International Journal of Assessment Tools in Education
2024, Vol. 11, No. 1, 88–108
https://doi.org/10.21449/ijate.1327082
journal homepage: https://dergipark.org.tr/en/pub/ijate
Research Article

Developing a game-based test to assess middle school sixth-grade students' algorithmic thinking skills

Emre Zengin[1*], Yasemin Karal[2]

[1]80th Year Republic Middle School, Gümüşhane, Türkiye
[2]Trabzon University, Faculty of Education, Department of Computer and Instructional Technologies Education, Trabzon, Türkiye

ARTICLE HISTORY

Received: July 13, 2023

Accepted: Jan. 10, 2024

Keywords:
Algorithmic thinking,
Test,
Middle school,
Game,
Embodied cognition.

Abstract: This study was carried out to develop a test to assess algorithmic thinking skills. To this end, the twelve steps suggested by Downing (2006) were adopted. Throughout the test development, 24 middle school sixth-grade students and eight experts in different areas took part as needed in the tasks on the project. The test was given to 252 students attending the sixth grade who were selected through purposeful sampling. The content validity of the test was ensured by means of obtaining expert opinion, whereas the construct validity was ensured by performing an independent sample t-test on the difference between the lower and upper groups. As a result, the algorithmic thinking skills assessment test was finalized with 22 main items and 2 sample items, totalling 24 items. The KR-20 reliability analysis proved a quite reliable test based on the reliability coefficient of 0.83. As mentioned earlier, the independent sample t-test was applied to the difference of lower and upper groups for construct validation of the test. It was seen that the test items are significant in discriminating the students in the lower and upper groups ($p<0.01$).

1. INTRODUCTION

Technology is developing and affecting social life. These effects are evident in a variety of spheres including education and health and social, commercial, and professional areas. This necessitates individuals to acquire different skills and maintain their individual development. With the acquisition of these skills, which are known as 21st-century skills, individuals are expected to be productive, problem-solving, critical, entrepreneurial, and able to think creatively (Geisinger, 2016; Keane, 2012; Khanlari, 2013). Under these skills lies algorithmic thinking skills, which are common skills that should be possessed by all individuals today (Futschek & Moschitz, 2010; Mumcu & Yıldız, 2018; Sarı et al., 2022). Individuals who have acquired algorithmic thinking skills can produce solutions by applying the algorithm structures in their minds to the problems they encounter in daily life (Thomas et al., 2015; Thomas et al., 2017). Bearing in mind the possibility of facing problems at any moment in life, individuals should also be prepared for these situations (Aytekin et al., 2018; Lertlapnon et al., 2022; Turchi et al., 2019). It is emphasized that algorithmic thinking skills should be taught in a planned and programmed way (Erümit et al., 2018). Several researchers have put forward steps to be followed in the process of developing algorithmic thinking skills (Zsako & Slavi, 2012, p. 55,

*CONTACT: Emre ZENGİN emre_zengin20@gmail.com 80th Republic Middle School, Gümüşhane, Türkiye

e-ISSN: 2148-7456 /

as cited in Szanto, 2002; Vasconcelos, 2007; Zsako & Slavi, 2012).

There are many studies on algorithmic thinking skills, and they include games (Lee et al., 2014; Paspallis et al., 2022; Sungkaew et al., 2022). According to research (Evripidou et al., 2021; Kiss & Arki, 2017; Pivec & Kearney, 2007), games have positive effects on students' attention and motivation (Apostolellis et al., 2014; Debabi & Bensebaa, 2016; Shang et al., 2019) and they boost skills such as problem-solving, flexibility, adaptability and creativity. In addition, it is stated that the actions and rules in games progress in parallel with the logic of algorithmic thinking (Karakasis & Xinogalos, 2020, Wangenheim et al., 2019; Yılmaz, 2020).

A literature review was conducted on studies dealing with the development of algorithmic thinking skills in connection with computer-based games and unplugged games. It was seen that data were collected during the students' achievement of the tasks given in games and assessment was predominantly carried out based on such data (Chuechote et al., 2020; Czakoova, 2020; Kazimoglu, 2020). Some other studies featuring computer-based games used achievement tests and questionnaires (Elshahawy et al., 2020; Hsu & Wang, 2018; Li et al., 2020; Tsukamoto et al., 2017). As for the studies using unplugged games, inferences were made from the data obtained from session recordings (Lin et al., 2020; Scharf et al., 2020) or interviews and observations (Chen & Chi, 2020). Other studies applied hybrid games, and it was seen that the assessment was done by using the data obtained through the think-aloud technique (Lee et al., 2014).

1.1. Algorithm and Algorithmic Thinking

The algorithm is a procedure of sequenced instructions fulfilled to complete a specific task (Borkulo et al., 2021). Algorithmic thinking is the ability to comprehend, execute, create, and evaluate algorithms (Brown, 2015). People attempt to solve every problem they encounter in their daily lives (Doleck et al., 2017; Kanaki & Kalogiannakis, 2022; Yadav et al., 2017). In the problem-solving process, it is regarded as important to reveal the most efficient solution by considering all reasonable possibilities (Hu, 2011; Jancec & Vujicic, 2021; Katai, 2015). Therefore, it is recommended that individuals be trained as good algorithmic thinkers (Czakoova & Udvaros, 2021; Figueiredo et al., 2021; Mezak & Papak, 2018).

1.2. Developing Algorithmic Thinking Skills

Researchers examining the development process of algorithmic thinking skills have suggested some models as seen in Table 1.

Table 1. *Models of algorithmic thinking skill development.*

Zsako and Slavi (2012)	Vascencolos (2007)	Szanto (2002)	Erümit et al. (2018)
1. Recognizing and Understanding Algorithms 2. Implementing Algorithms 3. Analyzing Algorithms 4. Making Algorithms 5. Realizing Algorithms 6. Modifying and Changing Algorithms 7. Designing Complex Algorithms	1. Read and Comprehend the Problem Statement 2. Select Theoretical Concepts That May Be Applied 3. Qualitative Description of the Problem 4. Formulization of a Solution Strategy 5. Test and Description of the Solution	1. Application/Coding 2. Algorithm Writing 3. Analogic Thinking 4. Being Able to Change Algorithm and Adapt to Current Situation 5. Production/ Derivation	1. Understand the Problem 2. Devise A Plan 3. Compare the Strategies 4. Devise and Algorithm 5. Code the Algorithm 6. Identify and Correct the Error in A Different Code 7. Prepare and Code New Algorithms

1.3. Games and Game-Based Learning

According to Huizinga (1955), a game is a sequence of activities that flow according to a set of pre-determined rules within a certain time and place. Yılmaz (2020) defines games as a type of behavior that children perform in order to adapt to the real world. Game-based learning is defined as a learning environment that enables students to achieve their learning goals and solve problems that they may encounter in daily life by providing a sense of achievement through gaming activities (Kim et al., 2009; McFarlane et al., 2002; Prensky, 2001).

1.4. Assessing Algorithmic Thinking

A review was conducted on the previous studies looking into the effect of game-based activities on algorithmic thinking skills. As a result, several data collection tools were found. Assessment was performed by using the data obtained from tests, questionnaires, interviews, observations, and session recordings. There were seven studies that assessed students' algorithmic thinking skills by using tests. These studies are listed in Table 2 below with details including the research title, year, author, level, and scope of the assessment tool.

