming ability does not represent their computational thinking competency.

The second aspect of computational thinking is that it can be separated from specific disciplines. It is regarded as a problem-solving skill in daily life, associated with problem-solving related skills and specific personalities, such as the perseverance and confidence that individuals exhibit when solving specific problems (Barr & Stephenson, 2011; Csizmadia et al., 2015; González, 2015; Lu & Fletcher, 2009; Selby & Woollard, 2013).

Computational thinking is regarded as a general method to solve problems that is not limited by one's specific discipline knowledge or programming ability. It also lays a theoretical foundation for the development of a general computational thinking scale or test.

1.2 The Assessment of Computational thinking in Higher Education

Most recent research on computational thinking has focused on K-12 education, while only a few studies involved higher education. Furthermore, little attention has been paid to computational thinking assessments (Cutumisu et al., 2019). There are several methods for assessing computational thinking in higher education.

Computational thinking assessments usually use multiple-choice questions or programming tests to examine learners' computer knowledge (if/else conditions, while conditions, loops, etc.) and programming ability, such as Scratch and Alice (González, 2015; Figl & Laue, 2015; Peteranetz et al., 2018; Romero et al., 2017; Topalli & Cagiltay, 2018; Dağ, 2019). These assessments are not suitable for students who have not been exposed to programming. Since it is restricted by teaching conditions, it is also unsuitable for general testing in some areas. In addition, some scholars have developed scales for subjectively assessing individuals' computational thinking ability, such as the computational thinking scale (Doleck et al., 2017; Korkmaz et al., 2017; Sentance & Csizmadia, 2017). This scale is easy to operate and understand, but it may only assess the ability that students think they have, not their real computational thinking ability. Therefore, although its construct validity is relatively high, its content validity is questionable.

However, an increasing number of researchers believe that computational thinking is a way of thinking that is connected to problem-solving in everyday life and a basic skill that everyone can possess (Wing, 2006). Therefore, existing assessments of computational thinking, which are based on computer or programming knowledge and designed to assess individuals while confining computational thinking to some fixed discipline, are not suitable for college students without computer knowledge or programming experience. "The Bebras International Challenge on Informatics and Computer Literacy" uses a life-based story to measure individuals' problem-solving ability, thereby removing the limitations imposed by assessments that test students' knowledge in specific disciplines. However, the problems are used only for competition, and measurable indicators (difficulty, discrimination, reliability and validity) have not been verified. Moreover, some problems are too difficult or too easy (Belletini et al., 2015) and are not suitable for testing the public. In addition, the problems

are provided by the participating countries with various languages and cultures. As a result, different students may have different understandings of the problems, leading to a mismatch between the measured skills (or the dimensions of these problems) and the previously proposed ones (Araujo et al., 2019; Izu et al., 2017).

This study describes the development of a computational thinking multidimensional assessment tool based on collecting materials related to computational thinking and teaching cases at home and abroad. The proposed tool considers problem-solving in real-life situations as the carrier, making it suitable for Chinese college students.

2 Method

2.1 The Five-Dimensional structure and operational definition of computational thinking

Weintrop & Wilensky (2015) expressed that researchers must decompose computational thinking into a set of well-defined and measurable skills, concepts or practices. Based on the above-mentioned literature review, we extracted the five core elements of computational thinking, namely abstraction (Abs), decomposition (Deco), algorithm (Alg), assessment (Eva) and generalization (Gen) five abilities (or dimensions). We also theoretically constructed the structure of college students' computational thinking competency as an efficient problem-solving ability and a general thinking ability that everyone could have. These five abilities adopt the terminology of computer science and conform to the initial definition of computational thinking while also removing the limitations of the discipline. Owing to the efforts of researchers, these five dimensions are no longer limited to computers. It is believed that these elements are critical for understanding, analyzing and solving problems. It is also believed that they are consistent with the general concept in the context of problem-solving. Thus, the five abilities or skills are used as the five dimensions of computational thinking assessment. Based on an overview of the five dimensions in the literature (Barr & Stephenson, 2011; IEA, 2016; Kalelioglu et al., 2016; Labusch & Eickelmann, 2017; Selby & Woollard, 2013; Wing, 2008), we describe the computational thinking assessment in more detail and turn it into an operational definition that is easy to measure, as shown in Table 1.

