Assessing young Korean children’s computational thinking: A validation study of two measurements

Jihyun Sung

Abstract
Computational thinking (CT) in young children has recently gained attention. This study verified the applicability of the Korean version of the Bebras cards and TACTIC-KIBO in measuring CT among young children in South Korea. A total of 450 children responded to the Bebras cards, TACTIC-KIBO, and Early Numeracy tasks that were used for the following analyses. Item response theory analysis, confirmatory factor analysis, correlation analysis, and calculation of Cronbach’s alpha were conducted to examine the psychometric properties of the validity and reliability of the two measurements. The results showed that these two measurements are acceptable for assessing CT among young children, demonstrating good validity and reliability, despite limitations such as the weak factor loadings of some items and low internal consistency of subfactors. These two CT measurements were significantly and positively correlated with early mathematical ability. Thus, these two measurements are acceptable for assessing CT among young children with varying CT ability, as they present good psychometric properties of the overall scores even though they have low internal consistency of subfactors and slightly weak correlations between subfactors.

Keywords Computational thinking · Mathematical ability · Assessment · Validity · Factor structure

1 Introduction

With global interest in the digital revolution era, formed by the convergence of information and communication technologies, education is also an area where significant changes are expected. To cope with the current context, students are expected to enjoy the process of proactively solving problems using the 4Cs (creativity, critical thinking,
communication, and collaboration) and the 3Rs (reading, writing, and arithmetic) as 21st-century skills (The Partnership for 21st Century Skills, 2016). Additionally, it has become a priority for national education policies worldwide to improve integrated science, technology, engineering, and mathematics (STEM) or art-incorporated (STEAM) education and practices (Kelly & Knowles, 2016). Countries such as the United States, the United Kingdom, Germany, and South Korea have proposed educational policies that promote STEM education (Korea Ministry of Education, 2015; National MINT (STEAM) Forum, 2014; NGSS Lead States, 2013; STEM learning, 2018). The purpose of integrated STEM education provides opportunities for students to learn and use high-level thinking skills such as computational thinking (CT), problem-solving, engineering design, technological literacy, scientific inquiry, and communication skills (Kelley & Knowles, 2016; Kennedy & Odell, 2014; Stohlmann et al., 2012).

CT refers to the ability to think analytically about problems, find solutions to problems, and disseminate results (Wing, 2006, 2011). In recent years, CT has incorporated skills related to the computer science field and cognitive abilities that can be applied to any domain, such as literacy, art, engineering, mathematics, and science (Grover & Pea, 2013; National Research Council, 2011; Royal Society, 2017; Selby & Wollard, 2013). CT can be cultivated through integrated education in diverse areas, such as programming education or robotics, by using technology. Despite debates on whether programming education should be experienced during early childhood (Cordes & Miller, 2000; Voogt et al., 2015), there is a growing consensus that teaching CT as a thought process is necessary from pre-primary education and important to start from early childhood (Bers et al., 2014; Kermani & Aldemir, 2015; Shute et al., 2017). As young children are still in the early stages of development and the conceptual boundary of the term CT remains unclear, studies have been undertaken to investigate CT among young children using interviews or project-based assessments (e.g., Angeli & Valanides 2020; Bers et al., 2019; Mioduser & Levy, 2010). Interview or project-based assessments allow for the in-depth examination of thinking and learning progress among children. However, these assessments are limited due to factors such as subjective evaluation, small sample size, and time consumption. It would be useful if there are measurements that can be used in pre-and post-test designs to evaluate the effectiveness of the curriculum or can be administrated outside of specific research settings. Therefore, the objective of this study was to explore the applicability of the Korean version of the two assessments developed to assess CT among young children. To do this, this study examines the psychometric properties of the content validity, construct, validity, concurrent and criterion validity, as well as the internal reliability of the measurements.

2 Background

2.1 Computational thinking in early childhood

According to Wing (2011) and the Royal Society (2012), the basics of CT allow ordinary people to think computationally in everyday life, formulate and solve
problems through information processing, and implement efficient solutions. Researchers have strongly agreed that CT is a complex thought process utilizing several components of “abstractions and pattern generalizations, systematic processing of information, symbol systems and representations, algorithmic notions of flow of control, structured problem decomposition (modularizing), iterative, recursive, and parallel thinking, conditional logic, efficiency and performance constraints, debugging and systematic error detection” (Grover & Pea, 2013, pp. 39–40). However, considering the cognitive development of young children, these subcomponents of CT have a developmental sequence and may be different from those of older children. Similar to studies on the unity and diversity models for executive function among young children (e.g., Collette et al., 2005; Miyake & Friedman, 2012), CT in young children also involves high-level thinking in which various factors work simultaneously. Therefore, it is necessary to review these models.

Bers (2017) described young children’s CT as “Seven Powerful Ideas” based on Papert’s (1980) concept of “powerful ideas.” In Papert’s (1980) book, he viewed the powerful ideas as “learning how” skills within a domain or discipline, consistent with Piaget’s epistemology on knowledge. The term “powerful ideas” refers to the CT skills of young children, which can be learned through tangible programming or robotic education during early childhood, and comprises a series of skills such as algorithms, modularity, control structure, representation, hardware and software, design process, and debugging (Bers, 2017).

These powerful ideas were connected with underlying concepts and skills in computer science, appropriate for early childhood development and the curriculum (Bers, 2017). Algorithms refer to a series of sequential steps required to solve a problem or achieve a final goal. Sequencing, in which objects or actions are planned and arranged in the correct order, is an important skill that is learned in early childhood. Modularity is breaking down complex tasks or procedures into simpler, more manageable units. Control structures involve understanding various control structures, such as repetition, circulation, conditions, events, and nested structures, including pattern recognition for repeated regularity and cause and effect. Representation refers to understanding representative symbols and learning how to express ideas using these symbols. Ideas for hardware and software include understanding systems and the relationships between their components and operating systems. Design process ideas are step-by-step iterative processes of creating a program or final product. Lastly, debugging is systematically analyzing and evaluating a program or product. Thus, early childhood CT can be understood as children’s internalized ability to think and use knowledge in new ways and to connect their knowledge meaningfully to other areas.

2.2 Unplugged computing

“Tangible” computing education refers to learning and applying programming and engineering skills to young children using visible objects (Ishii & Ullmer, 1997). This is also referred to as “unplugged” computing as it facilitates the understanding of computer science principles and concepts without being connected to media or
computers (Bell et al., 2009). Understanding programming language structure and developing programming skills are difficult for first-time students or young children. However, unplugged computing makes the basic concepts and informatics related to computing visualized to children. Furthermore, it teaches children computer sciences, with topics such as artificial intelligence, human-computer interfaces, and programming (Bell et al., 2015; CS Unplugged, 2019; Kirçali & Özdener, 2022; Rodriguez et al., 2017).