Table 2. *Studies using multiple-choice tests for assessing algorithmic thinking skills.*

Author	Research Title	Year	Level	Test Scope
Gürbüz et al.	*"What's the weather like today?": A computer game to develop algorithmic thinking and problem-solving skills of primary school pupils*	2017	8-10 year-olds	In a computer-based educational game, students are provided with known variables of sun, temperature, humidity, and wind. Students try to reach the correct answer among 144 possible answers by developing and executing algorithms by using the given values.
Tsukamoto et al.	*Evaluating Algorithmic Thinking Ability of Primary Schoolchildren Who Learn Computer Programming*	2017	3rd through 6th graders	There is a test which is comprised of 3 multiple-choice items on which students perform sequential tasks. Students use the basic structures and elements of the algorithm while completing this task.
Hsu and Wang	*Applying game mechanics and student-generated questions to an online puzzle-based game learning system to promote algorithmic thinking skills*	2018	4th graders	There are 3 test items in which students give commands containing the basic algorithm structures and elements such as conditions, loops, variables, etc. so that they can make sure the aircraft object arrives at the right position.
Elshahawy et al.	*Codaroutine: A serious game for introducing sequential programming concepts to children with autism*	2020	14, 12, 8 7 year-olds	In a three-phased game related to daily life tasks, students are asked questions about the tasks. The students have to use the basic algorithm structures and elements while trying to find the answer.
Li et al.	*Socially shared regulation of learning in game-based collaborative learning environments promotes algorithmic thinking, learning participation and positive learning attitudes*	2020	Middle schoolers	Students attend a six-week Kodu Game Lab training. Next, they try to answer 5 multiple-choice items and 1 open-ended item regarding the software. The questions are targeted at concept understanding, algorithm creation and complex game design. Kodu Game Lab is a specific software for designing 3D games.

Oluk and Cakir	*The Effect of Code.Org Activities on Computational Thinking and Algorithm Development Skills*	2021	6th Graders	In order to examine the effect of the applications on the development of algorithm development skills of students, the algorithm development achievement test was applied as an evaluation tool. The test includes questions within the framework of learning algorithm logic, choosing the best algorithm and editing faulty algorithms.
Dag et al.	*The effect of an unplugged coding course on primary school students' improvement in their computational thinking skills*	2023	3rd Graders and 4th Graders	In order to examine the effect of computer-free coding courses on students' computational thinking skills, evaluation was carried out with multiple-choice tests. While computational thinking is evaluated by covering different dimensions, 3 items in the test are related to algorithm design.

When we look at the current literature, it is pointed out that the brain is a part of the body and it shows the importance of the body in learning. The embodied cognition theory emphasizes the inseparable link between the brain, body, and the world. Advocates of the theory claim that the brain must be understood in the context of its physical body whilst, reciprocally, the active body can alter the function of the brain. Implications of embodied cognition theory in education have become a significant part of contemporary teaching and learning practices, under the umbrella of embodied learning (Anderson, 2003; Ayala et al., 2013; Paloma, 2017; Wilson, 2002). Participating in games through physical activities is seen as an important way to ensure embodied learning (Altakrouri & Schrader 2012; Iacolina et al. 2010). It has been demonstrated that by integrating this path into the educational environment, students' motivation and desires increase, their active participation is ensured, and their cognitive and academic performances are positively affected (Kosmas et al., 2018; Kosmas & Zaphiris, 2023). Kosmas et al. (2018) evaluated the use of physically based games on 31 primary school students from the framework of embodied learning. The results showed that the games had a positive impact on students' short-term memory and emotional states. In context-based evaluation, which is one of the types of educational evaluation, evaluations are carried out by focusing on the contexts in individuals' lives (La Belle et al., 1979; Taasoobshirazi & Carr, 2008). Based on this, it is stated that in evaluations carried out by selecting the situations in individuals' lives as context, individuals' mastery and active participation in the process will increase and the validity of the evaluation results will be high (Bellochi et al., 2016; Fensham & Rennie, 2013). Traditional games which are one of the cultural elements, are played by children in a fun way in natural environments. In the course of time, children transferred many features of the traditional game into their lives (Sümbüllü & Altınışık, 2016; Yılmaz, 2020). Departing from this, the present study aimed to develop a test in line to assess the algorithmic thinking skills of middle school sixth-grade students.

2. METHOD

This is a research study conducted to develop a test to assess algorithmic thinking skills through traditional games. In line with this purpose, the research questions to be answered in the study are as follows:

1. How to scope a test to evaluate algorithmic thinking skills?
2. What is the validity and reliability of the test developed to evaluate algorithmic thinking skills?

3. How is the criterion score of the test developed to evaluate algorithmic thinking skills determined?

The development process was implemented in compliance with the steps suggested by Downing (2006). The model developed by Downing (2006) consists of 12 steps that progress gradually. While different tasks are performed at each step, there are relational tasks that affect each other between different steps. Detailed explanations about the steps are given below by headings.

2.1. Participants

Various groups of participants took part in the study until the test development was completed. For each stage of the process, a different sample selection method was used. For example, the participants involved in the validity and reliability checks were selected with purposeful sampling. This method allows for choosing the most eligible individuals or groups who share a similar experimental background concerning the project objectives (Yıldırım & Şimşek, 2006). The inclusion criterion was that 24 6th-grade students who participated in the development phase of algorithmic thinking skill test items and 252 6th-grade students who participated in the evaluation phase completed the "Problem Solving and Coding" unit of the Ministry of National Education 6th Grade Information Technologies and Software Course Curriculum. It was explained that the aim was to evaluate algorithmic thinking skills with the items or tests presented to the students at the beginning of both stages. To ensure the content validity of the test, opinions of experts in the fields of Information Technologies, Turkish Language Literature and Mathematics Education were taken. A Turkish Language and Literature expert examined the grammatical and semantic suitability of the test items. The Information Technologies and Mathematics Education expert evaluated the suitability of the test items with the curriculum outcomes. The objectives of the 6th Grade Information Technologies and Software Curriculum Unit 5 Problem Solving and Coding Unit and four of the objectives of the 6th Grade Mathematics Unit 1 Numbers and Operations overlap with each other. The numbers of all the participants and their descriptive features are given along with the particular stage of assignment they were assigned in Table 3.

Table 3. *Study participants by stages of test development.*

	Stage	No of participants	Descriptive Features
1	Determining the game pool and checking the grammaticality and semantics of the items	1	Expert in Turkish Language and Literature
2	Determining the relationship between traditional games and suggested ways to develop algorithmic thinking skills	4	Expert in Information Technologies
3	Determining the content validity of the test	3	Expert in Information Technologies and Mathematics
		24	Middle School Sixth-Grade Students
4	Performing validity and reliability studies of the test	252	Middle School Sixth Grade Students
5	Determining the passing score of the test	3	Information Technologies Teacher

2.2. Data Collection Tools

2.2.1. *Document*

Models proposed by researchers such as Szanto (2002), Garner (2003), Futschek (2006), Vascencolos (2007), and Committee on Logic Education for the development of algorithmic thinking skills were examined. In order to analyze the test items in detail, during the document creation process, a form containing the game pool and models for model selection was created and the opinions of 2 information technologies field experts were taken. Field experts expressed a common opinion about the suitability of the model proposed by Vascencolos. Therefore, the test items created within the scope of this research were divided into weekly worksheets in accordance with the algorithmic thinking skill teaching method suggested by Vascencolos (2007) and applied during the course, by consulting the experts. Question items were added, and the students were expected to answer the following questions: a) What exactly is the problem you are expected to solve? State the problem in your own words; b) Clarify the problem situation by drawing a picture or diagram of the problem; c) What are the known algorithmic concepts or elements? d) What is/are the unknown or requested situation(s)? e) Illustrate your solution with a row algorithm or flowchart; f) Test your solution and explain your answer. Thus, information was collected about the students' response status to the test items and their ability to identify and use the basic structure and elements of the algorithm in the items.

