International Journal of Assessment Tools in Education
2025, Vol. 12, No. 3, 825–842
https://doi.org/10.21449/ijate.1588902
journal homepage: https://dergipark.org.tr/en/pub/ijate
Research Article

Validation and psychometric properties of the Computational Thinking Multidimensional Test

Ali Orhan[1*], **İnan Tekin**[2], **Sedat Şen**[3]

[1]School of Foreign Languages, Zonguldak Bülent Ecevit University, Zonguldak, Türkiye
[2]School of Foreign Languages, Zonguldak Bülent Ecevit University, Zonguldak, Türkiye
[3]Department of Educational Sciences, Faculty of Education, Harran University, Şanlıurfa, Türkiye

ARTICLE HISTORY

Received: Nov. 21, 2024

Accepted: June 28, 2025

Keywords:
Computational thinking assessment,
University students,
Multidimensional item response theory,
Item quality analysis,
Scale adaptation.

Abstract: In this study, it was aimed to translate and adapt the Computational Thinking Multidimensional Test (CTMT) developed by Kang *et al*. (2023) into Turkish and to investigate its psychometric qualities with Turkish university students. Following the translation procedures of the CTMT with 12 multiple-choice questions developed based on real-life situations, the data were collected from 359 university students studying in different disciplines of a state university in the northern part of Türkiye. The structure validity and item quality of the Turkish version of CTMT were verified using multidimensional item response theory and the results showed that the Turkish test has good psychometric quality. In addition, its structure and item difficulty and discrimination characteristics are similar to the original test. Therefore, it can be stated that the Turkish version of CTMT can be used as an effective and valid assessment tool to evaluate the computational thinking skills of Turkish university students.

1. INTRODUCTION

1.1. Computational Thinking

Computational thinking (CT) is a term referring to the concepts, skills, and ideas of computer science applied in problem-solving in an interdisciplinary way (Araujo *et al*., 2019). The use of CT, especially as a term, can be dated back to the time when humans first met computers and then Seymour Papert's pioneering attention to computing in the realm of education (Rode *et al*., 2015). This term was popularized by Jeanette Wing's usage of the term CT in 2006 and after that, various descriptions have been made for the definition of CT (Hanid *et al*., 2022). Wing (2006) regarded CT as a vital skill for everyone and didn't confine CT to only computer science and added that students' analytical skills should be supported by CT besides reading, writing, and arithmetic. In other words, CT is not just about programming, but it is about an entire way of thinking (Li *et al*., 2020). Therefore, CT could promote new perspectives and solutions to problems in many other disciplines (Buckley, 2012). CT refers to possessing skills,

*CONTACT: Ali ORHAN ali_orh_an@hotmail.com School of Foreign Languages, Zonguldak Bülent Ecevit University, Zonguldak, Türkiye

e-ISSN: 2148-7456

knowledge, and attitudes required for using computers to solve problems in real life for production purposes (Aho, 2012). Therefore, it can be pointed out that CT is not a skill that should be possessed only by individuals who are working in computer science. Instead, it is vital for everyone because it may be used to solve problems in daily life. Besides direct descriptions, by working on a program called Scratch which helps students with creating their own interactive stories, Brennan and Resnick (2012) developed a definition of CT involving three key dimensions: (1) Computational concepts that designers employ as they program such as sequences, loops, and events, (2) computational practices that designers develop as they program such as iterating, debugging, and abstracting, (3) computational perspectives that designers form about the world and themselves such as expressing, connecting, questioning helping them realize CT is a medium of creation and realization of being with others and empowering to ask questions. So, it can be inferred that computational concepts and computational practices are not merely enough for the students; they are also required to develop computational perspectives (He *et al.*, 2021).

CT is a kind of analytical thinking that shares the notion of problem-solving in mathematics, designing, and evaluating large systems and complex systems in engineering, and understanding computability, intelligence, and the mind and human behavior in scientific thinking (Wing, 2008). There are many ways to conceptualize CT. Common aspects are decomposition, abstraction, algorithms, evaluation, and generalization (Kang *et al.*, 2023). As Kang *et al.* (2023) stated, decomposition is about dividing a complex problem into small problems or several parts that can be understood and solved. Abstraction refers to removing unnecessary details to reduce the complexity of the task or problem. In other words, during the problem-solving process, it is about the extraction, collecting, or creating key data regarding problem-solving. The algorithm's dimension is about the transformation of the problem from symbolic expressions into a computational model. Evaluation dimension tries to assess the accuracy of the formulated rules, the effect of proposed solutions, and the appropriateness of the methodology. Finally, generalization includes summarizing the data collection and analysis process to reveal its specific problem-solving steps for developing a model for the imitating process. Also, Doleck *et al.* (2017) divided CT skills into six which are algorithmic thinking, cooperativity, creativity, critical thinking, and problem-solving. Social cooperation or collaboration is one of the core aspects of CT (Standl, 2016). As computational issues require large-scale networking and complex data processing, results are attained nearly only with cooperation and collaborative problem-solving (Doleck *et al.*, 2017).

Several 21st-century skills, like creativity, critical thinking, and problem-solving, are aligned with CT (Lye & Koh, 2014). The importance of creativity in the CT process is not only in producing creative output but also in developing new ways of thinking to solve challenges (Dagiene *et al.*, 2019). Moreover, creativity is also related to computer science in terms of motivation and interest in the information technologies field, serving as a catalyst to solve algorithmic problems, creating computational artifacts, and developing new knowledge (Hershkovitz *et al.*, 2019). Computational creativity is also regarded as how creativity is manifested in solutions within the learning environments (Israel-Fishelson *et al.*, 2021). Cooperativity is also one of the main parts of CT as it is a fundamental dimension of complex problem-solving existing in CT (Lemay *et al.*, 2021). As CT is a problem-solving process, it might be challenging for one to come up with a solution when facing complicated problems. Therefore, to solve a problem cooperatively, students must first know how to describe and decompose it into multiple components (Jiang & Li, 2021).

Algorithmic thinking is also one of the skills in CT. Knuth (1974) in those years put the algorithm at the heart of computer science and described it as "an algorithm is a clearly specified set of rules that tells how to produce specified output information from given input data in a finite number of steps." A program is a specific representation of an algorithm, much as the term "data" refers to a specific representation of "information". Choi *et al.* (2017) describe an algorithm as putting a variety of skills in order to be ready to solve a problem and then putting forward the ability to search for information among varied data. Algorithms can occur in many forms. They can be paragraphs, flow charts, images, or pictures, lists or any combination of these. It should be borne in mind for students that those algorithms are creative processes and there is no right or wrong algorithm if the steps make sense and end up with the correct product (Peel & Friedrichsen, 2018).

CT is also regarded as a complement of critical thinking as a way of reasoning to solve problems, make decisions, and interact with the world (Kules, 2016). CT also integrates critical thinking and existing knowledge to solve complex technological problems (Voskoglou & Buckley, 2012). CT provides students with essential critical thinking skills that allow them to conceptualize, analyze, and solve more complex problems (Tran, 2019). As it is seen, CT overlaps all the skills to solve technological, academic, or any other problems in people's daily lives in this digital era. According to Selby (2015), these skills are evaluation, algorithm design, abstraction (functionality, data), decomposition, and generalization (including pattern recognition). Moreover, Selby (2015) said that these skills can be assigned to the levels of Bloom's taxonomy. Evaluation can be assigned to the level of evaluation, algorithms to the level of synthesis, abstraction and decomposition to the analysis, and generalization to the level of application.