Given the developmental stage of young children, the purpose and nature of early childhood education, and teaching and learning methods for young children, “tangible computing education” or “unplugged computing” is more useful and developmentally appropriate for early childhood computing education. From this perspective, the Computer Science Education Research Group of the University of Canterbury of New Zealand, Code.org, and Kinderlab Robotics support unplugged computer science and programming activities (Bell et al., 2015; https://kinderlabrobotics.com). Block-based programming tools, such as Cubetto and KIBO, represent different commands and functions for programming and allow children to perform programming tasks by creating screen-free robot instruction sequences. In particular, KIBO kits consist of wooden programming blocks with barcodes, parameter cards for the condition of functions, modules, and sensors attached to the robot body. Programming KIBO allows children to become engineers by participating in the design process of operating the robot (Sullivan & Bers, 2018). When children code with KIBO in various play and activities, they learn STEAM-related skills, which enable them to learn CT skills in a playful and tangible way.

2.3 Assessments of computational thinking for young children

Despite the increasing interest in CT and education inclusive of CT, only a few studies have focused on CT in young children. Additionally, most previous studies assessing CT in the early age group have used unstructured assessment methods, such as interviews or observation of children’s performance in tasks or projects. Wang et al. (2014) observed children aged five to nine, with experience using programming tools to evaluate their CT. In this study, children who had completed two main programming sessions were asked questions listed in a questionnaire, such as “What other things do you think the blocks can do?” “What are these used for?,” and “Is this easy to use?” They also included photographs and videotapes. This evaluation method provides multidimensional qualitative data on CT skills, including abstraction, automation, problem decomposition and analysis, and creativity. However, it does not distinguish between individual quantitative CT abilities. Mioduser and Levy (2010) studied how children acquired programming skills using a project to construct a robot’s behavior. They studied children’s artificial adaptive behavior, construction processes, and the role of partners and the environment by analyzing explanations provided by kindergartners. Moore et al. (2020) examined the CT of three second-grade children who participated in an enrichment program that integrated STEM and CT using a task-based interview approach.
The method used to measure CT in these studies is advantageous because it enables a more practical evaluation of children's performance and understanding. However, evaluating CT of many children is time-consuming and subjective on the examiner's part. Additionally, the children's verbal expression abilities covary with the interviews for assessing CT. There is a limit to the parallel identification of the effects of programming education and the content of learning in schools. Recently, studies have been conducted to objectively measure CT abilities in young children. Marinus et al. (2018) developed the Coding Development (CODE) Test 3–6, a coding assessment tool for three-to-six-year-olds. This instrument is based on a small wooden robot “Cubetto,” a simplified form of the LOGO turtle programming developed by Papert (1980) and comprises two types of items: “build” items (10 items) and “debug” items (3 items). A child is given three attempts to correct each item. If a child succeeds in fewer attempts, they obtain a higher score.

Similarly, Relkin (2018) developed Tufts Assessment of Computational Thinking in Children-KIBO robot version (TACTIC-KIBO), a specific platform instrument for five-to-seven-year-old children. This instrument examines children's programming proficiency based on the concepts of the “Seven Powerful Ideas” proposed by Bers (2017). It comprised seven questions or tasks at each of the four programming levels derived from the Developmental Model of Coding (Vizner, 2017). Children were asked to complete questions or tasks using KIBO robot kits. The KIBO robotic platform was originally developed as a developmentally appropriate tool for teaching the basics of computer programming to 4–7-year-old children. To manipulate the KIBO, children arranged and scanned the barcodes on programming blocks and parameter stickers which included “if” or “repeat.” Although these instruments measure children's coding knowledge or CT objectively and quantitatively, they depend on children's prior experiences of a certain programming platform such as Cubetto or KIBO.

To evaluate young children's CT learning and the effectiveness of the curriculum in pre-and post-test designs, assessments that are not dependent on a specific programming platform are needed. The Bebras Challenge (www.bebras.org) is an international education community aimed at promoting informatics and CT among school students. South Korea (Bebras Korea; www.Bebras.kr) has been officially affiliated with the Bebras Challenge community since 2017. Countries that have joined the Bebras Challenge run contests to solve tasks to measure CT skills during the worldwide Bebras Week. However, these tasks are performed in schools using online platforms. The contents of these tasks are mainly targeted at students from elementary school to high school, and tests are organized by tasks of varying difficulty levels depending on the age group. Google commissioned King's College London to develop sets of Bebras cards similar to card games used in classrooms and for children aged 7–11 (Sentance, 2018; www.bebras.uk). These cards were adapted from those developed and tested in Lithuania (Dagienė et al., 2017). Each card includes one task that evaluates CT concepts such as patterns, algorithms, logic, and abstraction. The difficulty levels of the 48 items in the task are divided into easy (17 items), medium (16 items), and hard levels (15 items).

However, CT measurements in young children have been performed without any validation procedures or targeted platforms. To address such issues, Relkin et al.
(2020) developed a new CT assessment tool, TechCheck, based on developmentally appropriate CT concepts for young children, and reported good psychometric properties. This study was conducted before the TechCheck scale was developed, and Relkin et al. (2020) evaluated the psychometric properties of the TACTIC-KIBO using the scores of children in grades 1 and 2 of elementary school. Therefore, this study validates the Bebras cards and the TACTIC-KIBO for assessing CT concepts among 5-year-old children in South Korea. By confirming the factor structure of the Bebras cards and the TACTIC-KIBO for preschoolers and examining the correlation of these factors, the findings of this study may be used to assess young children’s CT skills in classes and further research on CT as the effectiveness of integrated STEM education.

2.4 Mathematical ability and computational thinking

As CT has been accepted as a “new literacy of the 21st century” (Wing, 2011), there have been increasing studies examining how CT relates to existing academic subjects. CT refers to analytical thinking with several commonalities with mathematical thinking (e.g., problem-solving), engineering thinking (e.g., evaluating design and processes), and scientific thinking (e.g., systematic analysis) (Berland & Wilensky, 2015; Bers, 2010). According to Brownell’s (1945) meaning theory, understanding a mathematical concept refers to knowing its function, structure, and relationships (Higgins & Wiest, 2006). Through one-on-one interviews, Higgins and Wiest (2006) found that mathematics experiences are similar to computational procedures among second-graders. Furthermore, a study on CT and mathematics among preschool children reported that CT learning, such as algorithmic thinking, modularity, and debugging, was conducted through a set of prototype play-based learning tasks, such as card ordering, animal shapes, and creating two identical bracelets, already a part of preschool class activities (Lavigne et al., 2020). Moreover, researchers have found that preschoolers use mathematical knowledge, such as placing the cards in a logical sequence (algorithmic thinking), combining individual shapes and two or more shapes that represent modules (modularity), as well as a pattern, counting, or comparison to solve a task (debugging) during mathematics learning tasks.