2.2 Method

First, materials related to computational thinking and its teaching cases at home and abroad were collected (such as *computational thinking for educators from Google* and *Bebras Unplugged in Australia*). These were then combined with problem-solving situations in life. We developed an initial test containing 20 life-story situations according to the Bebras problems and preliminarily determined the test framework and the dimensions that each question examined. After that, small-scale testing and interviews were conducted on the initial 20-question test. Items were added or deleted after qualitative (item description) and quantitative (item discrimination, pass rate)

Table 1 Five Dimensions of Computational Thinking and Operational Definitions

Dimension	Operational Definition
Abs	When dealing with complex problems or tasks, the complexity of the problem or task can be reduced by removing unnecessary details; in the process of problem-solving, it can extract, collect or create key data related to problem-solving; it can express real problems in symbolic ways.
Deco	A relatively complex problem can be divided into several parts or small problems that can be understood, solved and assessed. These small parts or problems can then be decomposed and analyzed, and the connection and logical sequence between the different parts can be established.
Gen	Summarizing includes collecting and analyzing data, considering its specific problem-solving steps, and developing a model to imitate processes. After problem-solving for a specific situation, ideas and solutions from one problem are applied to another.
Alg	The problem is transformed from symbolic expressions into a computational model (such as sequences and rules). When analyzing and solving the problem, the specific steps of problem-solving can be described, and the solution can be obtained by clearly defining each step.
Eva	The correctness of the formulated rules can be assessed. Moreover, the effectiveness of a solution and the use of appropriate methods to assess the optimal solution among multiple rules or algorithms to solve the problem can be determined.

analyses. Ultimately, 14 questions remained. Finally, we conducted a test among college students, collected data and analyzed the quality of the test.

2.2.1 Participants

To verify the applicability of the test, 450 college students studying different disciplines (e.g. computer science and technology, Chinese language and literature, mathematics, psychology) were randomly selected for the test. According to the Chinese education system, their age was 18~22 years old. After excluding invalid responses, we obtained 433 responses from 105 males, 322 females and 6 respondents who did not fill their gender. The data validity rate was 96%. The basic information provided by the college students is shown in Table 2.

2.2.2 Pre-Testing and Item Selection

In the initial test, the 20 story scenarios mainly included “Ants cross the bridge” (item 2), “False news” (item 4), “You ask, I guess” (item 5), “The stitches of the sewing

Table 2 Basic Information of Respondents

Feature	Category	Code	Number of Cases	Total
Gender	Male	1	105	433
	Female	2	322	
	Not filled	0	6	
Discipline	Math	1	47	433
	Chinese language and literature	2	91	
	Computer science and technology	3	31	
	Science education	4	142	
	Psychology	5	74	
	Geography	6	44	
	Not filled	0	4	

machine” (item 6) and “Energy conservation” (item 14). Here, we use “You ask, I Guess”, “Who will go home first”, and “Energy conservation” as examples to show the question context and dimensions:

You ask, I guess: *In a 20-question game, one player is thinking of something in his mind, and another player can guess what the other player is thinking by asking 20 right and wrong questions. Observe the picture in Fig. 1. What is the fewest number of questions you can ask the person to determine the animal they are thinking about?*
()

Fig. 1 The Question of “You Ask I Guess”

Analysis: If each question can eliminate half of the options, the first question eliminates four, and there are four remaining. The second question eliminates two, and two remain. After removing one of the four questions, only the last one is left. The number of questions to be asked can be obtained by calculating the log 2 of all possibilities or by calculating the power of 2 close to or equal to the total number of possibilities: $2^4=16$, $2^5=32$... $2^{33} \approx 850$ million. This is the so-called abstraction and algorithm.

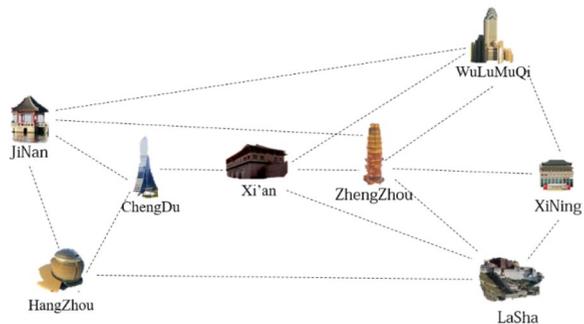
Who will go home first: The teacher played a game with the students when the class was about to end, and the winner could leave school first. The rules of the game are as follows: The school has a corridor with five doors in a row. The students lined up and took turns walking down the corridor. When they walk to an open door, students must close it and walk to the next door. When reaching a closed door, stu-

dents must open it and enter the classroom, leaving the door open until the teacher dismisses them. At the start of the game, all doors are closed. If a student finds that all the doors are open, he can go home! If the students are numbered from 1 to 35, which student leaves the school first?

Analysis: This question can be solved with binary counting, where 0 is a closed door and 1 is an open door. 00000 means all doors are closed. For the first student, the first door is closed. They open it and enter the classroom. The number is 00001. The second student shuts the first door. The second door is closed, so they open it and enter the room. The number is 00010. Only for the 32nd student will 11,111 occur. It is mainly decomposition.