In today's digitally-driven society, it is essential to introduce and cultivate CT skills in students from an early age (Hsu *et al.*, 2018). As digital technologies become increasingly integrated into every aspect of life and work, fostering CT skills prepares students not only for future careers but also for effective problem-solving and informed decision-making in a complex, technology-rich world (Akerfeldt *et al.*, 2024). Key elements of CT like abstraction and algorithms help learners understand and build systems of understanding which can cut across various academic disciplines such as language, science, or mathematics resulting in better learning outcomes (Grover & Pea, 2013). In their meta-analysis study investigating the effect of CT on academic achievement, Lei *et al.* (2020) concluded that CT and academic achievement are positively correlated. Along with this, CT has started to attract huge attention from researchers and educators (Rao & Bhagat, 2024), and it has been integrated into curriculums of various countries like the USA, France, Austria, etc. (Akerfeldt *et al.*, 2024).

2.2. How to Assess Computational Thinking

In recent years, computing has made lots of leaps of innovation and imagination as it eases our efforts to solve compelling problems such as curing or preventing diseases and makes us understand our nature as humans and aware of our relationships with the world. In return for this, these advances require computing-supported individuals to reach new horizons for problem-solving (Barr & Stephenson, 2011). In other words, computational processes to solve problems need CT ability. The developments in the CT concept can also be seen when looking at Google Scholar's search reports rising from the 90's (3 studies) to 2020 (2660 studies). Research and practice have increased, and as was aforementioned, researchers emphasized that CT is a core skill for all students, not only for students in computer sciences (Chen *et al.*, 2017a; Wing, 2006). As the importance of CT for students from all educational levels is clear, assessing CT using valid and reliable instruments is vital.

Most of the assessment tools for CT generally aim to assess students' programming abilities or computer knowledge, which are not suitable for students not being exposed to computer knowledge (Dağ, 2019; Romero *et al.*, 2017; Topallı & Çağıltay, 2018). Also, these measurement tools cannot be used for general testing in some of the disciplines as they are restricted by teaching conditions. On the other hand, there are some other scales that subjectively assess individuals' abilities for CT (Gülbahar *et al.*, 2019; Kılıç *et al.*, 2021; Korkmaz *et al.*, 2017; Kukul & Karataş, 2019; Tsai *et al.*, 2020; Yağcı, 2019). However, these tools have an important deficiency as they only measure the ability that individuals think they have, not their CT performance, although they can be easily administered. In other words, the majority of the scales aiming to assess CT in the literature are self-report Likert-type scales and they only measure the ability that individuals think they have. However, we need tests including multiple-choice questions to measure individuals' CT performance. By CT performance, we mean that individuals' CT skills are measured by performance tests, including multiple-choice questions. We think that these kinds of tests are more important than Likert-type scales to measure CT. Therefore, it can be said that previous assessment tools based on self-report scales and computer programming is easy and far from assessing students' CT abilities (Kang *et al.*, 2023). In their systematic review study, Tikva and Tambouris (2021) collected CT studies from 2006 to 2020, and they concluded that only 10 studies assessed CT ability by using multiple-choice and fill-in-the-blank questions, even in the programming field. Moreover, despite the interest in the field of CT in K-12, including fields outside of science, technology, and programming, there has been less interest for CT in higher education (Czerkawski & Lyman, 2015). As CT is a new research area, especially in an education environment, there is a clear need for new and to the point measurement tools that can assess university students' CT abilities in a valid and reliable way (Grover, 2017; Küçükaydın & Akkanat, 2022) because the existing tools based on programming or computer knowledge are not suitable for university students who do not have programming experience or computer knowledge. The newly developed College Students' Computational Thinking Multidimensional Test (CTMT) by Kang *et al.* (2023) is a measurement tool that has good psychometric quality, and it has high potential to satisfy the need for a discipline-free and to-the-point measurement tool designed to measure university students' CT abilities. The need for new and to-the-point measurement tools designed to assess university students' performance-based CT skills is also valid for the literature in Türkiye because the literature lacks these kinds of measurement tools that can assess university students' performance-based CT skills. Therefore, it can be said that the lack of new and to-the-point measurement tools designed to assess university students' CT performances is an important gap in CT literature.

This study aimed to translate and adapt the CTMT developed by Kang *et al.* (2023) into Turkish and to investigate its psychometric qualities with Turkish university students. Along with this aim, the following question was sought:

1. How are the psychometric qualities of the Turkish version of the CTMT?

It is widely accepted that CT is an important attribute of individuals in this century; however, there has been an ongoing debate regarding its definition and models, and this appears as an important barrier against assessing CT of individuals (Lai, 2021; Lai & Ellefson, 2023; Lu *et al.*, 2022). Therefore, it can be said that this study is important because it will contribute to this discussion regarding the conceptualization of CT and expand the CT literature by testing the multidimensional model consisting of five CT dimensions of decomposition, generalization, abstraction, algorithm, and evaluation proposed by Kang *et al.* (2023). Also, this study will provide cross-cultural evidence regarding the structure and item quality of the recently developed CTMT which has a great potential to fulfill the need for a discipline-free and to the point measurement tool designed to assess university students' performance-based CT skills. Besides, this study will pave the way for cross-cultural studies aiming to compare the CT skills of students from different countries and to investigate the effectiveness of cultural elements on CT skills by translating and adapting the CTMT into Turkish.

2. METHOD

In this study, ITC guidelines for translating and adapting tests (ITC, 2017) were followed to translate and adapt the CTMT into Turkish. First of all, the CTMT was translated into Turkish following forward translation, synthesis of translations, expert committee review, and back translation steps. Then, adaptation phase including cognitive interviewing and pilot testing was conducted. Finally, validation phase including the investigating psychometric properties of the Turkish version of CTMT was carried out.

2.1. College Students' Computational Thinking Multidimensional Test

The CTMT, aiming to assess university students' five dimensions of CT skills, namely, decomposition, generalization, abstraction, algorithm, and evaluation, was developed by Kang *et al.* (2023) based on life-story situations. The definitions and explanations for these five dimensions can be seen in Table 1. Kang *et al.* (2023) first wrote 20 story scenarios, and one question related to each scenario regarding the five dimensions of CT. Then, they conducted a pilot study with these 20 items, and after the pilot study, they excluded 6 items and finalized the test with 14 items. The structure validity and item quality of the CTMT, containing 14 multiple-choice items, were verified through multidimensional item response theory using the data collected from 450 college students studying in different disciplines in China. Firstly, the five-dimensional structure of the CTMT was verified by a model comparison using the multidimensional item response theory. After the five-dimension structure was verified, they conducted item quality analyses (difficulty, discrimination, item characteristic surface). Item quality analyses revealed that two items have poor psychometric standards, and the rest of the items have good quality regarding difficulty, multidimensional discrimination index (MDISC), and discrimination of various ability dimensions. Therefore, these two questions were excluded from the test, and the CTMT was finalized with 12 items. In short, the CTMT, containing 12 items, has good internal validity, high item quality, and can discriminate between university students' different disciplines. Therefore, it can be stated that the CTMT can be used as an effective tool to measure university students' CT skills.