Teacher candidates have thought that CT is naturally connected to mathematics and integrated to support and enhance the learning of existing subject areas (Gadanidis et al., 2017). An experimental study suggested that CT perspectives infused in mathematical problem-solving tasks improved kindergartners’ and first graders’ mathematics understanding and programming skills (Sung et al., 2017). Another review study on CT in K-12 mathematics classrooms considered linking mathematical concepts and CT with reporting. Notably, more studies incidentally linked mathematics learning with CT as opposed to creating explicit links (Hickmott et al., 2018). The aforementioned studies showed that multiple concepts in CT overlap significantly with elements of mathematics compared to other subjects. Therefore, this study hypothesized that the mathematical abilities among young children might demonstrate predictive validity for CT assessments.
2.5 The present study

The objective of this study was to verify the applicability of the Korean version of the Bebras cards and TACTIC-KIBO among young children in South Korea. This study has four aims: (1) to examine the content validity by performing the Item Information and Item Characteristic Curves for the Bebras cards and TACTIC-KIBO to evaluate the difficulty and discrimination power of each item; (2) to verify the factor structures of the Bebras cards and TACTIC-KIBO using confirmatory factor analysis (CFA); (3) to examine the concurrent and criterion validity of the Bebras cards and TACTIC-KIBO for young children; and (4) to assess the internal reliability of the two measurements using Cronbach’s alpha. Furthermore, Bebras cards that have good psychometric properties may be used on young children to assess CT skills and to determine the effectiveness of integrated STEM education and other programming and mathematical curricula in classrooms.

3 Method

3.1 Participants

The participants were recruited through information letters asking for their consent to participate in the STEAM education research for young children. This research was conducted in kindergartens and childcare centers in the cities of Seoul and Busan and Gyeonggi-do province in South Korea by incorporating programming education using unplugged robots. Participants included 450 five- and six-year-old children (mean age, 73 months; range, 62–81 months; boys, n=219; girls, n=231) from 30 different institutions (Table 1). Parents who agreed to participate in the study received information on STEAM education and the children’s test. Only children whose parents provided consent for them to participate in the test underwent the assessment tasks necessary for this study.

---

Table 1 Socio-Demographic features of participants (N=450)

| Variables              | Categories     | N (%) |
|------------------------|----------------|-------|
| Gender                 | Boys           | 219 (48.7) |
|                        | Girls          | 231 (51.3) |
| Area                   | Seoul          | 74 (16.4) |
|                        | Busan          | 138 (30.7) |
|                        | Gyeonggi-do    | 238 (52.9) |
| Types of Institutions  | Childcare center | 381 (84.7) |
|                        | Kindergarten   | 69 (15.3) |

---

1 The Republic of Korea has a two-tier system, comprising kindergartens and child-care centers, to provide early childhood education and care. Kindergartens serve children from age three through age five, and child-care centers serve children with working mothers from age zero through age five.
3.2 Procedures

The parents of all participants were informed about the procedures and provided written informed consent for their children to participate. All of the procedures performed in this study were approved by the institutional review board. The tasks on the easy-level cards of the Bebras Cards and the tasks and questions of the TACTIC-KIBO were translated into Korean and reviewed by three professors and field experts each, including the director of a childcare center and teachers with a master’s degree in the field of early childhood education and child studies. Children were administered the Bebras card game, TACTIC-KIBO, and the Early Numeracy Task as the pre-test of a larger study examining the effect of integrated STEAM education using a programming robot. Research assistants were trained to establish rapport with the children and to administer the assessments consistently. During free playtime in the institution, the children underwent one-on-one examinations with these research assistants in an independently prepared examination space or room separated from the classroom. Before the test, the child verbally consented to the test by responding to the following: “I am your teacher today. I will show you something new and fun, and I want to see how you work on it. Shall we do it together?” Each child participated in three sequential tasks. If the child was bored or unable to concentrate on the examination, the research assistant took a break between tasks and asked, ‘Shall we keep going?’ and continued or stopped and sent the child back to the classroom, depending on the child’s response. The average administration time for each task was approximately 15 min.

3.3 Measures

3.3.1 The Bebras cards

The Bebras cards present tasks related to CT concepts on the cards (Sentence, 2018; www.bebras.uk). Each card is assigned to one of the following CT concepts: Patterns, Algorithms, Logic, and Abstraction. The difficulty of the task on each card is divided into easy, medium, and hard levels. The easy-level task cards consist of 17 tasks that assess patterns, algorithms, and logic. Tasks exceeding the medium level were too difficult for six-year-olds in the pilot examination. Therefore, this study only used easy-level task cards even though they did not include abstraction tasks. When children say the order of pictures in Tasks 3 and 4, it may be difficult to express them verbally, so the actual picture cards of the pictures presented in the task were presented together (Fig. 1). Task 17 could be answered in various ways, so coin picture cards were provided so that the task could be solved by hand (Fig. 1). A maximum of three minutes was allocated for each task, and the test was stopped if the answers to five tasks were consecutively wrong or could not be provided. Each correct response was scored 1, whereas each incorrect response was scored 0.
Fig. 1 Photograph of the Bebras card tasks. Note: Panel a: Photograph of the Bebras Card task 4. Panel b: Photograph of the Bebras Card task 2. Panel c: Photograph of the Bebras Card task 3. Panel d: Photograph of the Bebras Card task 17.
3.3.2 The TACTIC-KIBO

The TACTIC-KIBO was used because it is time-tested and developmentally appropriate for assessing young children’s CT, even though it is programing platform-specific. The TACTIC-KIBO was developed and validated by Relkin (2018) to assess CT in children aged four to seven. It consists of four levels of activity using the KIBO kit, and each level contains questions and tasks associated with Seven Powerful Ideas: Control Structure, Hardware, Software, Representation, Algorithms/Modularity, Debugging, and Design Process. The research assistant programmed the KIBO in front of the child using the specified blocks, parameter cards, and additional sensors and modules for each task and verbally asked a question or a task related to the performance (e.g., question 1 [control structure] at level 2: [Scan blocks and press triangle on KIBO] Watch the KIBO robot. What is the robot doing now that is different from the first program?) (Fig. 2). Each response to the question or task was scored as 1 point for satisfactory and 0 for unsatisfactory. If the total numerical score at each level was higher than 3, the assessment was continued for the next level of questions and tasks.