Energy conservation: Domestic airlines set up several routes, as shown in Fig. 2. In response to the requirements of energy conservation and emission reductions, a company plans to cancel some routes on the premise of ensuring that passengers can reach all cities (passengers need to reach the destination city by connecting flights). How many routes can be cancelled at most? ()

Fig. 2 The Question of “Energy Conservation”



Analysis: Assuming there are two cities, at least one route is required; if there are three cities, then at least two routes are required; if there are five cities, at least seven routes are required. This problem can also be represented by a graph in which the city is the node and the route is the edge. If nine routes are assumed to be cancelled, then six routes will remain, and at least one of the eight cities will not be connected to any route. Students need to know the structure of the data and characterize it appropriately to form computable models to solve problems.

For the 20 situational items, a small-scale test and interview were conducted. After qualitative (item expression) and quantitative (item discrimination, pass rate) analyses, items with poor results were deleted. Nine items, such as preparing banquets, special signal towers and superstars, were too simple for college students, so they were deleted. For two items – dividing candy equally and pasting wallpaper – expressions were ambiguous, and the language was modified accordingly. Based on college students’ suggestions, we added three interesting questions: combination lock, farmland report and string the beads. In the end, a test containing 14 life-based story situations was formed. The dimensions of the abilities to be measured for these questions are shown in Table 3.

Table 3 14 Story Situations and Measured Dimensions

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Item 13	Item 14
Abs	√	√		√	√					√		√	√	√
Eva		√		√	√	√			√					
Alg		√	√	√	√	√	√		√		√		√	√
Deco	√			√			√	√	√					
Gen			√		√	√					√	√		

Note: "√" means that this dimension is measured

Item 1: The shortest time; Item 2: Ants cross the bridge; Item 3: Reduced ribbon; Item 4: False news; Item 5: You ask, I guess; Item 6: Stitches of a sewing machine; Item 7: Combination lock; Item 8: Who will go home first; Item 9: Divide candy equally; Item 10: Paste wallpaper; Item 11: String the beads; Item 12: Farmland report; Item 13: Tree -planting problem; Item 14: Energy conservation.

2.3 Statistical methods

First, we verified the five-dimension structure of the multidimensional test by a model comparison based on 433 data (regardless of subject) using the multidimensional item response theory (MIRT) package in R and examined its structural validity. Secondly, after the model structure was initially determined, we performed an item quality analysis (difficulty, discrimination, item characteristic surface) and deleted items that did not meet the psychometrics standards. We thus formed the final test framework and model. Finally, we randomly extracted 57% of the data from the total data (31 for computers, 97 for Chinese language and literature, 47 for mathematics and 74 for psychology). We used the final multidimensional test model and the MIRT package to estimate the multidimensional ability of college students' computational thinking, analyzed the differences between college students studying different subjects and provide the discrimination validity of the test.

3 Results

3.1 Structural validation of the Multidimensional tests

Table 3 shows the dimensions of the 14 items, each of which has one or more dimensions. When verifying the model, it is necessary to consider the impact of multiple dimensions on the item. Since each question is scored as 0 or 1, the relationship between the score, the item parameters and students' abilities is not linear. It is not suitable to use confirmatory factor analysis to verify the model. Therefore, we used MIRT to verify the fitness of the data.

Table 3 preliminarily determines the dimensions of each item, but the relationship between the dimensions has not been confirmed. Therefore, based on the relationship between the questions and the dimensions provided in Table 3, a pairwise correlation between the five dimensions was referred to as a fully correlated model (FCM). Meanwhile, a partial correlation among the five dimensions was referred to as a partial correlated model (PCM), and a lack of correlation among the five dimensions was referred to as an uncorrelated model (UCM). Figure 3 provides a comparison of the three models. Because the multidimensional (M-D) model is more complicated than the others, we only show the relationship between the dimensions in the figure; the relationship between the dimensions and the questions is shown in Table 3. The appropriate model is selected as the final model of the multidimensional test through the fit index. In addition, to verify the fitness of the multidimensional model, we added the results of the one-dimensional (O-D) model for comparison. We used the MIRT package in R for data analysis. The results are shown in Table 4.

Fig. 3 Fully Correlated, Partially Correlated and Uncorrelated Models

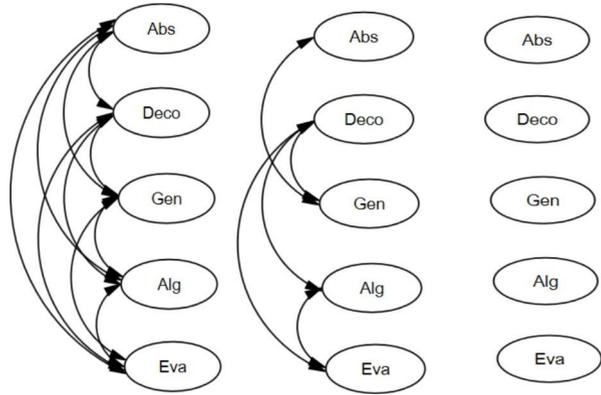


Table 4 Model Fit Index

Model		Fit Index					Information Index	
		M2	M2/DF	RMSEA	SRMSR	CFI	AIC	BIC
M-D model	FCM	111.136	1.793	0.043	0.069	0.890	7617.304	7833.053
	PCM	107.275	1.730	0.041	0.066	0.900	7609.711	7805.106
	UCM	107.876	1.740	0.041	0.062	0.890	7632.425	7807.467
O-D model		156.039	2.026	0.049	0.057	0.816	7644.834	7758.815