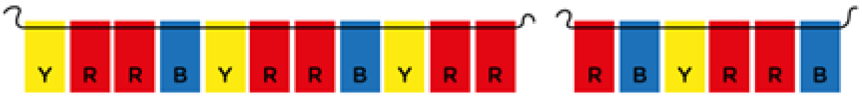
Table 1. *Five dimensions of computational thinking and their explanations*

Skills	Explanations
Evaluation	Correctness of formulated rules to be assessed. Furthermore, the efficacy of a solution and the application of appropriate methods to assess the best solution among multiple rules or algorithms to solve the problem can be determined (Kang *et al.*, 2023).
Algorithm	It comprises a series of clear and precise instructions that are identified and planned in a specific order for the resolution of a problem. A solution to the problem can be obtained by evaluating these clearly defined instructions (Fernandez *et al.*, 2018).
Abstraction	The core of abstraction is concentrating on significant details while dismissing unimportant information (Ezeamuzie *et al.*, 2022). The process of removing unnecessary details while dealing with complex problems or tasks. Throughout the problem-solving process, it can express real problems in symbolic ways (Kang *et al.*, 2023).
Decomposition	Decomposition is the breaking down of a process or thing into smaller components with the overarching purpose of making problems manageable and simplifying their solutions. These small parts or problems can then be decomposed and analyzed, and the relationship and logical sequence between them can be determined (Ezeamuzie *et al.*, 2022; Kang *et al.*, 2023).
Generalization	Generalization entails gathering and evaluating facts, thinking about problem-solving stages, and creating a model to imitate operations. Ideas and solutions from one problem are applied to another after problem-solving for a specific context (Kang *et al.*, 2023).

2.2. Translation Process of the CTMT

The original CTMT with 12 items was in Chinese and it was translated from Chinese into Turkish by two Turkish researchers who are pursuing their PhD degrees in China. Then, two versions of translations were compared with an expert in Türkiye who has high Chinese and Turkish language proficiency by selecting the best alternative for each scenario and item, and the best version was agreed on. After that, the Turkish version of CTMT was back translated into Chinese by another researcher who has high proficiency in two languages and had no idea about the original test. Then, comparing the back-translated version with the original test revealed that the test was accurately translated into Turkish. Finally, the name of the people and cities in the test was changed to Turkish names to increase the understandability of the test by Turkish students. After that, two faculty members who are working in the Mathematics field checked the Turkish version of CTMT to ensure the understandability of the questions in the test. Also, the Turkish version of CTMT was checked by a Turkish teacher to ensure the accuracy of Turkish wording and grammar. Then, a pilot study was conducted using the Turkish version of the test with 20 students. After this pilot study, the students were interviewed regarding the understandability and difficulty of the questions which showed that the questions in their current form were understandable and suitable for university students to be used for assessing CT. The sample questions from the Turkish version of CTMT can be seen in Table 2.

Table 2. *Sample questions from CTMT*

1. There are four people waiting for an interview at the same time in the same unit. The required time for each interview is 3 minutes, 6 minutes, 9 minutes, and 10 minutes, respectively, and only one person can be interviewed at a time. What is the minimum time required to shorten the combined interview time and waiting time for these four people? A. 50 B. 58 C. 63 D. 66
2. Simay has made a colorful paper strip. The paper strip has three different colors (yellow Y, red R, and blue B), and Berkay Mehmet has cut a section of the strip (pictured below). Berkay Mehmet says he will return the paper strip if Simay correctly guesses how many pieces of paper are in the cut section. So, how many pieces of paper have been cut? A. 30 B. 31 C. 32 D. 33 
3. A new circular area with a perimeter of 1.800 meters has been constructed within a park. Landscaping is now required. To maximize the space, the plan is to plant willow trees every 6 meters and a pine tree between every two adjacent willow trees. How many trees will be planted in total? A. 300 B. 600 C. 900 D. 1000

2.3. Participants

The data for this current study were collected from 359 university students studying in different disciplines of a state university in the northern part of Türkiye in the fall term of the 2023-2024 academic year. The students were randomly selected and they participated in the study on a voluntary basis. 168 of these students were female (46.8%) and 191 of them were male (53.2%). The mean age of the students was 20.36 (SD=1.66) ranging from 17 to 27. 33.1%, 18.4%, 17.5%, 17%, and 13.9% of them were freshman year, junior year, senior year, sophomore year, and English preparatory year students, respectively.

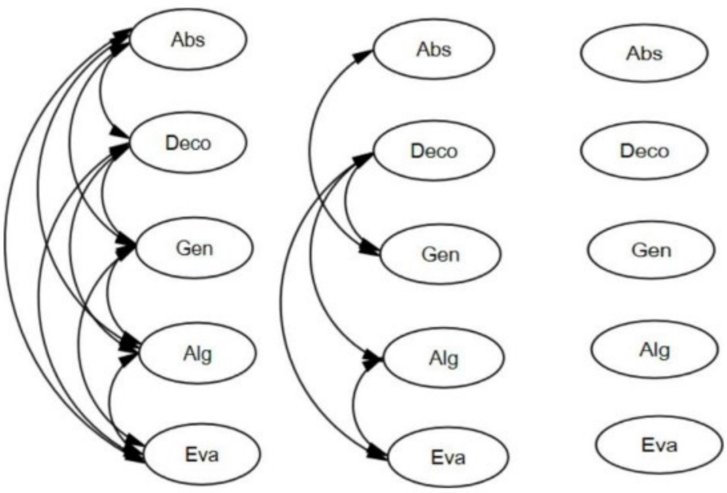
2.4. Data Collection

Following ethical committee approval from the first author's affiliated university, the data for this study were gathered in pen and paper format by the researchers. Privacy and confidentiality issues were shared with the students, and they were informed about the aim of this study and their right to withdraw anytime they want. The test was completed in approximately 50 minutes.

2.5. Data Analysis

The primary aim of this study was to validate the multidimensional structure of the CTMT in a Turkish university student sample. The use of multidimensional item response theory (MIRT) analysis directly supports this aim by enabling simultaneous estimation of dimensional structure and item quality (Ackerman *et al.*, 2003; McDonald, 2000; Reckase, 1997), providing evidence for both construct validity and item functioning in the Turkish adaptation. Also, if the assumption of unidimensionality is not satisfied in item response data, MIRT is more appropriate for modeling the interaction between items and examinees (Ackerman, 1994; Bonifay, 2019; Reckase, 2006). Given that the original CTMT has multidimensional structure and the original scale development study (Kang *et al.*, 2023) of the CTMT used MIRT, similar analyses were conducted with five dimensions in the present study to allow for comparability of results. In this study, multidimensional IRT analyses were employed using the MIRT (Chalmers, 2012) package in R software. The multidimensional two-parameter logistic (2PL) model, which allows not only difficulty but also the discrimination of the items to vary, was preferred. Each of the 12 items on the scale was dichotomously scored as "correct" or "incorrect". Thus, multidimensional item discrimination and item difficulty parameters were estimated for each item. Item and model fit values were also estimated using the MIRT package. Considering the five-dimensional structure of the CTMT scale, the QMCEM (quasi-Monte Carlo EM) algorithm was used to obtain parameter estimates. Following Kang *et al.* (2023), we tested three models (see Figure 1): (1) a fully correlated model where all dimensions are freely correlated; (2) a partial correlated model where only theoretically related dimensions are allowed to correlate; and (3) an uncorrelated model with orthogonal dimensions. These models reflect competing theoretical assumptions about the structure of CT: whether CT skills are fully integrated, partially overlapping, or independent. Testing all three allows us to assess which structure best fits the data and aligns with prior theory. As in Kang *et al.* (2023), Model 1 (fully correlated model) allows correlation between five dimensions, Model 2 (partially correlated model) allows correlation between some of the dimensions, and the correlation between dimensions is constrained to zero in Model 3 (uncorrelated model). The model fit indices estimated for each model include Akaike information criterion (AIC), Bayesian information criterion (BIC), sample-size adjusted BIC (SABIC), and Hannan-Quinn (HQ) criterion. It was assumed that models with smaller values for these indices showed better fit (Chen *et al.*, 2017b; Vrieze, 2012). Log likelihood (LL) values were also reported and LL values are expected to be negative; a higher (less negative) LL value indicates better model fit (Reckase, 2006). Comparisons between models were also made using the anova function in the R program. The model that best fits the data among the three models was selected and the results were reported based on this model. In addition to parameter estimates, the frequencies of correct responses for each item were calculated.