3.3.3 The early numeracy task

To establish criterion validities of the Bebras cards and the TACTIC-KIBO, the Early Numeracy Task was used. This task was part of the Early Years Toolbox (EYT), a battery of measures of young children’s executive function, self-regulation, language, and numeracy. The EYT was developed by Howard and Melhuish (2015) and validated in Korea by Chung et al. (2018). The EYT is a brief, game-like, valid, and reliable assessment tool used with young children aged three to six years and can be administered with the EYT app on an iPad. The Early Numeracy Task, one of the EYT task, assesses 12 different early numeracy abilities: spatial thinking and measurement, number concepts, cardinality, identifying numbers, matching numbers and digits, counting a subset, number order, conceptual subitizing, ordinality, word problems, patterning, and equations. Prior to starting the tasks, the child is provided with dynamic visual and auditory feedback from instructional and practice trials. This assessment is a format that helps the participant solve the problem and comprises 85 tasks. Scores are calculated as the number of correct items, and if the child responds incorrectly to five consecutive tasks, the assessment is stopped.

3.4 Data analyses

Data were analyzed using SPSS 26.0, SPSS Statistics Korea Plus Pack based on R software, and AMOS. First, items were examined using the two-parameter logistic (2-PL) Item Response Theory (IRT) and Item Characteristics Curves (ICC) to evaluate whether they functioned differently in the difficulty and discrimination power of individual questions. Through Item Information Curves (IIC), the amount of information yielded by each item measuring the underlying latent ability was examined.
Fig. 2. Example of the TACTIC-KIBO task assessment. Note: Panel a: Assessment sheet of level 2 of the TACTIC-KIBO. Panel b: Materials for level 2 tasks of the TACTIC-KIBO.
(Baker & Kim, 2017). Second, to investigate the factor structure, CFA of the Bebras cards and the TACTIC-KIBO was conducted. The average variance extracted (AVE) and construct reliability (CR) were calculated using the standardized regression weight and variance estimates of each variable from the CFA. Third, the correlations among the Bebras cards, TACTIC-KIBO, and the Early Numeracy task were calculated to examine concurrent and criterion validity. Lastly, internal consistency reliability coefficients were examined using Cronbach’s alphas for the Bebras cards and the TACTIC-KIBO.

### 4 Results

#### 4.1 Descriptive statistics

The average for the total Bebras cards score was 4.96 ($SD=2.76$) of 17 points, and the averages of the subdomains of the Bebras cards were as follows: Patterns 1.99 ($SD=1.05$), Algorithms 2.06 ($SD=1.53$), and Logic 0.97 ($SD=0.89$) (Table 2). Table 3 presents the total mean score for the TACTIC-KIBO, 9.55 ($SD=5.28$) of

| Item no. | Mean ($SD$) | Percent correct | Difficulty | Discrimination |
|----------|-------------|-----------------|------------|---------------|
| Patterns |             |                 |            |               |
| Item 1   | 0.27 (0.44) | 26.9            | 9.68       | 0.10          |
| Item 2   | 0.92 (0.27) | 92.2            | -10.18     | 0.25          |
| Item 9   | 0.25 (0.43) | 23.1            | 1.64       | 0.70          |
| Item 10  | 0.42 (0.49) | 37.6            | 0.30       | 1.74          |
| Item 11  | 0.14 (0.35) | 12.2            | 1.88       | 1.18          |
| Item 12  | 0.08 (0.27) | 6.9             | 6.85       | 0.36          |
| Algorithms |         |                 |            |               |
| Item 3   | 0.38 (0.49) | 38.2            | 0.40       | 1.51          |
| Item 4   | 0.28 (0.45) | 27.6            | 0.81       | 1.67          |
| Item 5   | 0.57 (0.50) | 56.7            | -0.57      | Bac.93        |
| Item 6   | 0.18 (0.38) | 17.8            | 3.92       | 0.36          |
| Item 7   | 0.25 (0.43) | 25.1            | 2.22       | 0.44          |
| Item 13  | 0.34 (0.48) | 30.0            | 0.87       | 0.82          |
| Item 15  | 0.13 (0.33) | 10.9            | 2.09       | 1.11          |
| Logic    |             |                 |            |               |
| Item 8   | 0.44 (0.50) | 40.9            | 0.18       | 1.18          |
| Item 14  | 0.08 (0.27) | 7.1             | 2.68       | 1.04          |
| Item 16  | 0.48 (0.50) | 41.3            | 0.10       | 1.29          |
| Item 17  | 0.01 (0.11) | 1.1             | 3.37       | 1.59          |

$N=450$
the 28 points, and the mean scores of each item. As shown in Table 3, the higher the TACTIC-KIBO level, the lower the average score. The subdomain averages of TACTIC-KIBO were as follows: Control Structure 1.95 (SD = 1.13), Hardware 1.29 (SD = 0.69), Software 1.89 (SD = 1.01), Representation 1.61 (SD = 1.20), Algorithms 0.90 (SD = 1.08), Debugging 0.94 (SD = 0.99), and Design Process 0.98 (SD = 0.63).

Table 3 Overall item difficulty and discrimination estimates of the TACTIC-KIBO based on the 2-parameter logistic model

| Item no. | Mean (SD) | Percent correct | Difficulty | Discrimination |
|----------|-----------|-----------------|------------|----------------|
| Level 1 (n = 332) | 4.19 (1.28) |                |            |                |
| Item 1  | 0.72 (0.45) | 71.9            | -2.05      | 0.87           |
| Item 2  | 0.89 (0.32) | 88.3            | -5.37      | 0.47           |
| Item 3  | 0.85 (0.36) | 84.4            | -3.05      | 0.98           |
| Item 4  | 0.71 (0.45) | 70.7            | -0.94      | 2.74           |
| Item 5  | 0.05 (0.22) | 5.1             | 4.65       | 0.54           |
| Item 6  | 0.34 (0.47) | 33.5            | 1.36       | 0.09           |
| Item 7  | 0.83 (0.38) | 82.3            | -5.06      | 0.70           |
| Level 2 (n = 283) | 3.40 (1.92) |                |            |                |
| Item 1  | 0.87 (0.34) | 73.4            | -2.15      | 1.87           |
| Item 2  | 0.31 (0.46) | 26.0            | 0.52       | 1.04           |
| Item 3  | 0.82 (0.39) | 69.2            | -1.65      | 1.62           |
| Item 4  | 0.63 (0.48) | 53.3            | -0.68      | 2.41           |
| Item 5  | 0.52 (0.50) | 44.3            | -0.59      | 1.61           |
| Item 6  | 0.48 (0.50) | 40.4            | -0.25      | 3.06           |
| Item 7  | 0.09 (0.28) | 7.5             | 1.49       | 1.22           |
| Level 3 (n = 217) | 1.75 (1.67) |                |            |                |
| Item 1  | 0.63 (0.48) | 40.7            | -0.64      | 2.16           |
| Item 2  | 0.03 (0.18) | 2.1             | 3.24       | 1.06           |
| Item 3  | 0.45 (0.50) | 29.0            | -0.34      | 1.94           |
| Item 4  | 0.49 (0.50) | 31.7            | -0.26      | 4.02           |
| Item 5  | 0.40 (0.49) | 25.7            | -0.23      | 3.41           |
| Item 6  | 0.19 (0.39) | 12.3            | 0.85       | 1.45           |
| Item 7  | 0.04 (0.19) | 2.4             | 3.77       | 0.72           |
| Level 4 (n = 108) | 0.84 (1.46) |                |            |                |
| Item 1  | 0.23 (0.42) | 7.5             | 0.87       | 2.70           |
| Item 2  | 0.35 (0.48) | 11.4            | 0.56       | 1.83           |
| Item 3  | 0.15 (0.36) | 4.8             | 1.35       | 2.07           |
| Item 4  | 0.13 (0.34) | 4.2             | 1.05       | 4.96           |
| Item 5  | 0.45 (0.50) | 14.7            | 0.22       | 2.04           |
| Item 6  | 0.21 (0.41) | 6.9             | 1.43       | 1.09           |
| Item 7  | 0.16 (0.37) | 5.1             | 1.90       | 1.03           |
significant difference between boys and girls ($t = -2.45, df = 448, p < .05$) in the overall mean score and in the algorithm ($t = -2.65, df = 448, p < .01$) and logic domains ($t = -2.71, df = 418, p < .01$) among the subdomains of the Bebras cards. Thus, the CT scores were significantly higher among the girls. The overall mean score of the TACTIC-KIBO for male and female participants was 9.94 ($SD = 5.33$) and 9.21 ($SD = 5.23$), respectively. Furthermore, there was no significant difference between the genders ($t = 1.26, df = 330, p > .05$) in the overall mean score. However, significant gender differences were found only in the control structure ($t = 2.48, df = 330, p < .05$) among the TACTIC-KIBO subdomains.