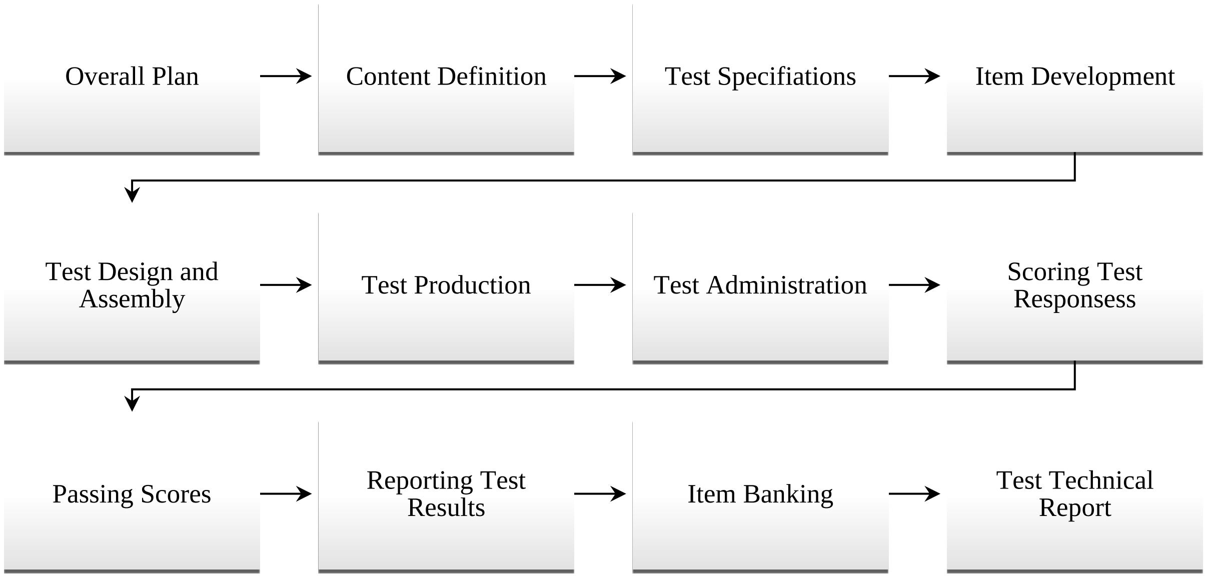
2.2.2. *Interview*

The interview is preferred in order to reveal the thoughts, attitudes, interests and beliefs of the source person or group for the research (Ocak, 2010). In this research, an unstructured interview, one of the interview types, was used. Unstructured interview: It enables the researcher to obtain detailed information about the situation by constantly asking different questions according to the answers he receives from the interviewee. (Gall et al., 1996; Türnüklü, 2000). Students and science experts were interviewed as needed at certain stages of the study. Sample questions are as follows: Can you give examples of traditional games that can be used in relation to algorithmic thinking skills during the research process for "Determining the Game Pool and Grammatics and Semantics of the Items" for Turkish Language Literature Field Expert? Are there any grammatical or semantic errors in the test items? How? To the Information Technologies Field Expert for "Determining the content validity", which of the basic structure and elements of the algorithm do you think are included in the relevant article? Information showing which topics were carried out and with whom the interview process was carried out is given in Table 3 above.

2.3. The Procedure

The test development process was based on Downing's (2006) test development model. The schematic depiction of the model is shown in Figure 1.

Figure 1. *Test development procedure (Downing, 2006).*



2.3.1. *Overall plan*

According to Downing (2006), the most crucial step of test development is to clearly state the purpose of the test in question. The purpose of the current test is to assess algorithmic thinking skills. The test items were planned in multiple-choice format to be applied in pen and paper style.

2.3.2. *Content definition*

In this step, the content of the test was described. The scope of the content was determined by including the basic algorithm structures and elements such as conditions, loops, operators, constants, and variables. These concepts were integrated within the framework of traditional games and game rules later in the procedure. A game pool was created, and question items were written. The games that make up the scope of the test are presented in Table 4.

Table 4. *Games covered in the test.*

Game Names			
Jump-rope	Hide-and-Seek	Hopscotch	Know and Keep Your Name
Relay	Day-and-Night	Musical Chairs	Puss-in-the-corner
Pounding Nails	Dodgeball	Handkerchief Grabber	Stop!
Word Derivation	Five Stones	Boom!	41 Sticks
Shopkeeper	Tawing	Captive	Grasshopper

2.3.3. *Test specifications*

In this step, the preferred test format, the number of items, the scoring rules, and the time limit are determined. It is recommended that the multiple-choice test type be preferred in cases of assessment at the lower levels of the cognitive field (Güler, 2015). This test type has two major advantages. Easy and objective scoring supports reliability, and the opportunity to measure a large number of outcomes ensures higher validity. It is also convenient as it can be applied to large audiences and it allows for rapid and precise assessment (Başol, 2019). Due to these strengths, a multiple-choice test format was selected for the tool in the study. The assessment tool was first drafted with a total of 27 items, 2 of which were sample items. In order to estimate the test duration, a pilot application was conducted on a group with similar characteristics as the sample group.

2.3.4. *Item development*

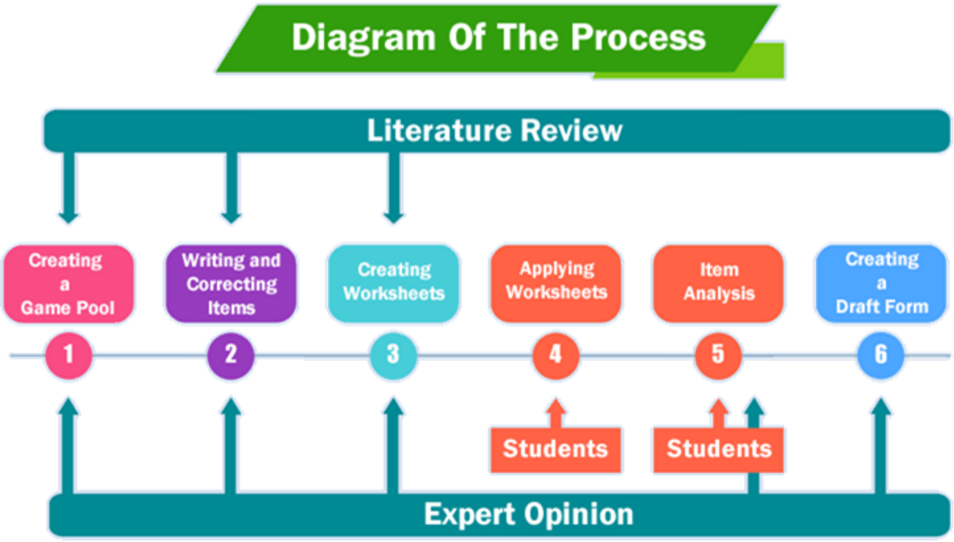
The actions undertaken at this stage are shown in Figure 2. Downing (2006) suggests writing items in the specified number and type and then having them reviewed by field experts. The experts' feedback is used to revise the items. In this study, the game was picked from the pool of games, and items were developed by taking a snapshot of the game played by certain characters within the framework of the game rules. The initial items consisted of three options "yes", "no" and "maybe". The items were reviewed by three experts. It was suggested by the experts to write options more relevant to the items. Thus, the items were revised accordingly. Worksheets were drawn up with the items and they were applied to seven students at the sixth-grade level for 7 weeks. It was checked whether the question items put the students through the steps in Vascencolos's model (2007). The worksheets were collected from the students, then they were reviewed, and matrices were created by two experts independently. The agreement rate between the experts was found 0.90 with the reliability formula of Miles and Huberman (1994). According to the reliability formula of Miles & Huberman (1994) used in the analysis of qualitative data, the fact that the result of Reliability = Consensus / (Consensus + Disagreement) above 0.80 indicates that the reliability level is high. The findings which were reached through consensus are presented in Table 5. It was seen that the students were able to determine the basic algorithm structures and elements in the games and transfer them to the line flow charts, and as a result, the question items were proved to be compatible with the model of Vascencolos (2007).