Figure 1. *Three multidimensional models estimated in this study.*



3. RESULTS

The descriptive statistics of the items on the CTMT are presented in Table 3. Preliminarily, we examined the frequency distribution of all item responses by calculating the proportion of correct responses (i.e., mean). The mean values of the items ranged from 0.253 (Item 7) to 0.763 (Item 5). The skewness values ranged between −1.244 to 1.138, which was within the acceptable range of -2 to + 2. The kurtosis values were between -2.011 and -0.456, which was also within the acceptable range of -7 to + 7 (Bryne, 2010). Inter-item correlations between CTMT items are also presented in Table A1 (see Appendix). As shown in Table A1, the correlation values ranged from -0.062 to 0.270.

Table 3. *Descriptive statistics for CTMT items*

Item #	*M*	*SD*	Skewness	Kurtosis
1	0.632	0.483	-0.551	-1.706
2	0.504	0.501	-0.017	-2.011
3	0.513	0.501	-0.050	-2.009
4	0.524	0.500	-0.095	-2.002
5	0.763	0.426	-1.244	-0.456
6	0.348	0.477	0.640	-1.599
7	0.253	0.436	1.138	-0.709
8	0.599	0.491	-0.405	-1.846
9	0.281	0.450	0.977	-1.052
10	0.674	0.469	-0.746	-1.452
11	0.607	0.489	-0.441	-1.816
12	0.448	0.498	0.208	-1.968

As in Kang *et al.* (2023), this study analyzed three different multidimensional models using the same data set. The results of the model-data fit indices of three multidimensional models are presented in Table 4. As shown in Table 4, Model 2 has the best fit among the tested models as it has the lowest values for AIC, BIC, SABIC, and HQ model fit indices. This result is consistent with Kang *et al.* (2023). In addition, the following model fit indices were obtained for Model 2: M_2=57.360 (*df*=36, *p*=.013), RMSEA=0.041 (90% CI [0.019, 0.059]), SRMR=0.062, TLI=0.861, CFI=0.924. TLI value (0.861) is slightly below the conventional cutoff for acceptable model fit. However, we would like to note that TLI is known to be sensitive to both model complexity and sample size (Bentler, 1990). As our model includes five dimensions with partially constrained correlations, this complexity may have affected the TLI. Despite this, other model fit indices such as RMSEA (0.041), SRMR (0.062), and CFI (0.924) all indicate acceptable to good fit. Most of these structural equation modeling-based indices show that Model 2 has an adequate fit except for significant M_2 (Maydeu-Olivares & Joe, 2006). According to Anova test comparisons, both Model 1 (chi-square=74.766, *df*=10, *p*<.001) and Model 2 (chi-square=66.24, *df*=5, *p*<.001) had statistically better fit than Model 3. However, no statistically significant (chi-square=8.526, *df*=5, *p*=.13) difference was found between the fit of Model 2 and the fit of Model 1. Although the difference between Model 1 and Model 2 was not statistically significant, Model 2 had the lowest AIC, BIC, SABIC, and HQ values (see Table 4), consistent with Kang *et al.* (2023). Moreover, Model 2 allows for partial correlations between dimensions, which is theoretically meaningful for the CTMT, as some dimensions are expected to be related while others may remain distinct. Therefore, Model 2 was selected as the best fitting model.

Table 4. *Fit indices and log-likelihood statistics for the three tested models.*

Model #	AIC	SABIC	HQ	BIC	LL
Model 1	5437.55	5470.96	5510.13	5620.07	-2671.78
Model 2	**5436.08**	**5465.93**	**5500.94**	**5599.18**	**-2676.04**
Model 3	5492.32	5518.62	5549.46	5636.00	-2709.16

Note. AIC=Akaike Information Criteria, BIC=Bayesian Information Criteria, SABIC=sample size adjusted BIC, HQ= Hannan-Quinn criterion, LL=loglikelihood statistics.

The correlations between the dimensions are presented in Table 5. In the partial correlated model, some dimensions are specified to be independent (r=0.000) based on prior theory and the original model structure of Kang *et al*. (2023). As shown in Table 5, the correlation estimates ranged from 0.105 to 0.932. The results indicated that the smallest correlation was between decomposition and generalization dimensions, while the highest relationship was between evaluation and decomposition dimensions. The correlation values among other dimensions varied between moderate and high based on Schober *et al.*'s (2018) study. The results indicated that most of the dimensions were highly correlated.

Table 5. *Correlations among MIRT multidimensional capabilities.*

	Abstraction	Evaluation	Algorithm	Decomposition	Generalization
Abstraction	-				
Evaluation	.000	-			
Algorithm	.000	.582	-		
Decomposition	.000	.932	.678	-	
Generalization	.894	.000	.000	.105	-

Table 6 presents the estimates of item parameters obtained under partial correlation model. The a1, a2, a3, a4, and a5 values represent the item slopes for abstraction, evaluation, algorithm, decomposition, and generalization, respectively. In Table 6, the zero (.000) values in the a1–a5 columns reflect the unidimensional item design of the CTMT, where each item is specified to measure a primary dimension only. Non-target dimensions are constrained to zero in the MIRT model to maintain interpretability and avoid parameter estimation instability. For example, for F2, which is associated with items M4, M5, M7, M8, item slope (a2) values for other items are reported as .000. The item slope estimates ranged from -0.531 to 0.557 for abstraction indicators, -6.338 to 2.011 for evaluation indicators, 0.184 to 1.498 for algorithm indicators, -1.401 to 8.220 for decomposition indicators, and -0.525 to 1.917 for generalization indicators. The item intercept value is indicated by the letter d in Table 6. Item intercept values ranged from -2.453 to 1.413. In Table 6, MDISC denotes the multidimensional discrimination index (Reckase, 2009), and MDIFF (Reckase, 2009) represents the multidimensional item difficulty index. The difficulty estimates ranged from -1.379 (Item 5) to 0.873 (Item 6) while discrimination estimates were between 0.184 (Item 3) and 10.379 (Item 7). There was heterogeneity in the estimates of discrimination parameters, and a majority of the items had good discrimination. The only problematic item was Item 3, which had a small discrimination parameter (0.184). This item needs to be improved. Item 7 was found to be the most discriminative item with the steepest slope. There was also heterogeneity in the estimates of difficulty parameters. Except for Item 5, all the items were within the range of -1.00 to +1.00 logits, indicating most of the items can be ranked at the medium and easy level. As shown in Table 6, 8 of the 12 items on the CTMT had negative MDIFF values, indicating that the majority of the items on the scale were easy.