### 4.2 Content validity (item response theory)

Table 2 presents the item difficulty and discrimination estimates for the Bebras cards. The IRT analysis results of the items of the Bebras cards showed that discrimination values were all above 0, and the mean discrimination index was 0.96, ranging between 0.1 and 1.74. Regarding the discrimination index, 4 of the 17 items showed higher than moderate discrimination. ICCs (Fig. 3) are S-shaped curves that show the probability of the correct response to the Bebras cards for children’s CT abilities. The location of the peak of the ICCs presents the level of difficulty, and the steepness indicates the question’s discrimination, with steeper curves discriminating better. Therefore, the various locations of the ICC peaks show the CT ability of five-to-six-year-old children, and the various slopes of the ICCs show acceptable levels of discrimination. The peak point of the IIC slopes (Fig. 4) indicates the point where there is a 50% chance of getting the item right. In this study, most peaks occurred either toward the middle or the right end of the x-axis (latent ability), indicating that the Bebras cards provide better information about children with average or higher CT ability.

![Image of Item Characteristic Curves](image-url)
Table 3 presents the IRT results for the TACTIC-KIBO. The results showed that discrimination values were above 0, and the mean discrimination index was 1.77, ranging from 0.09 to 4.96. The discrimination index of 16 of the 28 items showed above moderate discrimination (Baker & Kim, 2017). The ICCs (Fig. 5) present the probability of the correct response in the TACTIC-KIBO for children’s CT abilities. Given the location of the peak of the ICCs and the steepness of the ICCs, the TACTIC-KIBO tasks consisted of difficult and highly discriminating questions on the CT ability of children aged five to six years. In the IICs (Fig. 6), one item provided more information at high ability levels above $\theta = 5$, but almost no information about low ability levels, because the items were
already too hard for the participants. However, other items with low discriminability do not provide much information overall but cover a range of ability levels.

### 4.3 Construct validity

CFAs were employed to examine the model fit of the Bebras cards and the TACTIC-KIBO, developed based on concept models for the components of young children’s CT ability. The model fit indices for the Bebras cards were $\chi^2/df = 1.42, p = .00$, TLI = 0.83, CFI = 0.87, and RMSEA = 0.03. The indices of the Bebras model were composed of three theoretical constructs, showing a good model fit. Convergent validity is determined by the index of the AVE (Fornell & Larcker, 1981), the degree to which a latent construct and the observed variables are related, and by the index of CR, the internal consistency of scale items. The AVE index and the CR value of the Bebras three-construct model were 0.47 and 0.92, respectively, indicating sufficient convergent validity. However, some of the items loaded on each factor contained items with a very low loading value, such that the items needed to be modified later (Fig. 7).

Therefore, items with a factor loading lower than 0.30 with the theoretical construct were deleted, and only two measurement items remained in the patterns and logic constructs. Based on the suggestion that three and more items are necessary to “provide minimum coverage of the construct’s theoretical domain” (Hair et al., 2010, p. 676), Patterns and Logic constructs were combined into a pattern recognition construct. This has been considered a cornerstone of CT (BBC Bitesize, 2022, https://www.bbc.co.uk/bitesize/guides/zp92mp3/reVision/1). The model fit indices for the Bebras cards with two constructs were $\chi^2/df = 3.51, p = .00$, TLI = 0.72, CFI = 0.85, and RMSEA = 0.07; these indices showed a good model fit. Furthermore, all items loaded in each
factor were higher than 0.30 (Fig. 8). The AVE index and the CR value of the Bebras two-construct model were 0.59 and 0.91, respectively, indicating sufficient convergent validity.

The model fit indices for the TACTIC-KIBO were $\chi^2/df = 1.59$, $p = .00$, TLI = 0.80, CFI = 0.84, and RMSEA = 0.03. The indices of the TACTIC-KIBO model, comprised seven theoretical factors, which showed good model fit. The AVE index and the CR value of the TACTIC-KIBO seven-factor model were 0.63 and 0.98, respectively, indicating very good convergent validity. However, there were items with significant but low factor-loading values (Fig. 9).
Concurrent and criterion validity analyses were conducted by calculating correlations among the Bebras card scores, TACTIC-KIBO scores, and early numeracy scores (see Tables 4 and 5).

Several significant positive correlations emerged as evidence of concurrent validity between the Bebras cards and the TACTIC-KIBO (see Table 4). There was a significant positive association among all the subfactors of the Bebras cards. Scores on Pattern Recognition ($r = .22$, $p < .001$) and Algorithms ($r = .15$, $p < .01$) of the Bebras cards were positively correlated with Representation among the TACTIC-KIBO subdomains. The Bebras scores on Pattern Recognition ($r = .28$, $p < .001$) and Algorithms ($r = .19$, $p < .01$) were positively associated with Algorithms among the TACTIC-KIBO subdomains. The total score of the Bebras cards was positively associated with Hardware, Software, Representation, and Algorithms among the subdomains (ranging from $r = .11$, $p < .05$ to $r = .28$, $p < .001$), and the total score of the TACTIC-KIBO ($r = .18$, $p < .01$), except for the Control Structure, Debugging, and Design Process subdomains. There were significant positive correlations between all subfactors of the TACTIC-KIBO.