Table 5. *Results of test item check against Vascencolos's (2007) model.*

	Participant						
Theme	P1	P2	P3	P4	P5	P6	P7
Read and Comprehend the Problem Statement	7	7	7	7	7	7	7
Select Theoretical Concepts That May Be Applied	7	6	7	7	7	6	6
Qualitative Description of the Problem	7	6	7	7	6	5	6
Formulization of a Solution Strategy	6	6	7	6	7	6	7
Test and Description of the Solution	7	7	7	7	7	7	7

At the end of this stage, a discussion was held with the participant group regarding the items and the number of options, and it was decided to increase the number of options to 4.

Figure 2. *Diagram of item development process.*



2.3.5. *Test design and assembly*

According to Downing (2006), the items are included in the test form in a way that minimizes the cognitive load on the respondents. The practices and the data obtained up to this stage were discussed with two experts. As a result, the draft form was obtained. The suitability of the items in the draft form and the algorithmic structures and elements embedded in the items were reviewed by three experts. The test was finalized accordingly. A sample item is shown below. Once the draft form was ready, it was given to 14 students in the sixth grade in order to appraise the time needed to answer the test items and to make a preliminary assessment of the items before starting the pilot application. Based on this trial, it was decided to fix the test duration to 40 minutes.

Sample Item: Five players stand around a circle. Ali starts counting by saying "One!" and counting continues clockwise one by one. The players corresponding to 5 and its multiples have to shout “Boom!” instead of “Five” or so on. Those who fail to do so are eliminated and the counting continues from where it left off. When it is Ali's turn for the third time, one person has been eliminated from the game.

In this case, which of the following is true?

A) Ali has said “Boom!” once.
B) Ali has said “Boom!” twice.
C) Ali has said “Boom!” three times.
D) Ali has never said “Boom!”.

2.3.6. *Test production*

At this stage, the algorithmic thinking skill assessment test was published. The printout was checked carefully for typos.

2.3.7. *Test administration*

To make sure that the physical conditions were equal for all students during the application of the test, the students were given the test in their classrooms, and they were supervised by their teachers. Application permission from the Directorate of National Education, ethics committee report from Trabzon University and informed consent form from student parents were obtained for students to participate in the process. The advisors who supported the test in different classes during the evaluation process of the test were asked to inform the students that the exam would be completed within 1 class hour (40 minutes), that no course scores of the students would change with the test results, and that the purpose of the test was to measure the algorithmic thinking skills of the students. Thus, it is aimed to help students easily transfer their existing knowledge to paper without worrying emotionally during the evaluation phase. Similar information was given to the students during the application of the test, in which the researcher participated as an evaluator.

2.3.8. *Scoring test responses*

Since the 24 6th-grade students who participated in the development phase of the algorithmic thinking skill test items and the 256 6th-grade students who participated in the evaluation phase were in different schools in different provinces, there was no interaction between them. During the item development phase and evaluation process of the test, all students completed the evaluation simultaneously within one class hour (40 minutes). However, the only difference is that the evaluation for the item development phase of the test is made at a time before the test evaluation process. The students’ responses in the test were coded as 0 or 1 in an MS Excel sheet by the researcher depending on the meaning of the responses. Then, the item difficulty indexes and item discrimination indexes were calculated, and a transaction was performed to determine the items to be included in the final test. Additionally, the KR-20 reliability calculation method was used to compute the reliability of the test.

2.3.9. *Passing scores*

In Downing’s model (2006), the passing score of a test is calculated according to the absolute or relative criteria rules. In this study, the opinions of three field experts were asked and it was decided to determine the score value of each item separately in calculating the passing score of the test. It was advised by the experts to carry out two actions in determining the scores of the items. The first thing to do is to classify the items by difficulty level based on the item difficulty index value. In the second step, points are appointed to the basic algorithm structures and elements in the item by judging the importance of those elements and the structures for the solution of the problem. According to Başol (2019), item difficulty index values between 0.85 and 1.00 refer to very easy items, those between 0.61 and 0.84 are considered easy, items ranging from 0.40 to 0.60 are considered medium difficulty, difficulty index between 0.16 and 0.39 refer to difficult items, and the values between 0.00 and 0.15 indicate very difficult items. According to the item difficulty index calculation, 16 of the items were found at medium difficulty and six items were easy. The difficulty levels of the items and their classification according to the elements they cover are shown in Table 6 (After the item statistics were made, three test items were removed and item numbers were recoded between 1-22.).

Table 6. *Test items by difficulty level and contained elements.*

Item Difficulty Level	Algorithm Structures and Elements Contained in the Item	No of Items	Item No
Medium Difficulty	Condition, Loop, Constant, Variable, Operator	12	4,7,8,9,10,14,15,18,19,20,21,22
	Condition, Loop, Variable, Operator	2	1,5
	Condition, Constant, Variable, Operator	2	2,17
Easy	Condition, Loop, Constant, Variable, Operator	1	12
	Condition, Constant, Variable, Operator	2	11,16
	Condition, Loop, Operator	1	13
	Condition, Variable, Operator	2	3,6

Definitions: Condition: Probability situation, Loop: Recurring situation, Constant: Unchanged in value, Variable: Changed value, Operator: Mathematical operations.

The algorithm structures and elements contained in each item were evaluated relative to each other. According to the experts in the study, condition and loop were of equal importance and more important than the other elements of the algorithm. The remaining elements, which are constant, variable and operator, were attached to the same level of importance. The scores of the items were calculated by taking into consideration the scores given by the experts for the basic structures and elements of the algorithm.

The scores for the basic algorithm structures and elements embedded in the items of medium difficulty are given in Table 7. The scores of the items of medium difficulty varied between 3.5 and 5.5 points depending on the basic structures and elements of the algorithm contained in the respective items.

Table 7. *Basic algorithm structures and element scores in items of medium difficulty*.

Basic Algorithm Structure and Element	Score
Condition	2
Loop	2
Constant	0.5
Variable	0.5
Operator	0.5

The scores for the basic structures and elements of the algorithm covered in easy items are presented in Table 8. It can be seen that the scores of the easy-level items range from 2.75 to 4 points, depending on the structures and elements contained in the respective items.

Table 8. *Basic algorithm structure and element scores in easy items.*

Basic Algorithm Structure and Element	Score
Condition	1.25
Loop	1.25
Constant	0.5
Variable	0.5
Operator	0.5

A student who answers all the test items correctly gets 100 points. Table 9 gives details about the scores of the test items.

Table 9. *Scores of test items.*

Items of Medium Difficulty		
No of Items	Item No	Item Score
12	4,7,8,9,10,14,15,18,19,20,21,22	5.5
2	1,5	5
2	2,17	3.5
Easy Items		
1	12	4
1	13	3
2	11,16	2.75
2	3,6	2.25

The Angoff method was preferred to determine the cut-off score of the test. It is the most widely used method in determining the cut-off score for tests (Demir & Köse, 2014). In the Angoff method, experts analyze the test items one by one and estimate the correct answer rate for each item for 100 students. Then, the average of the experts' estimates for each item is calculated to set the minimum passing score for the items separately. Finally, the passing score of the test is determined by taking the average of the minimum passing scores of all items. In this study, the estimates were made by three experts, and they are shown in Table 10 below. As instructed in the method above, the minimum passing scores for all of the items were determined by taking the average of the estimates reported by the experts for the items in the first place. Based on the average of the minimum passing scores, the passing score of the algorithmic thinking skill assessment test was set as 67.23.