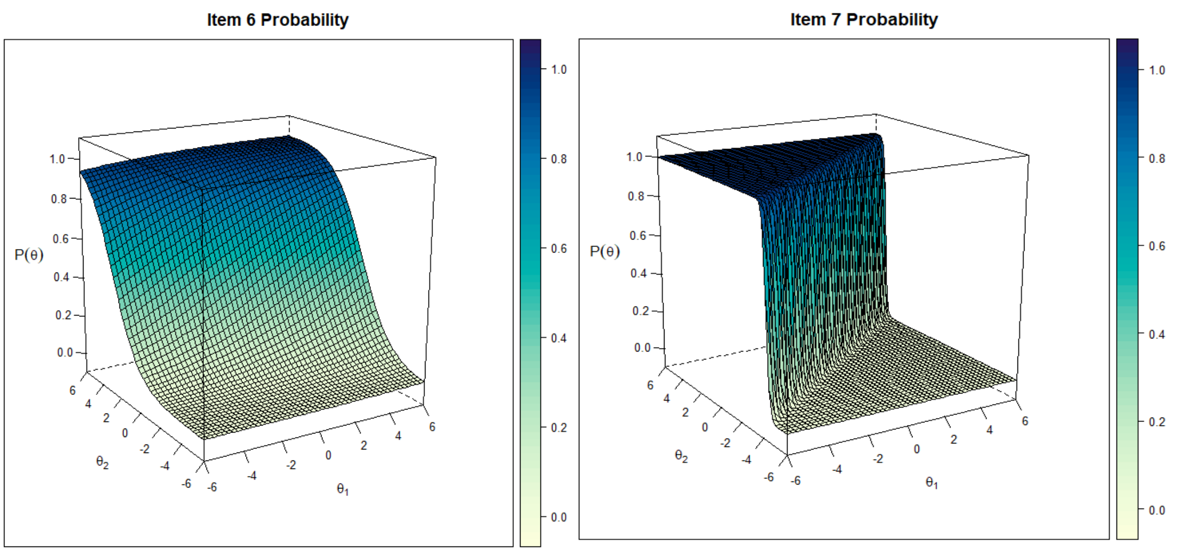
Based on the discrimination and difficulty parameters in Table 6, two items (most difficult and most discriminative) were selected to present their item characteristic surface (see Figure 2). These surfaces illustrate how the probability of a correct response varies as a function of the latent trait(s). The axes (theta1 and theta2) represent two of the five CT dimensions for each item, selected for visualization. The slope of the surface reflects item discrimination on these dimensions. The vertical axis indicates the probability of a correct response. The steepness of the surface indicates item discrimination, while position reflects item difficulty.

Table 6. *Item parameter estimates based on partial correlated models.*

Item #	MDIFF	a1	a2	a3	a4	a5	*d*	MDISC
1	-0.658	0.000	0.000	0.000	0.998	-0.024	0.657	0.998
					(0.248)	(0.258)	(0.140)	
2	-0.026	0.461	0.000	0.000	0.719	0.000	0.022	0.854
		(0.285)			(0.228)		(0.122)	
3	-0.273	0.000	0.000	0.184	0.000	0.000	0.050	0.184
				(0.155)			(0.106)	
4	-0.045	0.000	2.011	0.000	-1.401	0.506	0.113	2.503
			(1.214)		(1.021)	(0.281)	(0.126)	
5	-1.379	0.000	0.017	0.879	0.000	-0.525	1.413	1.024
			(0.373)	(0.421)		(0.311)	(0.195)	
6	0.873	0.000	0.000	0.000	0.226	0.801	-0.726	0.832
					(0.253)	(0.266)	(0.137)	
7	0.236	0.000	-6.338	0.000	8.220	0.000	-2.453	10.379
			(6.050)		(7.261)		(1.629)	
8	-0.388	-0.272	1.418	0.000	0.000	0.000	0.561	1.444
	(0.337)	(0.546)					(0.176)	
9	0.749	0.000	0.000	0.838	0.000	1.917	-1.568	2.092
				(0.565)		(0.822)	(0.436)	
10	-0.662	-0.531	0.000	1.498	0.000	0.000	1.053	1.589
		(0.494)		(0.506)			(0.240)	
11	-0.518	0.557	0.000	0.869	0.000	0.000	0.535	1.032
		(0.319)		(0.248)			(0.137)	
12	0.213	0.405	0.000	1.155	0.000	0.000	-0.260	1.224
		(0.341)		(0.289)			(0.139)	

Note. Standard errors are in parentheses.

Figure 2. *Item characteristic surfaces of the most difficult (Item 6) and most discriminative (Item 7) items.*



Note. P(θ)=logit (i.e., Probability of a correct response), θ1= Evaluation, θ2= Decomposition.

4. DISCUSSION and CONCLUSION

In this study, it was aimed to translate and adapt the CTMT, which is a measurement tool with good psychometric quality, into Turkish and to investigate its psychometric qualities with Turkish university students. As Kang *et al.* (2023) did in their study, first of all, three different models assuming five dimensions, namely abstraction, evaluation, algorithm, decomposition, and generalization, were estimated using multidimensional IRT and these three different multidimensional models which are Model 1 (fully correlated model), Model 2 (partial correlated model), and Model 3 (uncorrelated model) were compared. The results showed that Model 2 has the best fit among the tested models which is consistent with the results of Kang *et al.*'s (2023) study. Therefore, it can be said that the structure of the Turkish CTMT is similar to the original test. Also, this study showed that the correlation coefficients indicated moderate-to-high correlations between the dimensions which shows us that this study reveals higher correlation coefficients between the dimensions than Kang *et al.*'s (2023) study which concluded low-to-moderate correlation coefficients between the dimensions. Cultural factor and sample differences can be shown as a possible explanation for this difference regarding inter-dimensional correlations. As China and Türkiye have different cultures and the students from these countries possess different cultural backgrounds, the difference regarding inter-dimensional correlations between this study and Kang *et al.*'s (2023) study may be explained with this. The results of the multidimensional item discrimination and item difficulty parameters based on a partially correlated model showed that a majority of the items have good discriminations except for Item 3 which had a small discrimination parameter. This shows us that Item 3 should be revised and improved in the Turkish version. According to the classification of Ding *et al.* (2012), five of the items in the Turkish test had excellent MDISC values. Also, three of them had good MDISC values while another three questions had medium values. Therefore, it can be said that the items in the Turkish version of CTMT had similar discrimination parameters with Kang *et al.*'s (2023) study indicating 3, 7, and 2 items had excellent, good, and medium MDISC values, respectively. This study showed that Item 3 had the smallest discrimination parameter and Item 7 was the most discriminative item in the test. However, Kang *et al.* (2023) concluded that Item 3 had good value for MDISC. Item 7 showed an excellent MDISC value in the Turkish and original test, and it was the most discriminative item in the Turkish version, while it was the third most discriminative item in the original version.