Table 5 Correlations Among MIRT Multidimensional Capabilities

Dimension	Abs	Deco	Gen	Alg	Eva
Abs	1	-	0.378	-	-
Deco		1	0.15	0.472	0.279
Gen			1	-	-
Alg				1	0.45
Eva					1

Table 4 shows that the results of the multidimensional model are better than those of the one-dimensional model. Specifically, the indicators of the three kinds of models are better in the multidimensional model. A comprehensive comparison revealed that the M2/DF, RMSEA, CFI, AIC and BIC of the partially correlated model are better than those of the other two models. Therefore, a partially correlated model was used for subsequent analyses. Table 5 shows the correlation between five dimensions. The correlation coefficient is between [0.15, 0.47], indicating low-to-medium correlations between these dimensions.

3.2 Item quality analysis

3.2.1 Item index

The mathematical model of MIRT includes multidimensional ability parameters that describe individual characteristics, as well as the difficulty and discrimination of each item in multiple ability dimensions. It also provides the synthesized item discrimina-

Table 6 Item Parameters Based on Partial Related Models

Item	d	α					b	MDISC
		Abs	Eva	Alg	Deco	Gen		
T01	2.517	0	0	0	1.211	1.38	-1.602	1.836
T02	0.463	0.174	0	0	1.272	0	-0.366	1.284
T03	-0.078	0	0	1.431	0	0	0.058	1.431
T04	1.636	0	0.468	0	0.895	0.212	-1.539	1.032
T05	-0.598	0	0	0.234	0.260	0	1.683	0.350
T06	1.064	0	0.332	0.927	0	0.768	-0.902	1.249
T07	0.1	0	0	0	0.725	0.127	-0.144	0.736
T08	-1.308	0	1.386	0	0.868	0	0.835	1.635
T09	0.13	0.186	0	0	0.358	0	-0.303	0.404
T10	0.377	0.718	0.850	0	0	0	-0.304	1.113
T11	-0.052	0	0	0.511	0	0.980	0.0391	1.105
T12	0.426	2.128	0	0.381	0	0	-0.194	2.162
T13	0.567	0.927	0	0.686	0	0	-0.460	1.153
T14	-0.013	0.682	0	0.218	0	0	0.029	0.716

tion known as multidimensional discrimination (MDISC). The results of the item parameters used in this study based on a partially correlated model are shown in Table 6.

In the table, d is the intercept, α is the dimensional discrimination, b is the difficulty and MDISC is the discrimination of the item. According to the classification of Ding et al., (2012), an MDISC above 1.5 is excellent, 1.0~1.5 is good, 0.5~1.0 is medium and 0.5 is poor. Among the 14 items in this test, three had excellent MDISC values, accounting for 20% (T01, T08, T12). Another seven questions (50%) had good MDISC values (T02, T03, T04, T06, T11, T10, T13). Two questions (15%) had medium values (T07, T14), and two more (15%) had poor values (T05, T09).

When calculating difficulty in MIRT, the d parameter can be converted into the difficulty parameter b according to the MDISC index. The larger the value of b , the more difficult the problem. T01 and T04 are relatively simple, T05 and T08 are relatively difficult and the others have medium difficulty.

3.2.2 Item characteristic surface

According to the discrimination and difficulty parameters in Table 5, three questions with item quality of “excellent”, “good” and “poor” were selected to present its item characteristic surface (Fig. 4). As shown in Fig. 4, T12 has a steeper surface in dimensions θ_1 and θ_3 , meaning this question has high discrimination in the abstract and algorithm dimensions. T10 has a steep surface in dimension θ_2 , while its surface in dimension θ_1 is relatively flat. It can be seen that the question has high discrimination in the assessment dimension, while the discrimination in the abstract dimension is normal. T09 is relatively flat in both dimensions, indicating that this question has mediocre discrimination on abstraction and decomposition dimensions. Fig. 5 is a graph of an equiprobability curve corresponding to the item characteristic surfaces of T12 and T09. The numbers marked on each line represent the same amount of item