Table 10. *Angoff method results for the algorithmic thinking skill test.*

Item No	Expert 1	Expert 2	Expert 3	Minimum Passing Score
Item 1	50	60	70	60
Item 2	60	40	50	50
Item 3	80	75	80	78.3
Item 4	75	65	70	70
Item 5	55	65	60	60
Item 6	72	64	80	72
Item 7	76	70	60	68.6
Item 8	55	60	75	63.3
Item 9	50	55	65	56.6
Item 10	80	75	75	76.6
Item 11	60	70	75	68.3
Item 12	85	80	85	83.3
Item 13	70	80	85	77.5
Item 14	40	55	65	78.3
Item 15	45	50	50	48.3
Item 16	60	60	65	61.6
Item 17	65	50	70	61.6
Item 18	55	55	65	58.3
Item 19	70	55	65	63.3
Item 20	75	70	80	78.3
Item 21	70	65	75	70
Item 22	70	80	75	75
Final Test Cut-off Score				67.23

2.3.10. *Reporting test results*

Downing (2006) states that students who take a test have the right to receive a report on their test performance. It is considered important to give a feedback report on student mistakes in clear and understandable language. For this reason, a report was created after assessing the student responses.

2.3.11. *Item banking*

According to Downing (2006), it is essential to safely store the items that are regarded as effective in evaluations in case they are needed for developing a new test form or developing a different version of the test. Therefore, the test items were saved in a safe platform.

2.3.12. *Test technical report*

Downing (2006) points out that all data regarding test development activities must be reported in complete. Again, the entire development process along with the findings concerning test reliability and validity are elaborated here.

3. FINDINGS

3.1. Validity and Reliability Studies

Test development becomes final upon the collection of data for validity and reliability. Validity means the determination of the extent to which the assessment tool measures the intended construct without confusing it with other features (Büyüköztürk et al., 2020). Reliability is defined as the quality of the assessment tool being free from random errors (Baykul et al., 2003; Güler, 2015). In this study, the results of the validity and reliability analyses of the test are presented under the following headings.

3.1.1. *Content validity*

Content validity relates to the extent the test items are competent in measuring the behaviors intended to be measured. The number and quality of test items are important to create a test with high content validity. For content validity, expert opinion is frequently sought regarding the suitability and ability of the item to measure the intent situation. A table of specifications is created in light of the expert feedback (Büyüköztürk et al., 2020). In this study, 4 Information Technologies experts examined the 25 items of the test and produced the table. The specification table revealed that among the basic algorithm concepts, the condition was available in 24 items, the loop was in 16 items, the constant in 19 items, the variable in 25 items, and the operator was available in 24 items.

3.1.2. *Construct validity*

Construct validity concerns the adequacy of the test scores in measuring the construct targeted to be measured by the test (Büyüköztürk et al., 2020). To check construct validity, the items were analyzed according to the lower-upper group differences. The upper and lower groups were identified by applying the "27% rule" offered by Kelley (1939). In this scope, first of all, the total scores obtained from the test were calculated and the students were ranked from the highest to the lowest. The 27% extreme group that got the highest scores (68 people with the highest scores) was placed in the upper group, while the other 27% extreme group that got the lowest scores (68 people with the lowest scores) was named as the lower group. Finally, an independent sample t-test was applied on the difference between the upper and lower groups of the items. A significant difference between the groups shows that the items have enough discriminatory capacity to tell the proficient from the nonproficient students in terms of algorithmic thinking (Yıldırım & Şimşek, 2006). The data obtained from the independent sample t-test performed on the final test are given in Table 11. The results revealed that the items were significant (p<0.01) in distinguishing the students in the lower group from those in the upper group.

Table 11. *Independent sample t-test results based on lower-upper group difference.*

Item No	27% Lower-Upper Group Difference p Value
Item 1	0.000**
Item 3	0.000**
Item 4	0.000**
Item 5	0.000**
Item 6	0.000**
Item 7	0.000**
Item 8	0.000**
Item 9	0.000**
Item 11	0.000**
Item 12	0.000**
Item 13	0.000**
Item 14	0.000**
Item 15	0.000**
Item 16	0.000**
Item 17	0.000**
Item 19	0.000**
Item 20	0.000**
Item 21	0.000**
Item 22	0.000**
Item 23	0.000**
Item 24	0.000**
Item 25	0.000**

$^{**}p$<0.01

3.2. Reliability Analysis

3.2.1. *Item discrimination index*

Item discrimination index expresses the capacity of each test item to distinguish a high performer from a low performer. In other words, it is answering an item correctly by a high-achieving student whereas being answered incorrectly by a low-achieving student. The item discrimination index is calculated by subtracting the number of respondents with correct answers in the upper group from the number of those in the lower group and then dividing the result by half of the whole group. Item discrimination index value varies between -1 and 1 (Bayrakçeken, 2015).

Item discrimination index value (rjx);

rjx>=0.40 signifies a very good item,
0.30<=rjx>=0.39 signifies a good item that can be kept in scale without amendment,
0.20<=rjx>=0.29 signifies an item that needs correction and improvement,
0.19<rjx signifies an item that ought to be omitted (Büyüköztürk et al., 2020). The item discrimination index results obtained for the test in this study are given in Table 12. It was found that Items No 2, 10 and 18 were not good items so they were removed from the scale.

3.2.2. *Item difficulty index*

Item difficulty index indicates the correct answer rate for each item in an assessment tool. It is calculated by summing up the respondents in both the upper and lower group providing a correct answer for a given item and then finding the ratio of this sum to the whole group. Item difficulty index can take a value between 0 and 1. A value close to 0 marks a difficult item, while values close to 1 signal easy items (Bayrakçeken, 2015). The item difficulty index results for the test in this study are given in Table 12. The average item difficulty index was found to be 0.52, which implies that the items in the assessment tool were of medium difficulty.

Table 12. *Item difficulty index and item discrimination index results for the test.*

Item No	Item Difficulty Index	Item Discrimination Index
Item 1	0.59	0.42
Item 2	**0.13**	**0.10**
Item 3	0.40	0.42
Item 4	0.69	0.55
Item 5	0.44	0.33
Item 6	0.52	0.44
Item 7	0.75	0.30
Item 8	0.45	0.32
Item 9	0.52	0.55
Item 10	**0.16**	**0**
Item 11	0.40	0.39
Item 12	0.50	0.67
Item 13	0.63	0.51
Item 14	0.73	0.47
Item 15	0.78	0.36
Item 16	0.40	0.54
Item 17	0.40	0.60
Item 18	**0.22**	**0.17**
Item 19	0.61	0.64
Item 20	0.46	0.51
Item 21	0.41	0.50
Item 22	0.47	0.48
Item 23	0.51	0.55
Item 24	0.45	0.58
Item 25	0.40	0.51
Item Difficulty Index Mean: 0.52, Discriminatory Index Mean: 0.48		

3.2.3. *KR-20 reliability analysis*

The KR-20 formula is applicable in cases where the responses to test items are scored as 0 (wrong) or 1 (correct) (Büyüköztürk et al., 2020). As the KR-20 value approaches 1, it is assumed that the internal consistency increases and the test is a homogeneous tool that measures similar features, corresponding to higher levels of reliability. By contrast, the test is regarded to have low reliability as the KR-20 value approaches 0 (Çetin, 2019). The KR-20 result for the current test is shown in Table 13. As can be understood from the table, the reliability of the test was high (KR-20= 0.836> 0.70).

Table 13. *KR-20 reliability analysis result.*

Test	KR-20	No of Items
Algorithmic Thinking Skill Assessment Test	0.836	22

4. DISCUSSION and CONCLUSION

In this study, an algorithmic thinking skill assessment test was developed through contextualization with traditional games. The test ultimately consisted of 20 conventional children's games for answering 22 relevant items. Either the rules of the games were explained or snapshot situations were taken from the games, and scenarios were invented accordingly. Problem situations were given along with correct and incorrect options. Consequently, a measurement tool was designed whereby students' algorithmic thinking skills were measured in the context of common traditional games. Previous studies assessing algorithmic thinking skills through games were also investigated. To begin with, in Gürbüz et al. (2012), the students were provided with the knowns like the sun, temperature, humidity and wind and they were asked to prepare a weather forecast in reference to the given facts and figures. In the study of Zhao and Shute (2019), there was a character in the game and the students were supposed to bring it to the target point by giving the right commands. Similarly, in the study by Czakoova (2020), a three-level game was presented and the students were expected to animate the character and thus collect points by giving the right commands during the first two levels. In the third level, the character must reach the target most effectively and shortly. Again, in Kazimoglu's (2020) study, there was a robot and the students were instructed to make it to the target in the most effective way by means of giving commands. Another example was a study by Chen and Chi (2020). Two groups of students acted as pirate gangs trying to seize the treasure and they had to develop strategies to achieve their goal.