Also, it was found that most of the items in the Turkish version can be ranked at the medium and easy level. 8 of the 12 items on the CTMT had negative MDIFF values, indicating that the majority of the items on the test were easy. Kang *et al.* (2023) also revealed that 9 of the 12 items had negative MDIFF values. Therefore, it can be said that the proportion of the items in the translated and original tests in terms of difficulty is similar. This study also showed that Item 5 and Item 6 were the easiest and the most difficult items in the Turkish test, respectively. Kang *et al.* (2023) found that Item 5 was the third easiest item, while Item 6 had medium difficulty.

Therefore, it can be said that the translated and original test possessed mostly similar characteristics in terms of item discrimination and difficulty parameters. However, they also have some small differences regarding discriminative power and difficulty of items, which might be explained by sample differences and translation effects. As Türkiye and China are different countries in terms of culture and education systems, the university students from these countries may also possess different characteristics. As the original CTMT was developed for Chinese students, the different sample characteristics stemmed from the cultural and educational system differences between the two countries may result in small differences regarding discriminative power and difficulty of items.

In short, the CTMT developed by Kang *et al.* (2023), based on real-life story situations to assess university students' five dimensions of CT skills, namely, decomposition, generalization, abstraction, algorithm, and evaluation, was translated and adapted into Turkish. The structure validity and item quality of the Turkish version of CTMT were verified using MIRT, and it was seen that the Turkish test has good psychometric quality. Also, its structure and item difficulty, and discrimination characteristics are similar to the original test. Therefore, it can be said that this study provides cross-cultural evidence for the structure and item quality of the CTMT, as this adaptation study echoed similar results as in the original study. Besides, this study shows that the Turkish version of CTMT can be used as an effective and valid assessment tool to evaluate the CT skills of Turkish university students. Although the importance of CT for individuals is clear in this century, there has been an ongoing discussion regarding its definition and models, which emerges as an important barrier against assessing CT (Lai, 2021; Lai & Ellefson, 2023; Lu *et al.*, 2022). Therefore, it can be said that this study is important because it verified the multidimensional model that emerged in Kang *et al.*'s (2023) study with Turkish university students, and hence, it contributes to the ongoing discussion regarding the definition and models of CT. The multidimensional model consisting of five CT dimensions of decomposition, generalization, abstraction, algorithm, and evaluation can be seen as a verified structure to assess the CT skills of university students. These five dimensions use computer science terminology and align with the original definition of CT while also overcoming the limitations of the discipline (Kang *et al.*, 2023). This five-dimensional model of CT which was confirmed in China by Kang *et al.* (2023) and in Türkiye by this study provides an operational structure of CT which can be easily measured. There are also some other models defined by different organizations for CT. For example, CSTA (Computer Science Teachers Association) proposed a model for CT including sub dimensions, namely, problem decomposition, data analysis, algorithmic thinking, abstraction, automation, and testing and refinement (Lee, 2019). It can be said that the CSTA framework primarily positions CT within the context of computer science and offers limited guidance or examples on how CT can be integrated into other disciplines like mathematics, science, and language, etc. On the other hand, the ISTE (International Society for Technology in Education) (2024) approach offers a slightly broader and interdisciplinary framework for CT than the CSTA. The ISTE defines CT not as a technical skill specific to computer science, but as a mental tool that all students can use in interdisciplinary problem-solving processes and the core elements of CT are decomposition, gathering and analyzing data, abstraction, algorithm design (ISTE, 2024). It can be said that the five-dimensional model of CT which was emerged in Kang *et al.*'s (2023) study and confirmed in this study includes similar core elements like the other models proposed by different organizations like the ISTE and CSTA. Also, this model does not position CT only within the context of computer science and proposes a framework that can be used within different disciplines which can be seen as an important advantage for the CTMT. In addition, the CTMT developed based on this five-dimensional model offers a unique and comprehensive assessment tool for researchers and practitioners to assess performance-based CT skills of students.

4.1. Limitations and Recommendations for Future Research

This study has provided evidence for structure validity and item quality of the Turkish version of CTMT; however, it lacks the investigation of the empirical validity of Turkish CTMT which can be examined by calculating the correlations between the Turkish CTMT and other several outcomes or antecedents of CT like academic achievement, critical thinking, and problem solving, etc. Therefore, future studies can be conducted to investigate the relation between the scores produced by the Turkish CTMT and other outcomes or antecedents of CT to verify the Turkish test's empirical validity. Also, it would be a good idea to translate and adapt the CTMT into different languages and cultures because these adaptation studies can provide additional evidence for the structure and item quality of the CTMT. Besides, although this study contributes to the ongoing discussion regarding the conceptualization of CT by verifying the multidimensional model consisting of five CT dimensions, there is still a need for other studies that may contribute to this issue. Therefore, more theoretical studies focusing on the definition, dimensions, cultivation, and assessment of CT should be conducted. In addition, the psychometric qualities of the CTMT can be tested with participants from different education levels (e.g., high school students). Also, longitudinal studies can be conducted to investigate the development of CT skills of students in time.

Declaration of Conflicting Interests and Ethics

The authors declare no conflict of interest. This research study complies with research publishing ethics. The scientific and legal responsibility for manuscripts published in IJATE belongs to the authors. **Ethics Committee Number**: Zonguldak Bülent Ecevit Üniversitesi, Human Research Ethics Committee, 24.10.2023/368263, Protocol no. 322.

Contribution of Authors

All authors contributed to the manuscript equally.

Orcid

Ali Orhan https://orcid.org/0000-0003-1234-3919
İnan Tekin https://orcid.org/0000-0002-5616-4325
Sedat Şen https://orcid.org/0000-0001-6962-4960

REFERENCES

Ackerman, T.A., Gierl, M.J., & Walker, C.M. (2003). Using multidimensional item response theory to evaluate educational and psychological tests. *Educational Measurement: Issues and Practice, 22*, 37-51. https://doi.org/10.1111/j.1745-3992.2003.tb00136.x

Ackerman, T.A. (1994). Using multidimensional item response theory to understand what items and tests are measuring. *Applied Measurement in Education*, *7*(4), 255-278. https://doi.org/10.1207/s15324818ame0704_1

Aho, A.A. (2012). Computation and computational thinking. *The Computer Journal, 55*(7), 832e835. https://doi.org/10.1093/comjnl/bxs074

Akerfeldt, A., Kjällander, S., & Petersen, P. (2024). A research review of computational thinking and programming in education. *Technology, Pedagogy and Education*, *33*(3), 375–390. https://doi.org/10.1080/1475939X.2024.2316087

Araujo, A.L.S.O., Andrade, W.L., Guerrero, D.D.S., & Melo, M.R.A. (2019, February). How many abilities can we measure in computational thinking? A study on bebras challenge. In E.K. Hawthorne, *et al.* (Eds.), *SIGCSE '19 Proceedings of the 50th ACM Technical Symposium on Computer Science Education* (pp. 545-551).

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*(1), 48-54. https://doi.org/10.1145/1929887.1929905

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

Bonifay, W. (2019). *Multidimensional item response theory*. Sage Publications.

Brennan, K., & Resnick, M. (2012, April). *New frameworks for studying and assessing the development of computational thinking* [Conference presentation]. Annual Meeting of the American Educational Research Association, Vancouver, Canada.

Bryne, B. (2010). *Structure equation modeling with AMOS: Basic concepts applications and programs*. Routledge Taylor and Francis Group.