Several significant positive correlations emerged, establishing the criterion validity (Table 5). The Bebras cards’ subfactors, total score, TACTIC-KIBO’s subfactors, and total score were positively correlated with most of the subfactors of early mathematics ability and total score. We found that the CT ability among young children was somewhat related to their mathematical abilities.
4.5 Internal reliability

The internal reliabilities of the Bebras cards and TACTIC-KIBO were examined using Cronbach’s alpha. The internal reliability of the subfactors of Bebras was low, with **pattern recognition** $\alpha = 0.52$, and **algorithms** $\alpha = 0.49$. However, the reliability coefficient of all items was $\alpha = 0.66$, indicating a high and acceptable internal reliability. Similarly, the reliability coefficients of the subfactors of the TACTIC-KIBO were in the low to medium magnitude range, indicating **Control Structure** $\alpha = 0.46$, **Hardware** $\alpha = 0.11$, **Software** $\alpha = 0.51$, **Representation** $\alpha = 0.74$, **Algorithms** $\alpha = 0.64$, **Debugging** $\alpha = 0.37$, and **Design Process** $\alpha = 0.39$. However, the Cronbach’s coefficient of the total TACTIC-KIBO items was very high and satisfactory, with $\alpha = 0.88$. 

![Fig. 9 Confirmatory factor analysis for the TACTIC-KIBO. Note: The statistics presented are standardized regression coefficients, and all coefficients are significant ($p < .01$)](image-url)
**Table 4** Concurrent validity

|                | 1       | 2       | 3       | 4       | 5       | 6       | 7       | 8       | 9       | 10      | 11      |
|----------------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|---------|
| 1. Bebras_Pattern Recognition | 1       |         |         |         |         |         |         |         |         |         |         |
| 2. Bebras_Algorithms          | 0.45*** | 1       |         |         |         |         |         |         |         |         |         |
| 3. Bebras_Total              | 0.86*** | 0.87*** | 1       |         |         |         |         |         |         |         |         |
| 4. KIBO_Control Structure    | 0.12*   | 0.04    | 0.09    | 1       |         |         |         |         |         |         |         |
| 5. KIBO_Hardware             | 0.13*   | 0.09    | 0.13*   | 0.58*** | 1       |         |         |         |         |         |         |
| 6. KIBO_Software             | 0.14*   | 0.06    | 0.11*   | 0.64*** | 0.47*** | 1       |         |         |         |         |         |
| 7. KIBO_Representation       | 0.22*** | 0.15**  | 0.22*** | 0.62*** | 0.50*** | 0.67*** | 1       |         |         |         |         |
| 8. KIBO_Algorithms           | 0.28*** | 0.19**  | 0.28*** | 0.64*** | 0.55*** | 0.60*** | 0.69*** | 1       |         |         |         |
| 9. KIBO_Debugging            | 0.13*   | 0.02    | 0.09    | 0.55*** | 0.42*** | 0.51*** | 0.54*** | 0.58*** | 1       |         |         |
| 10. KIBO_Design Process      | 0.10    | −0.03   | 0.04    | 0.40*** | 0.32*** | 0.42*** | 0.40*** | 0.41*** | 0.48*** | 1       |         |
| 11. KIBO_Total               | 0.21*** | 0.11    | 0.18**  | 0.84*** | 0.69*** | 0.81*** | 0.84*** | 0.84*** | 0.76*** | 0.59*** | 1       |

*N* of Bebras respondents = 450. *N* of TACTIC-KIBO respondents = 332. *p* < .05. **p** < .01. ***p** < .001
|                             | Early Numeracy Ability |
|-----------------------------|------------------------|
|                             | Number concept | Spatial and measurement | Counting a subset | Identifying numbers | Matching numbers and digits | Number order | Ordinality | Cardinality | Conceptual subitizing | Patterning | Number word | Equation | Total |
| Bebras_ Pattern recognition | 0.11*           | 0.13**                | 0.22***             | 0.12*              | 0.12*                    | 0.18***      | 0.12*      | 0.15**      | 0.23***             | 0.27***     | 0.26***     | 0.28*** |
| Bebras_ Algorithms          | 0.09             | 0.21**                | 0.19**              | 0.16**             | 0.12*                    | 0.16**       | 0.16**     | 0.15**      | 0.27***             | 0.25***     | 0.23***     | 0.29*** |
| Bebras_ Total               | 0.13**           | 0.21**                | 0.22***             | 0.18***             | 0.15**                   | 0.21***      | 0.17***    | 0.18***      | 0.30***             | 0.30***     | 0.29***     | 0.34*** |
| KIBO_ Control Structure     | 0.14*            | 0.15**                | 0.18**              | 0.15**             | 0.16**                   | 0.16**       | 0.20**     | 0.14*       | 0.21***             | 0.22***     | 0.26***     |        |
| KIBO_ Hardware              | 0.12*            | 0.12*                 | 0.14*               | 0.12*              | 0.12*                    | 0.18**       | 0.14**     | 0.08        | 0.12*               | 0.18**      | 0.15**      | 0.21*** |
| KIBO_ Software              | 0.21***          | 0.19**                | 0.19***             | 0.19***             | 0.20***                  | 0.20***      | 0.25***    | 0.18**      | 0.25***             | 0.20***     | 0.28***     | 0.33*** |
| KIBO_ Representation        | 0.16**           | 0.13**                | 0.18**              | 0.23***             | 0.22***                  | 0.20***      | 0.20***    | 0.18**      | 0.20***             | 0.26***     | 0.29***     | 0.31*** |
| KIBO_ Algorithms            | 0.17**           | 0.20***               | 0.20***             | 0.21***             | 0.21***                  | 0.17**       | 0.19**     | 0.23***     | 0.23***             | 0.28***     | 0.25***     | 0.32*** |
| KIBO_ Debugging             | 0.14*            | 0.13*                 | 0.20***             | 0.13*              | 0.19**                   | 0.20***      | 0.24***    | 0.12*       | 0.13*               | 0.24***     | 0.23***     | 0.27*** |
Table 5 (continued)

| Early Numeracy Ability | Number concept | Spatial and measurement | Counting a subset | Identifying numbers | Matching numbers and digits | Number order | Ordinality | Cardinality | Conceptual subitizing | Patterning | Number word | Equation | Total |
|------------------------|----------------|-------------------------|-------------------|---------------------|----------------------------|--------------|------------|-------------|----------------------|------------|-------------|---------|-------|
| KIBO_ Design process   | 0.10           | 0.11*                   | 0.13*             | 0.13*               | 0.09                       | 0.15**       | 0.17**     | 0.15**      | 0.06                 | 0.12*      | 0.15**      | 0.18**  |       |
| KIBO_ Total            | 0.19***        | 0.19**                  | 0.23***           | 0.22***             | 0.22***                    | 0.24***      | 0.25***    | 0.20***     | 0.21***              | 0.30***    | 0.29***      | 0.35*** |       |

N of early numeracy respondents = 416. N of Bebras respondents = 416. N of TACTIC-KIBO respondents = 332. *p < .05. **p < .01. ***p < .001
5 Discussion

This study examined the validity and reliability of the Korean version of the Bebras cards and TACTIC-KIBO for young children aged 5–6 years in South Korea. This study contributes to the existing psychometric analyses of two measurements: a measurement that requires familiarity with a particular coding platform, TACTIC-KIBO, and a measurement that does not require it, the Bebras cards. Overall, these two measurements demonstrated moderate-to-good psychometric properties, except for items that were too difficult for the age level and based on unfamiliar tools.