With the collaboration of the relevant field experts, it was understood that the majority of the test items included condition, loop, constant, and variable and operator structures of the algorithm. Some of the items contained three of the algorithm structures and elements and some others contained four of them. Moreover, the study data demonstrated that "condition" became the most frequently used algorithm element evident in 22 items, but "loop" was the least seen element with 16 appearances. Thanks to this property, this new assessment test of algorithmic thinking skills seems to be an outstanding one in the literature. In contrast to the current study, the previous studies generated algorithmic thinking skill assessment tools that address only one or two of the basic algorithm structures and elements at one time (Hsu & Wang, 2018; Tsukamoto et al., 2017). For instance, Tsukamoto et al. (2017) developed a measurement instrument for primary school students' algorithmic thinking skills. It was built around sequential operations, conditional branches, and iterative operations. There were three items in the tool, each intending to measure one specific concept. Likewise, Hsu and Wang (2018) introduced an achievement test to assess algorithmic thinking skills, and the item added for debugging targeted conditionals as an algorithm structure. Another item assigned a task that entails using both condition and loop among basic algorithm structures and elements.

The resulting test is a valid and highly reliable instrument contextualized in the most common local children's games. According to Yılmaz (2020), when all the steps in a traditional game process are sequentially used and associated with a concept chosen from any discipline, students will have unconsciously learned the target concept while completing the game task in order to achieve success. Hence, it seems beneficial to integrate traditional games, which have survived as cultural heritage to the present day, into the education process.

In the study, content validation of the test was performed by taking expert opinion. A team of eight experts provided support for this particular job. Of the experts, seven were specialists of computer and instructional technology education and they were responsible for the content and applicability of the items. The other expert was a specialist in Turkish Language and Literature for checking the spelling and grammar rules in the texts. This step was in congruence with the literature as there were other studies seeking expert opinion to ensure the content validity of the test (Kocagül-Sağlam & Ünal-Çoban, 2018). As for construct validity, several methods were found in the literature, but one of the most extensively used ones was to perform an independent sample t-test on the difference between the upper and lower groups, as conducted in the present case (Özden & Yenice, 2021). Thus, both content and construct validation methods used in this study are in agreement with the literature.

When it comes to determining the difficulty indexes of the items in the algorithmic thinking skill assessment test, the values were noted between 0.40 and 0.78, resulting in the average difficulty index equal to 0.52 for the entire test. These figures reveal that the algorithmic thinking skill assessment test developed here is difficult at the intermediate level. Furthermore, the discrimination indexes of the individual items varied between 0.30 and 0.67, and the average discrimination index value of the test was found to be 0.48. In general, it is desirable to have a measurement tool with high validity and reliability. The meeting of these criteria is checked by looking at test results, particularly at the item discrimination index and item difficulty index values of each item (Güler, 2015). Although the item difficulty index values are not equal to 0.50, it is favourable to have the average test difficulty close to 0.50 for a successful test development process (Bayrakçeken, 2015). Discrimination index values between 0.30 and 0.40 classify good items, values greater than 0.40 classify very good items, and values lower than 0.20 imply that the items are too weak to remain on the scale (Özçelik, 2013). These threshold values support the quality of the test in that it consists of only very good and good items and it is a highly discriminatory instrument as a whole.

Lastly, the KR-20 coefficient was 0,83 which is far greater than the acceptable lower limit of 0.70. It is thus obvious that the reliability of the algorithmic thinking skills assessment test is at a satisfactory level (Büyüköztürk et al., 2020). The Kuder-Richardson (KR-20) reliability determination method is used to examine the internal consistency between test scores, in which responses to test items are computed as 0 (false) and 1 point (correct). This study is congruent with the related literature since the KR-20 reliability analysis was used for test reliability in most studies (Karatay & Doğan, 2016; Özden & Yenice, 2021; Şardağ & Kocakülah, 2016). Like many other aspects, the reliability calculation method used in this study is also similar to the previous studies in the literature.

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**: Trabzon University, 13.08.2021, E-81614018-000-704.

Contribution of Authors

The authors contributed equally to all the stages of the study.

Orcid

Emre Zengin https://orcid.org/0000-0002-6643-7550
Yasemin Karal https://orcid.org/0000-0003-4744-4541

REFERENCES

Altakrouri, B., & Schrader, A. (2012, September). Towards dynamic natural interaction ensembles. In *the 26th BCS Conference on Human Computer Interaction 26* (pp. 1-10). http://dx.doi.org/10.14236/ewic/HCI2012.0

Anderson, M.L. (2003). Embodied cognition: A field guide. *Artificial intelligence*, *149*(1), 91-130. https://doi.org/10.1016/S0004-3702(03)00054-7

Apostolellis, P., Stewart, M., Frisina, C., & Kafura, D. (2014). RaBit EscAPE: A board game for computational thinking. In *Proceedings of the 2014 conference on Interaction design and children* (pp. 349-352). http://dx.doi.org/10.1145/2593968.2610489

Ayala, N.A.R., Mendívil, E.G., Salinas, P., & Rios, H. (2013). Kinesthetic learning applied to mathematics using kinect. *Procedia Computer Science*, *25*, 131-135. https://doi.org/10.1016/j.procs.2013.11.016

Aytekin, A., Çakır, F.S., Yücel, Y.B., & Kulaözü, İ. (2018). The place and importance of algorithms in our lives. *Eurasian Journal of Social and Economic Studies*, *5*(7), 143-150. https://dergipark.org.tr/tr/pub/asead/issue/41013/495619

Başol, G. (2019). Measurement and evaluation in education. *Pegem Citation Index*, 001-307.

Baykul, Y., Gelbal, S., & Kelecioğlu, H. (2003). *Measurement and evaluation in education for Anatolian teacher high schools*. National Education Printing House.

Bayrakçeken, S. (2015). Test development. E. Karip (Ed.), *In Measurement and Evaluation* (s. 292-322). Pegem Academy.

Bellocchi, A., King, D.T., & Ritchie, S.M. (2016). Context-based assessment: Creating opportunities for resonance between classroom fields and societal fields. *International Journal of Science Education*, *38*(8), 1304-1342. https://doi.org/10.1080/09500693.2016.1189107

Borkulo, S., Chytas, C., Drijvers, P., Barendsen, E., & Tolboom, J. (2021). Computational thinking in the mathematics classroom: Fostering algorithmic thinking and generalization skills using dynamic mathematics software. In *The 16th Workshop in Primary and Secondary Computing Education* (pp. 1-9). https://doi.org/10.1145/3481312.3481319.

Brown, W. (2015). *Introduction to algorithmic thinking.* https://raptor.martincarlisle.com/Introduction%20to%20Algorithmic%20Thinking.doc

Büyüköztürk, Ş., Kılıç-Çakmak, E., Akgün, Ö.E., Karadeniz, Ş., & Demirel, F. (2020). *Scientific research methods*. Pegem Publications.