Buckley, S. (2012, October). *The role of computational thinking and critical thinking in problem solving in a learning environment* [Conference presentation]. European Conference on e-Learning, Oxfordshire, England.

Chalmers, R.P. (2012). Mirt: A multidimensional item response theory package for the R environment. *Journal of Statistical Software, 48*(6), 1-29. https://doi.org/10.18637/jss.v048.i06

Chen, G., Shen, J., Barth-Cohen, L., Jiang, S., Huang, X., & Eltoukhy, M. (2017a). Assessing elementary students' computational thinking in everyday reasoning and robotics programming. *Computers & Education, 109*, 162-175. https://doi.org/10.1016/j.compedu.2017.03.001

Chen, Q., Luo, W., Palardy, G.J., Glaman, R., & McEnturff, A. (2017b). The efficacy of common fit indices for enumerating classes in growth mixture models when nested data

structure is ignored: A monte carlo study. *Sage OPEN, 7*(1), 1-19. https://doi.org/10.1177/2158244017700459
Choi, J., Lee, Y., & Lee, E. (2017). Puzzle based algorithm learning for cultivating computational thinking. *Wireless Personal Communications, 93*, 131-145. https://doi.org/10.1007/s11277-016-3679-9
Czerkawski, B.C., & Lyman, E.W. (2015). Exploring issues about computational thinking in higher education. *TechTrends*, 59, 57–65. https://doi.org/10.1007/s11528-015-0840-3
Dagiene, V., Futschek, G., & Stupurienė, G. (2019). Creativity in solving short tasks for learning computational thinking. *Constructivist Foundations, 14*(3), 382-396.
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
Ding, S., Luo, F., & Tu, D. (2012). *Special topic research on new progress in project response theory*. Beijing Normal University Press.
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*, 355-369. https://doi.org/10.1007/s40692-017-0090-9
Ezeamuzie, N.O., Leung, J.S., & Ting, F.S. (2022). Unleashing the potential of abstraction from cloud of computational thinking: A systematic review of literature. *Journal of Educational Computing Research, 60*(4), 877-905. https://doi.org/10.1177/07356331211055379
Fernandez, J.M., Zúñiga, M.E., Rosas, M.V., & Guerrero, R.A. (2018). Experiences in learning problem-solving through computational thinking. *Journal of Computer Science and Technology, 18*(2), 136-142. https://doi.org/10.24215/16666038.18.e15
Grover, S., & Pea, R. (2013). Computational thinking in K–12: A review of the state of the field. *Educational Researcher 42*(1). 38-43. https://doi.org/10.3102/0013189X12463051
Grover, S. (2017). Assessing algorithmic and computational thinking in K-12: Lessons from a middle school curriculum. In P. Rich & C. Hodges (Eds.), *Emerging research, practice, and policy on computational thinking* (pp. 269–288). Springer.
Gülbahar, Y., Kert, S.B., & Kalelioğlu, F. (2019). The self-efficacy perception scale for computational thinking skill: Validity and reliability study. *Turkish Journal of Computer and Mathematics Education, 10*(1), 1–29.
Hanid, M.F.A., Said, M.N.H.M., Yahaya, N., & Abdullah, Z. (2022). The elements of computational thinking in learning geometry by using augmented reality application. *International Journal of Interactive Mobile Technologies, 16*(2), 28-41. https://doi.org/10.3991/ijim.v16i02.27295
Hershkovitz, A., Sitman, R., Israel-Fishelson, R., Eguíluz, A., Garaizar, P., & Guenaga, M. (2019). Creativity in the acquisition of computational thinking. *Interactive Learning Environments, 27*(5-6), 628-644. https://doi.org/10.1080/10494820.2019.1610451
He, Z., Wu, X., Wang, Q., & Huang, C. (2021). Developing eighth-grade students' computational thinking with critical reflection. *Sustainability, 13*, 1-21. https://doi.org/10.3390/su132011192
Hsu, T.C., Chang, S.C., & Hung, Y.T. (2018). How to learn and how to teach computational thinking: Suggestions based on a review of the literature. *Computers & Education, 126*, 296–310.
International Society for Technology in Education (ISTE). (2024, June 20). *Computational thinking competencies*. https://iste.org/standards/computational-thinking-competencies
International Test Commission (2017, December 25). *The ITC guidelines for translating and adapting tests* (Second edition). https://www.intestcom.org/files/guideline_test_adaptation_2ed.pdf
Israel-Fishelson, R., Hershkovitz, A., Eguíluz, A., Garaizar, P., & Guenaga, M. (2021). The associations between computational thinking and creativity: The role of personal characteristics. *Journal of Educational Computing Research, 58*(8), 1415-1447. https://doi.org/10.1177/0735633120940954

Jiang, B., & Li, Z. (2021). Effect of Scratch on computational thinking skills of Chinese primary school students. *Journal of Computers in Education, 8*(4), 505-525. https://doi.org/10.1007/s40692-021-00190-z

Kang, C., Liu, N., Zhu, Y., Li, F., & Zeng, P. (2023). Developing college students' computational thinking multidimensional test based on life story situations. *Education and Information Technologies, 28*(3), 2661-2679. https://doi.org/10.1007/s10639-022-11189-z

Kılıç, S., Gökoğlu, S., & Öztürk, M. (2021). A valid and reliable scale for developing programming-oriented computational thinking. *Journal of Educational Computing Research, 59*(2), 257-286. https://doi.org/10.1177/0735633120964402

Knuth, D.E. (1974). Computer science and its relation to mathematics. *The American Mathematical Monthly, 81*(4), 323-343. https://doi.org/10.2307/2318994

Korkmaz, Ö., Çakır, 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

Kukul, V., & Karatas, S. (2019). Computational thinking self-efficacy scale: Development, validity and reliability. *Informatics in Education, 18*(1), 151-164. https://doi.org/10.15388/infedu.2019.07

Kules, B. (2016). Computational thinking is critical thinking: Connecting to university discourse, goals, and learning outcomes. *Proceedings of the Association for Information Science and Technology, 53*(1), 1–16.

Küçükaydın, M., & Akkanat, Ç. (2022). Adaptation into Turkish of the Computational Thinking Test for Primary School Students. *Problems of Education in the 21st Century, 80*(6), 765-776. https://doi.org/10.33225/pec/22.80.765

Lai, R.P.Y. (2021). Beyond programming: A computer-based assessment of computational thinking competency. *ACM Transactions on Computing Education, 22*(2), 14–27. https://doi.org/10.1145/3486598

Lai, R.P.Y., & Ellefson, M.R. (2023). How multidimensional is computational thinking competency? A bi-factor model of the computational thinking challenge. *Journal of Educational Computing Research, 61*(2), 259–282. https://doi.org/10.31219/osf.io/e8n3h

Lee, M.G. (2019, May 30). *Teaching computational thinking in early elementary*. https://csteachers.org/teaching-computational-thinking-in-early-elementary/

Lei, H., Chiu, M.M., Li, F., Wang, X., & Geng, Y.J. (2020). Computational thinking and academic achievement: A meta-analysis among students. *Children and Youth Services Review, 118*, 1-8. https://doi.org/10.1016/j.childyouth.2020.105439