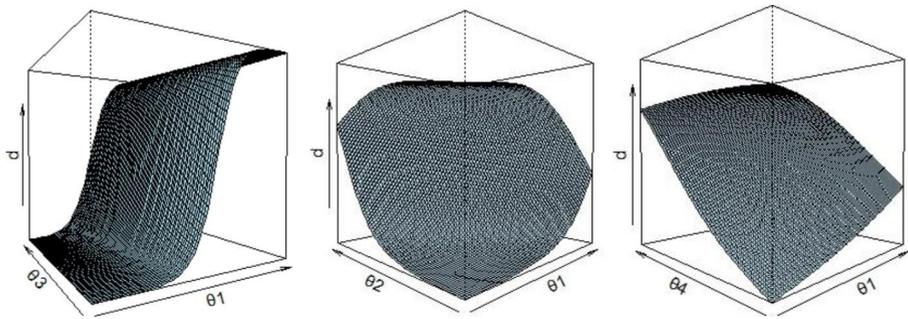


Fig. 4 Item Characteristic Surfaces of Three Questions. (T12 “Excellent” Discrimination T10 “Good” Discrimination T09 “Poor” Discrimination.)

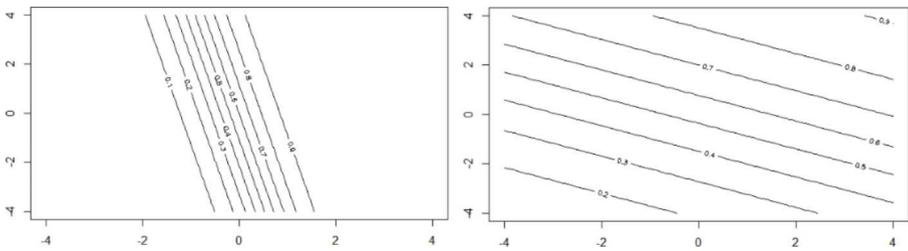


Fig. 5 Item Equiprobability Curve. (T12- equiprobability curve, T09- equiprobability curve.)

information for all participants on the line. The equiprobability curve reflects the discrimination: the denser the curve, the greater the discrimination.

3.2.3 The final multidimensional test

Based on the above MIRT analysis, the α and MDISC index of items 5 and 9 are both poor, so they were deleted from the test. The multidimensional test structure has been adjusted after being verified. For example, the ninth question, “Who will go home first”, originally examined the decomposition, but after MIRT analysis, the evaluation was also examined. A subsequent discussion of the item revealed that in addition to solving the question step by step through decomposition, there was also a need to evaluate the steps in this process to verify the rationality of the method. Thus, the item measured two dimensions. The adjusted dimensional structure is shown in Table 7. This table shows the number of questions measured in each dimension, which forms the final college student computational thinking test.

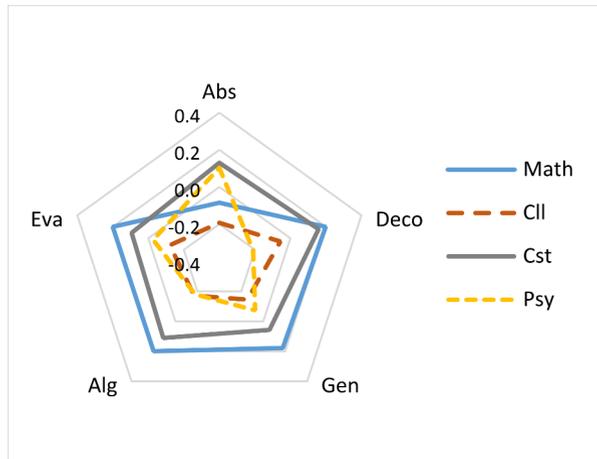
3.3 Differences in multidimensional abilities of College students in different disciplines

Based on the final test model in Table 4, the ability of college students in each ability dimension of computational thinking can be estimated, and the discipline differences between them can be analyzed. Figure 6 shows the average of the five dimensions of

Table 7 Multidimensional Structure of the Questions After Verification

	Item 1	Item 2	Item 3	Item 4	Item 5	Item 6	Item 7	Item 8	Item 9	Item 10	Item 11	Item 12	Number of Questions
Abs	√							√		√	√	√	5
Eva				√	√		√	√					4
Alg			√		√				√	√	√	√	6
Deco	√	√		√	√	√	√						5
Gen	√			√	√	√		√	√				5

Fig. 6 Radar Chart of Discipline Ability



college students in different disciplines. Students of computer science and technology (Cst) and mathematics (Math) performed the best for these dimensions and had the most balanced scores. Chinese language and literature (CII) students had negative scores for almost all dimensions, indicating that these students' performance was poor. Although Psychology (Psy) students performed as well as computer science and technology students in the abstract dimension, they did not perform as well in the other four dimensions.