Chen, K.Z., & Chi, H.H. (2022). Novice young board-game players' experience about computational thinking. *Interactive Learning Environments*, *30*(8), 1375-1387. https://doi.org/10.1080/10494820.2020.1722712

Chuechote, S., Nokkaew, A., Phongsasithorn, A., & Laosinchai, P. (2020). A neo-piagetian analysis of algorithmic thinking development through the "sorted" digital game. *Contemporary Educational Technology*, *12*(1), 1-15. http://dx.doi.org/10.30935/cet.685959

Czakóová, K. (2020). *Developing algorithmic thinking by educational computer games*. Paper presented at the Conference eLearning and Software for Educationtional, Romania. http://dx.doi.org/10.12753/2066-026X-20-003

Czakóová, K., & Udvaros, J. (2021). Applications and games for the development of algorithmic thinking in favor of experiential learning. In *EDULEARN21 Proceedings* (pp. 6873-6879). IATED. https://doi.org/10.21125/edulearn.2021.1389

Çetin, B. (Ed.) (2019). *Measurement and evaluation in education*. Anı Publishing.

Debabi, W., & Bensebaa, T. (2016). Using serious game to enhance learning and teaching algorithmic. *Journal of e-Learning and Knowledge Society*, *12*(2). http://dx.doi.org/10.20368/1971-8829/1125

Demir, O., & Köse, İ.A. (2014). Comparison of cutoff scores determined by Angoff, Nedelsky and Ebel standard setting methods. *Journal of Mersin University Faculty of Education*, *10*(2), 14-27. https://dergipark.org.tr/en/pub/mersinefd/issue/17394/181823?publisher=mersin

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. https://doi.org/10.1007/s40692-017-0090-9

Downing, S.M. (2006). Twelve steps for effective test development. *Handbook of test development*, *3*, 25. https://doi.org/10.4324/9780203874776.ch1

Elshahawy, M., Aboelnaga, K., & Sharaf, N. (2020). Codaroutine: A serious game for introducing sequential programming concepts to children with autism. *In 2020 IEEE Global Engineering Education Conference (EDUCON)* (pp. 1862-1867). IEEE. http://dx.doi.org/10.1109/EDUCON45650.2020.9125196

Erümit, K.A., Karal, H., Şahin, G., Aksoy, D.A., Aksoy, A., & Benzer, A.İ. (2018). A model proposal for teaching programming: programming in seven steps. *Education and Science*, *44*(197), 1-29. http://dx.doi.org/10.15390/EB.2018.7678

Evripidou, S., Amanatiadis, A., Christodoulou, K., & Chatzichristofis, S.A. (2021). Introducing algorithmic thinking and sequencing using tangible robots. *IEEE Transactions on Learning Technologies*, *14*(1), 93-105. https://doi.org/10.1109/TLT.2021.3058060

Fensham, P.J., & Rennie, L.J. (2013). Towards an authentically assessed science curriculum. In *Valuing assessment in science education: Pedagogy, curriculum, policy* (pp. 69-100). Springer.

Figueiredo, M., Amante, S., Gomes, H.M.D.S.V., Gomes, M.A., Rego, B., Alves, V., & Duarte, R.P. (2021). Algorithmic thinking in early childhood education: Opportunities and supports in the Portuguese context. In *EDULEARN21 Proceedings* (pp. 9339-9348). IATED. https://doi.org/10.22521/edupij.2022.112.3

Futschek, G., & Moschitz, J. (2010). Developing algorithmic thinking by inventing and playing algorithms. *Proceedings of the 2010 constructionist approaches to creative learning, thinking and education: Lessons for the 21st century (constructionism 2010)*, 1-10.

Gall, M.D., Barg, W.R., & Gall, j.P. (1996). Educational research: an introduction (6th ed.). Longman Publishing.

Geisinger, K.F. (2016). 21st century skills: What are they and how do we assess them?. *Applied Measurement in Education*, *29*(4), 245-249. https://doi.org/10.1080/08957347.2016.1209207

Güler, N. (2015). *Measurement and evaluation in education* (7th ed.). Pegem Academy Publishing.

Gürbüz, H., Evlioğlu, B., Erol, Ç.S., Gülseçen, H., & Gülseçen, S. (2017). "What's the weather like today?": A computer game to develop algorithmic thinking and problem solving skills of primary school pupils. *Education and Information Technologies*, *22*(3), 1133-1147. https://doi.org/10.1007/s10639-016-9478-9

Hsu, C.C., & Wang, T.I. (2018). Applying game mechanics and student-generated questions to an online puzzle-based game learning system to promote algorithmic thinking skills. *Computers & Education*, *121*, 73-88. http://dx.doi.org/10.1016/j.compedu.2018.02.002

Hu, C. (2011). Computational thinking: What it might mean and what we might do about it. In Proceedings of the 16th annual joint conference on innovation and technology in computer science education (pp. 223-227). ACM. http://dx.doi.org/10.1145/1999747.1999811

Huizinga, J. (1955). Homo ludens: A study of the play element in culture. The Beacon Press. https://doi.org/10.2307/2087716

Iacolina, S.A., Lai, A., Soro, A., & Scateni, R. (2010). Natural interaction and computer graphics applications. In *Eurographics italian chapter conference* (pp. 141–146).

Jančec, L., & Vujičić, L. (2021). Project "Algorithmic Thinking Skills through Play-Based Learning for Future's Code Literates". In *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)* (pp. 641-644). IEEE. https://doi.org/10.23919/MIPRO52101.2021.9597078

Kanaki, K., & Kalogiannakis, M. (2022). Assessing algorithmic thinking skills in relation to gender in early childhood. *Educational Process: International Journal*, *11*, 44-59. https://dx.doi.org/10.22521/edupij.2022.112.3

Karakasis, C., & Xinogalos, S. (2020). BlocklyScript: Design and pilot evaluation of an RPG platform game for cultivating computational thinking skills to young students. *Informatics in Education*, *19*(4), 641-668. https://doi.org/10.15388/infedu.2020.28

Karatay, R., & Doğan, F. (2016). Development of science process skill scale of 7th grade secondary school students. *Dicle Üniversitesi Ziya Gökalp Eğitim Fakültesi Dergisi*, *27*, 1-8. http://dx.doi.org/10.14582/DUZGEF.548

Kátai, Z. (2015). The challenge of promoting algorithmic thinking of both sciences-and humanities-oriented learners. *Journal of Computer Assisted Learning*, *31*(4), 287-299. https://doi.org/10.1111/jcal.12070

Kazimoglu, C. (2020). Enhancing confidence in using computational thinking skills via playing a serious game: A case study to increase motivation in learning computer programming. *IEEE Access*, *8*, 221831-221851. http://dx.doi.org/10.1109/ACCESS.2020.3043278

Keane, T. (2012). Leading with technology: 21st century skills= 3Rs+ 4Cs. *Australian Educational Leader*, *34*(2), 44. https://search.informit.org/doi/10.3316/informit.895597893475453

Khanlari, A. (2013). Effects of robotics on 21st century skills. *European Scientific Journal*, *9*(27). https://doi.org/10.19044/esj.2013.v9n27p%25p

Kim, B., Park, H., & Baek, Y. (2009). Not just fun, but serious strategies: Using meta-cognitive strategies in game-based learning. *Computers & Education*, *52*(4), 800-810. https://doi.org/10.1016/j.compedu.2008.12.004

Kiss, G., & Arki, Z. (2017). The influence of game-based programming education on the algorithmic thinking. *Procedia-Social and Behavioral Sciences*, *237*, 613-617. http://dx.doi.org/10.1016/j.sbspro.2017.02.020

Kocagül-Sağlam, M. & Ünal-Çoban, G. (2019). Developing the reasoning skills test for science teachers and teacher candidates. *Elementary Education Online*, *17*(3). https://doi.org/10.17051/ilkonline.2018.466374

Kosmas, P., Ioannou, A., & Retalis, S. (2018). Moving bodies to moving minds: A study of the use of motion-based games in special education. *TechTrends*, *62*, 594-601. https://doi.org/10.1007/s11528-018-0294-5

Kosmas, P., & Zaphiris, P. (2023). Improving students' learning performance through Technology-Enhanced Embodied Learning: A four-year investigation in classrooms. *Education and Information Technologies*, 1-24. https://doi.org/10.1007/s10639-022-11466-x