5.1 Content validity

In the IRT results of the two measurements, the Bebras cards showed a much larger difference in difficulty between the items than that of the TACTIC-KIBO. Items that are too easy or difficult should be newly developed or adjusted to the appropriate difficulty levels. The level of difficulty of the Bebras cards was somewhat high for five-to six-year-olds and was greater than that of the TACTIC-KIBO. Furthermore, the IRT results of the TACTIC-KIBO verified that the difficulty of measurement gradually escalated depending on the level. Five items on the Bebras cards showed low discrimination indices, whereas the remaining 12 items demonstrated moderate discrimination indices. This study found that the Bebras measurement could assess the average or higher CT ability of children. In the IRT results of the TACTIC-KIBO, only three items of Level one showed low discrimination indices, while the remaining items presented moderate to higher discrimination indices. Thus, the IRT results of the CT measurements showed acceptable psychometric properties for capturing the CT abilities among young children.

5.2 Construct validity

The CFA analyses evaluated how well the measured items represented the theoretical constructs of the Bebras cards and TACTIC-KIBO. The CFA results confirmed that the hypothesized models of the two measurements had good model fits. The TACTIC-KIBO was developed based on the Seven Powerful Ideas theory of CT in young children (Bers, 2017). Although the model fit, AVE, and CR indices of the two measurements were within the acceptable range, the loading values of some items of the subfactors were lower than 0.5. Scholars have suggested that factor-loading estimates should be higher than 0.5, ideally, 0.7 or higher (Chen & Tsai, 2007; Hair et al., 2010). Weak factor loadings in CFA may occur when the specified model is incorrect or is not specified correctly in the population (Ximénez, 2009). Upon deleting the items loaded with values less than 0.30 at the guideline cut-off point for identifying significant factor loadings based on sample size by Hair et al. (2010), the model fit indices of the Bebras two-construct model were acceptable, and item loading estimates improved.
In mathematics, a pattern refers to a repeated arrangement such as of numbers, shapes, and colors. For example, considering the contents of the items linked to patterns, the first subfactor of the Bebras cards, Items 1 and 2 are items that find and connect similar features; Items 9 and 12 are questions related to a typical pattern that finds a sequence based on repeated colors and shapes. Items 10 and 11 classified the same common features based on color, size, or shape. These can be viewed as typical questions and tasks related to patterns in mathematics. In terms of the logic of the Bebras, logical reasoning refers to the process of applying rules to solve problems in computer science. Easy-level logic items ask to find a picture that has an identical description (Item 8), guess missing pictures (Item 16), or solve problems using numbers (Items 14 and 17). Even though Items 14 and 17 were designated as easy, the passing rates of Items 14 and 17 were 7.1% and 1.1%, respectively; these items were very difficult for 5-6-year-olds to solve (see Table 2). Items 8 and 16, except for two items of the logic construct, are similar to pattern recognition skills, which look for similarities among and within problems as one of the CT skills. Therefore, the construct created by combining patterns and logic is referred to as pattern recognition. Young children who are still cognitively developing and lack specific high-level thinking skills seem to be at a stage before CT’s major higher-order thinking functions are subdivided. They apply ordinary thinking skills, such as pattern recognition, to identify problems and develop possible solutions.

However, items of TACTIC-KIBO asking about specific functions or knowledge of KIBO programming showed weak factor loadings. This shows that an assessment based on a specific platform may be inappropriate for measuring the CT of young children who have no experience with a specific platform. Therefore, to clarify the cause of the weak factor loadings, further studies are needed to examine the items of these two measurements in various countries and wide age groups. Additionally, further studies are necessary to establish a model for the constituent components of CT in young children, as the composition of the subfactors of CT in the two measurements is composed of other theoretical constructs excluding the algorithm.

5.3 Concurrent and criterion validity

The analyses of the concurrent validity of the two measurements indicated that both the subfactors and total scores of the two measurements were not large despite significant intercorrelation with each other. In terms of the subfactor items of the TACTIC-KIBO that did not correlate with the subfactors of the Bebras cards, the items comprised questions that inquired about the actual operation method or related knowledge about KIBO robotics, such as control structure, hardware, software, and design process. In the subdomains such as representation and algorithms, where young children were presented with tangible KIBO kits in the task situation and asked to solve the problem, all the items in the representation and algorithms were found to have a significant correlation with the subfactors of the Bebras cards. Furthermore, they seemed to assess the CT talent among young children. Therefore, Bebras cards may be more useful in evaluating
CT abilities among children in educational settings or research and in identifying children who need educational support.

The results of the criterion validity of the two measurements proved that CT abilities were associated with and predicted by mathematical abilities such as logical sequence, identifying a subset, or matching sets, and mathematical knowledge such as patterns, counting, or comparison to solve a problem (Gadanidis et al., 2017; Hickmott et al., 2018; Lavigne et al., 2020; Sung et al., 2017). The correlation coefficient ranged from 0.22 to 0.37, indicating that the correlation was weak despite a significant positive relationship. CT ability and mathematical ability were measured simultaneously. Thus, it was observed that mathematical concepts and knowledge facilitate computational procedures in analytical thinking and problem-solving. Additionally, CT is likely linked to other disciplines such as scientific thinking, engineering thinking, and mathematics. This supports the argument that CT can be viewed as a 21st-century ability that is consistent with the concepts and thinking skills prevalent in various disciplines and integrated with STEAM education (Kelly & Knowles, 2016; Kennedy & Odell, 2014; Stohlmann et al., 2012).

### 5.4 Internal reliability

The internal consistency results indicated that all items on the two measurements uniformly measured the CT abilities among young children. However, the internal consistency of each subfactor item was low. This implies that both these measurements are more reliable when the aggregated and single total scores are used as opposed to using the individual scores of various subfactors. The results of the CFA also support this finding. Given that CT is not a fully unified construct but a complex mixture of several domains of thinking and problem-solving abilities, the composite scores of the two measurements may validly and reliably reflect the CT ability among young children. Further studies are required to identify common constructs that capture the CT ability among children.