4 Discussion

4.1 The multidimensional test of College Students' computational thinking has a high quality

The computational thinking assessment tool for college students proposed in this study is scientific, and the verified structure can be used to assess the computational thinking abilities of college students in China. This study uses abstraction, decomposition, generalization, algorithm and evaluation as the five ability dimensions of computational thinking and further refines and defines the operational definition of each dimension (Barr & Stephenson, 2011; IEA, 2016; Kalelioglu et al., 2016; Labusch & Eickelmann, 2017; Selby & Woollard, 2013; Wing, 2008). Subsequently, according to the standard test development procedures, we collected item materials, conducted pre-test and item analyses and finally formed a multidimensional test of college students' computational thinking with 12 item scenarios.

The MIRT analysis results clearly and intuitively show the dimensions of each question and the measurement indicators of each dimension, which were considered to detect the detailed structure of item attributes or skills (Ackerman et al., 2003; Embretson, 2007). This provides important information for preparing the test and selecting different cognitive components of questions (Tu, Cai, Dai & Ding, (Tu et al. 2011)). The results of this study show that the five ability dimensions are structured reasonably, and some dimensions have low-to-medium-degree correlations. The fifth

and ninth questions performed poorly according to the MDISC and α index. Table 5 shows that the difficulty of the fifth question is relatively high, which led to its low discrimination in the two dimensions and its low MDISC value. Students perceived the ninth question as ambiguous due to the item condition setting. Therefore, its MDISC value was relatively low. We deleted the fifth and ninth questions to ensure the quality of the tool in future applications. The rest of the questions performed well regarding their difficulty, discrimination of various ability dimensions and MDISC. They also conformed to measurement standards. The quality of the multidimensional assessment tool for college students' computational thinking is high.

4.2 The computational thinking Multidimensional Test has good discriminative validity

Aho (2012) pointed out that problem-solving based on computational thinking is closely related to computational models. An important part of computational thinking is identifying appropriate computational process models to define problems and find solutions. Hu (2011) described the relationship between computational thinking and mathematical thinking and believed that there is some similarity between the two. Computational thinking absorbs part of the theories and methods of mathematical thinking to solve problems, such as the conceptualization, abstraction and modularity of things, as the two have a great positive correlation (Yu et al., 2018). Sırakaya, Alsancak Sırakaya and Korkmaz (2020) found that disciplinary thinking influences computational thinking. The main components of the thinking of liberal arts students are imaginal thinking, perceptual thinking and strong imagination; however, they exhibit weak abstract thinking, rational thinking and operational ability (Luo & Liu, 2017). Abstract and algorithmic thinking are the core dimensions of computational thinking.

Theoretically speaking, students of mathematics and computer science should have better computational thinking skills than students of psychology and Chinese language and literature. The results of this research verify this inference. The results of discipline analysis show that students in computer science and mathematics performed the best in four dimensions, while students in Chinese language and literature had the weakest ability in all five dimensions. Psychology students' abilities were between those of computer science and mathematics students and Chinese language and literature students. This outcome indicates that the multidimensional test of computational thinking for college students has good discrimination and can identify college students in different disciplines.

4.3 The practical application of the multidimensional test of computational thinking

Computational thinking is not possessed only by computer science majors but by all college students (Ministry of Education College Computer Course Teaching Steering Committee, 2013). In 2010, the first "Nine Schools Alliance Computer Basic Courses Seminar" issued a statement pointing out that "the cultivation of computational thinking" is the core task of computer teaching and improved the curriculum system by

focusing on cultivating computational thinking. Based on the existing literature, this study linked computational thinking with solving practical problems and developed a multidimensional computational thinking test for college students. This test can comprehensively assess the multidimensional competency of college students' computational thinking, the structure of multidimensional ability and the strengths and weaknesses of each student. It can also detect the features of computational thinking based on discipline, gender and other variables, such as the combination of strengths and weaknesses in the multidimensional abilities of college students of different disciplines and genders. Furthermore, it can assess teaching and training practices and can be used for theoretical research. In short, the development of the multidimensional test of computational thinking for college students promotes assessments and diagnoses of computational thinking ability and the theory and practical research of computational thinking. It could also contribute to the cultivation and research on the problem-solving abilities of college students.

5 Conclusions

In this study, a multidimensional test of computational thinking for college students was developed based on the context of daily life situations. The structure validity and item quality of the tool were verified by MIRT. Several conclusions can be drawn from the results. First, the college students' computational thinking test was constructed reasonably, and partial correlations between dimensions are the most appropriate. Second, among the test questions developed based on 14 life situations, there are two excellent questions, seven good questions and two medium questions. There were also two poor questions, which were deleted. Third, college students exhibited obvious differences in the five dimensions of computational thinking depending on their discipline. Students of mathematics and computer science performed significantly better than students in psychology and Chinese language and literature, indicating that the college students' computational thinking test has good discrimination. Thus, the test developed in this study can be used as an effective assessment tool for college students.