Lemay, D.J., Basnet, R.B., Doleck, T., Bazelais, P., & Saxena, A. (2021). Instructional interventions for computational thinking: Examining the link between computational thinking and academic performance. *Computers and Education Open*, 2, 1-6. https://doi.org/10.1016/j.caeo.2021.100056

Li, Y., Schoenfeld, A.H., DiSessa, A.A., Graesser, A.C., Benson, L.C., English, L.D., & Duschl, R.A. (2020). Computational thinking is more about thinking than computing. *Journal for STEM Education Research, 3*, 1–18. https://doi.org/10.1007/s41979-020-00030-2

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, 34,* 416-461. https://doi.org/10.1007/s12528-021-09305-y

Lye, S.Y., Koh, J.H.L. (2014). Review on teaching and learning of computational thinking through programming: What is next for K-12?. *Computers in Human Behavior*, *41*, 51–61. https://doi.org/10.1016/j.chb.2014.09.012

Maydeu-Olivares, A., & Joe, H. (2006). Limited information goodness-of-fit testing in multidimensional contingency tables. *Psychometrika, 71*(4), 713-732. https://doi.org/10.1007/s11336-005-1295-9

McDonald, R.P. (2000). A basis for multidimensional item response theory. *Applied Psychological Measurement, 24*(2), 99-114. https://doi.org/10.1177/01466210022031552

Peel, A., & Friedrichsen, P. (2018). Algorithms, abstractions, and iterations: Teaching computational thinking using protein synthesis translation. *The American Biology Teacher*, *80*(1), 21-28. https://doi.org/10.1525/abt.2018.80.1.21

Rao, T.S.S., & Bhagat, K.K. (2024). Computational thinking for the digital age: a systematic review of tools, pedagogical strategies, and assessment practices. *Education Technology Research and Development, 72*, 1893–1924. https://doi.org/10.1007/s11423-024-10364-y

Reckase, M.D. (2006). Multidimensional item response theory. In C.R. Rao & S. Sinharay (Eds), Handbook of statistics (pp. 607-642). Elsevier.

Reckase, M.D. (2009). *Multidimensional item response theory* (1st ed.). Springer.

Reckase, M.D. (1997). The past and future of multidimensional item response theory. *Applied Psychological Measurement*, *21*(1), 25- 36. https://doi.org/10.1177/0146621697211002

Rode, J.A., Weibert, A., Marshall, A., Aal, K., von Rekowski, T., Elmimouni, H., & Booker, J. (2015, September). *From computational thinking to computational making* [Conference presentation]. ACM International Joint Conference on Pervasive and Ubiquitous Computing, Osaka, Japan.

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*(42), 1-15. https://doi.org/10.1186/s41239-017-0080-z

Schober, P., Boer, C., & Schwarte, L.A. (2018). Correlation coefficients: Appropriate use and interpretation. *Anesthesia & Analgesia, 126*(5), 1763-1768. https://doi.org/10.1213/ANE.0000000000002864

Selby, C.C. (2015, November). Relationships: computational thinking, pedagogy of programming, and Bloom's Taxonomy. *Proceedings of the workshop in primary and secondary computing education*, 80-87.

Standl, B. (2016, April). A case study on cooperative problem solving processes in small 9th grade student groups. *Proceedings of IEEE Global Engineering Education Conference*, 961-967. IEEE.

Tikva, C., & Tambouris, E. (2021). A systematic mapping study on teaching and learning Computational Thinking through programming in higher education. *Thinking Skills and Creativity, 41*(100849), 1-18. https://doi.org/10.1016/j.tsc.2021.100849

Topallı, D., & Cağıltay, 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

Tran, Y. (2019). Computational thinking equity in elementary classrooms: What third-grade students know and can do. *Journal of Educational Computing Research, 57*(1), 3-31. https://doi.org/10.1177/0735633117743918

Tsai, M.J., Liang, J.C., & Hsu, C.Y. (2021). The computational thinking scale for computer literacy education. *Journal of Educational Computing Research, 59*(4), 579-602. https://doi.org/10.1177/0735633120972356

Vrieze, S.I. (2012). Model selection and psychological theory: A discussion of the differences between the Akaike information criterion (AIC) and the Bayesian information criterion (BIC). *Psychological Methods, 17*(2), 228–243. https://doi.org/10.1037/a0027127

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.

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

Voskoglou, M.G. & Buckley, S. (2012). Problem solving and computers in a learning environment. *Egyptian Computer Science Journal*, *36*(4), 28-46.

Yağcı, M. (2019). A valid and reliable tool for examining computational thinking skills. *Education and Information Technologies, 24*(1), 929–951. https://doi.org/10.1007/s10639-018-9801-8

APPENDIX

Table A1. *Correlation matrix for CTMT items.*

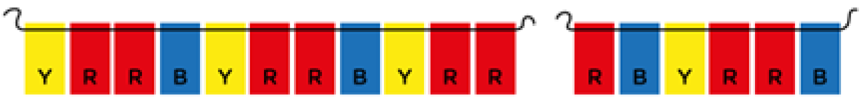
	M1	M2	M3	M4	M5	M6	M7	M8	M9	M10	M11	M12
M1	1.000											
M2	.006	1.000										
M3	-.004	.036	1.000									
M4	.094	.203	-.004	1.000								
M5	.092	.024	.139	.020	1.000							
M6	.072	.140	-.024	.112	.022	1.000						
M7	.179	.181	-.047	-.021	.068	.098	1.000					
M8	.248	.109	-.002	.096	.092	-.022	.137	1.000				
M9	.079	.137	.028	.076	-.016	.219	.234	.019	1.000			
M10	.074	.131	.107	.039	.270	-.041	.159	.122	.012	1.000		
M11	.108	.127	.072	.032	-.005	.037	.115	.098	.199	.159	1.000	
M12	.177	.121	-.062	.053	.081	.093	.234	.167	.183	.161	.198	1.000

Table A2. *Turkish sample questions from CTMT.*

1. Aynı birimde aynı anda mülakat bekleyen dört kişi vardır. Her bir görüşme için gereken süre sırasıyla 3 dakika, 6 dakika, 9 dakika ve 10 dakikadır ve bir seferde sadece bir kişiyle görüşme yapılabilmektedir. Bu dört kişi için görüşme süresi ve bekleme süresinin toplamını en kısa yapmak için gereken minimum süre nedir?

A. 50   B. 58   C. 63   D. 66

2. Simay renkli bir kağıt şerit yapmıştır. Kağıt şeridin üç farklı rengi vardır (sarı Y, kırmızı R ve mavi B) ve Berkay Mehmet şeridin bir kısmını kesmiştir (aşağıda resmedilmiştir). Berkay Mehmet, Simay'ın kesilen bölümde kaç kağıt parçası olduğunu doğru tahmin etmesi halinde kağıt bandı iade edeceğini söyler. O halde kaç parça kağıt kesilmiştir?



A. 30
B. 31
C. 32
D. 33

3. Bir parkın içine çevresi 1.800 metre olan yeni bir dairesel alan inşa edilmiştir. Şimdi alanın etrafının peyzajının yapılması gerekmektedir. Alandan mümkün olduğunca fazla yararlanmak için, her 6 metrede bir söğüt ağacı ve her iki bitişik söğüt ağacının arasına bir akçaağaç ağacı dikilmesi planlanmaktadır. Bu durumda alana toplam kaç ağaç dikilecektir?

A. 300   B. 600   C. 900   D. 1000