La Belle, T.J., Moll, L.C., & Weisner, T.S. (1979). Context-based educational evaluation: A participant research strategy. *Educational Evaluation and Policy Analysis*, *1*(3), 85-94. https://doi.org/10.2307/1164160

Lee, T.Y., Mauriello, M.L., Ahn, J., & Bederson, B.B. (2014). CTArcade: Computational thinking with games in school age children. *International Journal of Child-Computer Interaction*, *2*(1), 26-33. http://dx.doi.org/10.1016/j.ijcci.2014.06.003

Lertlapnon, T., Lueangrungudom, N., & Vittayakorn, S. (2022). Protobot: An Educational Game for Algorithmic Thinking. In *2022 14th International Conference on Information Technology and Electrical Engineering (ICITEE)* (pp. 79-84). IEEE. https://doi.org/10.1109/ICITEE56407.2022.9954081

Li, J., Lin, Y., Sun, M., & Shadiev, R. (2020). Socially shared regulation of learning in game-based collaborative learning environments promotes algorithmic thinking, learning participation and positive learning attitudes. *Interactive Learning Environments*, 1-12. https://doi.org/10.1080/10494820.2020.1857783

Lin, S.Y., Chien, S.Y., Hsiao, C.L., Hsia, C.H., & Chao, K.M. (2020). Enhancing computational thinking capability of preschool children by game-based smart toys. *Electronic Commerce Research and Applications*, 44, 101011. https://doi.org/10.1016/j.elerap.2020.101011

McFarlane, A., Sparrowhawk, A., & Heald, Y. (2002). *Report on the educational use of games*. TEEM (Teachers evaluating educational multimedia). Cambridge.

Mezak, J., & Papak, P.P. (2018). Learning scenarios and encouraging algorithmic thinking. In *2018 41st International Convention on Information and Communication Technology, Electronics and Microelectronics (MIPRO)* (pp. 0760-0765). IEEE. https://doi.org/10.23919/MIPRO.2018.8400141

Mumcu, H.Y., & Yıldız, S. (2018). The investigation of algorithmic thinking skills of 5th and 6th graders according to different variables. *MATDER Journal of Mathematics Education*, *3*(2), 18-26. https://dergipark.org.tr/en/pub/med/issue/41621/453450

Özçelik, D.A. (2013). *Assessment and Evaluation in Schools Teacher's Handbook*. Pegem Academy.

Özden, B., & Yenice, N. (2021). Developing a scientific inquiry skills test for secondary school 7th and 8th grade students. *Journal of Mersin University Faculty of Education*, *17*(1), 112-131. https://doi.org/10.17860/mersinefd.726360

Paloma, G.F. (Ed.). (2017). *Embodied Cognition. Theories and applications in education science*. Nova Science Publishers. https://hdl.handle.net/11386/4712878

Paspallis, N., Kasenides, N., & Piki, A. (2022). A Software Architecture for Developing Distributed Games that Teach Coding and Algorithmic Thinking. In *2022 IEEE 46th Annual Computers, Software, and Applications Conference (COMPSAC)* (pp. 101-110). IEEE. https://doi.org/10.1109/COMPSAC54236.2022.00023

Pivec, M., & Kearney, P. (2007). Games for learning and learning from games. *Informatica*, *31*(4).

Prensky, M. (2003). Digital game-based learning. *Computers in Entertainment (CIE)*, *1*(1), 21-21.

Sarı, U., Pektaş, H.M., Şen, Ö.F., & Çelik, H. (2022). Algorithmic thinking development through physical computing activities with Arduino in STEM education. *Education and Information Technologies*, 1-21. https://doi.org/10.1007/s10639-022-10893-0

Scharf, F., Winkler, T., & Herczeg, M. (2008). Tangicons: Algorithmic reasoning in a collaborative game for children in kindergarten and first class. Paper presented at the 7th International Conference on Interaction Design and Children, USA. http://dx.doi.org/10.1145/1463689.1463762

Shang, J., Ma, S., Hu, R., Pei, L., & Zhang, L. (2019). *Game-based learning in future school. In shaping future schools with digital technology*. Springer. http://dx.doi.org/10.1007/978-981-13-9439-3_8

Sungkaew, K., Lungban, P., & Lamhya, S. (2022). Game development software engineering: digital educational game promoting algorithmic thinking. *International Journal of Electrical and Computer Engineering (IJECE)*, *12*(5), 5393-5404. http://dx.doi.org/10.11591/ijece.v12i5.pp5393-5404

Şardağ, M., & Kocakülah, A. (2016). Developing a science process skills test for eighth grade students. *Journal of Sakarya University Faculty of Education*, *31*, 1-32. https://dergipark.org.tr/tr/pub/sakaefd/issue/24690/261073

Taasoobshirazi, G., & Carr, M. (2008). A review and critique of context-based physics instruction and assessment. *Educational Research Review*, *3*(2), 155-167. https://doi.org/10.1016/j.edurev.2008.01.002

Thomas, J.O., Odemwingie, O.C., Saunders, Q., & Watlerd, M. (2015). Understanding the difficulties African-American middle school girls face while enacting computational algorithmic thinking in the context of game design.

Thomas, J.O., Rankin, Y., Minor, R., & Sun, L. (2017). Exploring the difficulties African-American middle school girls face enacting computational algorithmic thinking over three years while designing games for social change. *Computer Supported Cooperative Work (CSCW)*, *26*(4), 389-421. https://doi.org/10.1007/s10606-017-9292-y

Tsukamoto, H., Oomori, Y., Nagumo, H., Takemura, Y., Monden, A., & Matsumoto, K.I. (2017). Evaluating algorithmic thinking ability of primary schoolchildren who learn computer programming. In *2017 IEEE Frontiers in Education Conference (FIE)* (pp. 1-8). IEEE. http://dx.doi.org/10.1109/FIE.2017.8190609

Turchi, T., Fogli, D., & Malizia, A. (2019). Fostering computational thinking through collaborative game-based learning. *Multimedia Tools and Applications*, *78*(10), 13649-13673. https://doi.org/10.1007/s11042-019-7229-9

Turnuklu, A. (2000). A qualitative research technique that can be used effectively in educational research: Interview. Educational management in theory and practice, *24*(24), 543-559.

Vasconcelos, J. (2007). *Basic Strategy for Algorithmic Problem Solving*. http://www.cs.jhu.edu/~jorgev/cs106/ProblemSolving.html

Wangenheim, C., Medeiros, G., Missfeldt Filho, R., Petri, G., da Cruz Pinheiro, F., Ferreira, M.N., & Hauck, J.C. (2019). SplashCode--A Board Game for Learning an Understanding of Algorithms in Middle School. *Informatics in Education*, *18*(2), 259-280. http://dx.doi.org/10.31235/osf.io/2qbnp

Wilson, M. (2002). Six views of embodied cognition. *Psychonomic bulletin & review*, *9*, 625-636. https://doi.org/10.3758/BF03196322

Yadav, A., Gretter, S., Good, J., McLean, T. (2017). *Computational Thinking in Teacher Education*. In: Rich, P., Hodges, C. (eds) Emerging Research, Practice, and Policy on Computational Thinking. Educational Communications and Technology: Issues and Innovations. Springer, Cham. https://doi.org/10.1007/978-3-319-52691-1_13

Yıldırım, A., & Şimşek, H. (2006). *Qualitative research methods in the social sciences*. Seçkin Publishing.

Yılmaz, E.A. (2020). *By the power of games: an introduction to the science of gamification*. Epsilon Publishing House. https://dergipark.org.tr/tr/pub/tk/issue/56680/784921

Zsakó, L., & Szlávi, P. (2012). ICT Competences: Algorithmic Thinking. *Acta Didactica Napocensia*, *5*(2), 49-58.