Computational thinking has become an essential skill in the 21st century. However, considering the multitude of computational thinking definitions and models, assessing computational thinking has become a major problem (Lu et al., 2022). This problem has led to notable differences in research results, as different definitions have been used in research on the cultivation and assessment of computational thinking. Therefore, more theoretical research on computational thinking is needed. While computational thinking is beginning to gain popularity in multiple disciplines, more effort is still needed to differentiate computational thinking from computer science and explore how computational thinking can be involved in daily life in a wide range of domains (Tikva & Tambouris, 2021). In the future, the influencing factors of the computational thinking ability of college students (e.g. cognitive development, teaching environment, family environment, existing experiences) can be considered to improve their computational thinking.

Acknowledgments:

Acknowledgements Supported by the Open Research Foundation of Zhejiang Key Laboratory of Intelligent Education Technology and Application (JYKF20050).

Funding: Supported by the Open Research Foundation of Zhejiang Key Laboratory of Intelligent Education Technology and Application (JYKF20050).

Conflict of interest The authors received funding from the Open Research Foundation of Zhejiang Key Laboratory of Intelligent Education Technology and Application (JYKF20050). Data Availability not applicable.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Ackerman, T. A., Gierl, M. J., & Walker, C. M. (2003). Using multidimensional item response theory to assess educational and psychological tests. *Educational Measurement: Issues and Practice*, 22(3), 37–51. doi: <https://doi.org/10.1111/j.1745-3992.2003.tb00136.x>
- Aho, A. V. (2012). Computation and computational thinking. *The Computer Journal*, 55(7), 832–835. doi: <https://doi.org/10.1093/comjnl/bxs074>
- Araujo, A. L. S. O., Andrade, W. L., Guerrero, D. D. S., & Melo, M. R. A. (2019). *How many abilities can we measure in computational thinking? A study on Bebras challenge*. Paper presented at the proceedings of the 50th ACM technical symposium on computer science education
- Barr, V., & Stephenson, C. (2011). Bringing computational thinking to K-12: What is involved and what is the role of the computer science education community? *ACM Inroads*, 2. doi: <https://doi.org/10.1145/1929887.1929905>
- Belletini, C., Lonati, V., Malchiodi, D., Monga, M., Morpurgo, A., & Torelli, M. (2015). *How challenging are Bebras tasks? An IRT analysis based on the performance of Italian students*. Paper presented at the Proceedings of the 2015 ACM conference on innovation and technology in computer science education
- Brennan, K., & Resnick, M. (2012). *New frameworks for studying and assessing the development of computational thinking*. Paper presented at the Proceedings of the 2012 annual meeting of the American Educational Research Association, Vancouver, Canada
- Chen, X., & Ma, Y. (2020). The construction and exploration of the assessment index system of localized computational thinking—Based on the sample analysis and verification of 1,410 high school students. *Journal of Distance Education*, (5)
- National Research Council. (2010). *Report of a workshop on the scope and nature of computational thinking*. National Academies Press.
- Csizmadia, A., Curzon, P., Dorling, M., Humphreys, S., Ng, T., Selby, C., & Woollard, J. (2015). *Computational thinking - a guide for teachers*. Retrieved from <https://eprints.soton.ac.uk/424545/>
- Cutumisu, M., Adams, C., & Lu, C. (2019). A scoping review of empirical research on recent computational thinking assessments. *Journal of Science Education and Technology*, 28(6), 651–676
- Ding, S., Luo, F., & Tu, D. (2012). *Special topic research on new progress in project response theory*. Beijing: Beijing Normal University Press
- Dağ, F. (2019). Prepare pre-service teachers to teach computer programming skills at K-12 level: Experiences in a course. *Journal of Computers in Education*, 6(2), 277–313. <https://doi.org/10.1007/s40692-019-00137-5>

- Doleck, T., Bazelais, P., Lemay, D. J., Saxena, A., & Basnet, R. B. (2017). Algorithmic thinking, cooperativity, creativity, critical thinking, and problem solving: Exploring the relationship between computational thinking skills and academic performance. *Journal of Computers in Education*, 4(4), 355–369
- ECT (2017). CT Overview. Retrieved from <http://edu.google.com/resources/programs/exploring-computationalthinking/#!CT-overview>
- Embretson, S. E. (2007). Construct validity: A universal validity system or just another test assessment procedure? *Educational Researcher*, 36(8), 449–455
- Figl, K., & Laue, R. (2015). Influence factors for local comprehensibility of process models. *International Journal of Human-Computer Studies*, 82, 96–110. <https://doi.org/10.1016/j.ijhcs.2015.05.007>
- González, M. R. (2015). *Computational thinking test: Design guidelines and content validation*. Paper presented at the Proceedings of EDULEARN15 conference
- Grover, S., & Pea, R. (2018). Computational thinking: A competency whose time has come. *Computer science education: Perspectives on teaching and learning in school*, 19
- Hu, C. (2011). *Computational thinking: what it might mean and what we might do about it*. Paper presented at the Proceedings of the 16th annual joint conference on innovation and technology in computer science education
- IEA. The IEA's international computer and information literacy study (ICILS) 2018. What's next for IEA's ICILS in 2018? Retrieved December 12, 2017, from http://www.iea.nl/fileadmin/user_upload/Studies/ICILS_2018/IEA_ICILS_2018_Computational_Thinking_Leaflet.pdf
- Izu, C., Mirolo, C., Settle, A., Mannila, L., & Stupuriene, G. (2017). Exploring Bebras tasks content and performance: A multinational study. *Informatics in Education*, 16(1), 39–59
- Kalelioglu, F., Gulbahar, Y., & Kukul, V. (2016). A framework for computational thinking based on a systematic research review
- 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
- Labusch, A., & Eickelmann, B. (2017). *Computational thinking as a key competence—A research concept*. Paper presented at the Conference Proceedings of International Conference on Computational Thinking Education
- Li, L. (2012). Computational Thinking—Concepts and Challenges. *China University Teaching*, 01, 9–14
- Lu, C., Macdonald, R., Odell, B., Kokhan, V., Epp, D., C., & Cutumisu, M. (2022). A scoping review of computational thinking assessments in higher education. *Journal of Computing in Higher Education*. Doi: <https://doi.org/10.1007/s12528-021-09305-y>
- Lu, J. J., & Fletcher, G. H. (2009). *Thinking about computational thinking*. Paper presented at the Proceedings of the 40th ACM technical symposium on computer science education
- Luo, H., Liu, J., & Luo, Y. (2019). The necessary mental literacy in the era of artificial intelligence: Computational thinking. *Modern Educational Technology*, (06), 26–33
- Lyon, J. A., & Magana, J. A. (2020). Computational thinking in higher education: A review of the literature. *Computer Applications in Engineering Education*, 28(5), 1174–1189
- Peteranetz, M. S., Flanigan, A. E., Shell, D. F., & Soh, L. (2018). Helping engineering students learn in introductory computer science (CS1) using computational creativity exercises (CCEs). *IEEE Transactions on Education*, 61(3), 195–203. <https://doi.org/10.1109/TE.2018.2804350>
- Qualls, J. A., & Sherrell, L. B. (2010). Why computational thinking should be integrated into the curriculum. *Journal of Computing Sciences in Colleges*, 25(5), 66–71
- Romero, M., Lepage, A., & Lille, B. (2017). Computational thinking development through creative programming in higher education. *International Journal of Educational Technology in Higher Education*, 14(1), 42. <https://doi.org/10.1186/s41239-017-0080-z>
- Selby, C., & Woollard, J. (2013). Computational thinking: The developing definition
- Sentance, S., & Csizmadia, A. (2017). Computing in the curriculum: Challenges and strategies from a teacher's perspective. *Education and Information Technologies*, 22(2), 469–495. <https://doi.org/10.1007/s10639-016-9482-0>
- Sırakaya, M., Sırakaya, A., D., & Korkmaz, Ö. (2020). The impact of STEM attitude and thinking style on computational thinking determined via structural equation modeling. *Journal of Science Education and Technology*, 29, 561–572
- Tikva, C., & Tambouris, E. (2021). Mapping computational thinking through programming in K-12 education: A conceptual model based on a systematic literature review. *Computers & Education*, 162, 104083
- Ministry of Education College Computer Course Teaching Steering Committee. (2013). Declaration of computational thinking teaching reform. *China University Teaching*, 07, 7–17

- Topalli, D., & Cagiltay, N. E. (2018). Improving programming skills in engineering education through problem-based game projects with scratch. *Computers & Education, 120*, 64–74. <https://doi.org/10.1016/j.compedu.2018.01.011>
- Tu, D., Cai, Y., Dai, H., & Ding, S. (2011). Multidimensional item response theory: Parameter estimation and application in psychological testing. *Acta Psychologica Sinica, 43*(11), 1329–1340
- Weintrop, D., & Wilensky, U. (2015). *Using commutative assessments to compare conceptual understanding in blocks-based and text-based programs*. Paper presented at the ICER
- Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33–35
- 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
- Yu, X., Xiao, M., & Wang, M. (2018). Computational thinking training in progress: Practical methods and assessment at the K-12 stage. *Journal of Distance Education, 02*, 18–28. doi:<https://doi.org/10.15881/j.cnki.cn33-1304/g4.2018.02.002>

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.