### 5.5 Limitation and directions for future research

The current study has several limitations. Although the Bebras cards and TACTIKIBO are applicable and acceptable measurements for preschoolers in South Korea, solving items or performing tasks of the measurement may have been challenging for five-to six-year-old children. In addition, this study had a relatively small sample size. Future studies should expand performance and response results to a larger sample of five-to-six-year-old children with corresponding measurements. Thus, the developmental and performance levels of CT ability in young children should be reported and analyzed further using comparative studies and an increased accumulation of information on various races, ages, and countries.

The two measurements in this study are based on the theoretical components of CT ability. However, the CFA results showed low loadings of several items to subfactors. When five experts in related fields examined the face validity of the items and the tasks of the two measurements, they evaluated that the various items...
and tasks reflected the subdomain to be measured but could be difficult for the target age groups. Additionally, as some items in the TACTIC-KIBO inquired about the method of a specific programming platform, some responses were skewed in the current sample and showed low internal reliability in subfactors such as *Hardware* and *Debugging* that specifically inquired about solving the programming problem and the mechanical function of the KIBO. Considering the developmental level of five-to-six-year-old children, it might be because the item or the task may be double-loaded to various factors or a common factor, or because it requires additional abilities among children, such as the ability to follow instructions well, focus on a specific task, or comprehend language.

### 5.6 Conclusions

This study provides valuable information for measuring and evaluating young children’s CT abilities under elementary graders. The Bebras cards and TACTIC-KIBO have acceptable psychometric properties and are applicable for measuring different CT abilities among young children, despite demonstrating low internal consistencies of subfactors and slightly weak correlations between subfactors. The Bebras cards, which do not require any coding platform familiarity, are more suitable for application in research and educational settings than the TACTIC-KIBO, which requires a particular coding platform such as KIBO robot kits. The TACTIC-KIBO can be used by teachers and researchers to measure children’s understanding or acquired knowledge after using the KIBO, rather than a pre-test of CT ability. As the Bebras cards and TACTIC-KIBO were not designed to quantify CT ability in each of the subfactors, it would be preferable to use scores that combine the results of multiple subfactors into a single total score.

Additionally, as the TACTIC-KIBO comprises more subfactors than the Bebras cards among the two CT measurement tools, further discussion is required to underscore the definition and composition of the CT concept among young children. The results of the two measurements of CT among young children suggest designing and implementing assessments for young children’s CT ability. Furthermore, pertinent conversations between professionals and field experts may be indispensable regarding the CT concept and the appropriate CT level for the development of young children that can be linked to the school curriculum. Accordingly, attempts to revise and develop CT measurements are required, as are empirical studies on the reliability and validity of new measurements targeting various groups.

**Acknowledgements** I sincerely appreciate the research manager, Ji Young Lee, research assistants, Jiyoung Park and Jeongsu Noh, and all the undergraduate research assistants who made this project possible. I would also like to thank all of the kindergarten and childcare center educators, parents, and young children who participated in this study as well as collaborators of the research project, Drs. Hui Young Chun and Soyeon Park. Finally, I gratefully acknowledge the grant support of the Korean National Research Foundation for this work.

**Funding** This work was supported by the Korean National Research Foundation (NRF-2018S1A5A2A03036263) grant entitled “The development and effectiveness of an unplugged activities curriculum on early childhood STEAM education: A mixed-method approach.”
Data availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval All procedures performed in the study involving human participants were in accordance with the ethical standards of the Kosin University Institutional Review Board (KU IRB no. 2019-0008) and with ethical standards laid down in the 1964 Declaration of Helsinki and its later amendments or comparable ethical standards.

Consent to participate Informed consent was obtained from educators and parents/guardians of the participating children. The children verbally provided assent to inclusion.

Consent for publication The authors who developed the tools used in the study provided permission to use the measurement scales for the study and publish the results.

Conflicts of interest/Competing interests The author has no conflict of interest to disclose.

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

References

Angeli, C., & Valanides, N. (2020). Developing young children’s computational thinking with educational robotics: An interaction effect between gender and scaffolding strategy. Computers in Human Behavior, 105, 105954. https://doi.org/10.1016/j.chb.2019.03.018

Baker, F. B., & Kim, S. H. (2017). The basics of item response theory using R. Springer.

BBC Bitesize (2022). KS3: Introduction to computational thinking. BBC. https://www.bbc.co.uk/bitesize/guides/zp92mp3/revision/1

Bell, T., Alexander, J., Freeman, I., & Grimley, M. (2009). Computer science unplugged: School students doing real computing without computers. The New Zealand Journal of Applied Computing and Information Technology, 13(1), 20–29.

Bell, T., Witten, I. H., & Fellows, M. (2015). CS Unplugged. An enrichment and extension programme for primary-aged students. University of Canterbury. CS Education Research Group Version 3.2.2.

Berland, M., & Wilensky, U. (2015). Comparing virtual and physical robotics environments for supporting complex systems and computational thinking. Journal of Science Education and Technology, 24(5), 628–647. https://doi.org/10.1007/s10956-015-9552-x

Bers, M. U. (2010). The TangibleK Robotics program: Applied computational thinking for young children. Early Childhood Research & Practice, 12(2), 1–20. http://ecrp.uiuc.edu/v12n2/bers.html

Bers, M. U. (2017). Coding as a playground: Programming and computational thinking in the early childhood classroom. Routledge.

Bers, M. U., Flannery, L., Kazakoff, E. R., & Sullivan, A. (2014). Computational thinking and tinkering: Exploration of an early childhood robotics curriculum. Computers & Education, 72, 145–157. https://doi.org/10.1016/j.compedu.2013.10.020

Bers, M. U., González-González, C., & Armas–Torres, M. B. (2019). Coding as a playground: Promoting positive learning experiences in childhood classrooms. Computers & Education, 138, 130–145. https://doi.org/10.1016/j.compedu.2019.04.013
Brownell, W. (1945). When is arithmetic meaningful? The Journal of Educational Research, 38(7), 481–498. Retrieved May 20, 2021, from http://www.jstor.org/stable/27528612
Chen, C. F., & Tsai, D. (2007). How destination image and evaluative factors affect behavioral intentions? Tourism management, 28(4), 1115–1122. https://doi.org/10.1016/j.tourman.2006.07.007
Chung, S., Kim, I., Kim, H., Ma, Y., & Park, B. (2018). Validation study of the Korean version of early years toolbox (EYT). Korean Journal of Child Studies, 39(6), 131–142. https://doi.org/10.5723/jkcs.2018.39.6.131
Collette, F., Van der Linden, M., Laureys, S., Delfiore, G., Degueldre, C., Luxen, A., & Salmon, E. (2005). Exploring the unity and diversity of the neural substrates of executive functioning. Human Brain Mapping, 25(4), 409–423. https://doi.org/10.1002/hbm.20118
Cordes, C., & Miller, E. (2000). Fool’s gold: A critical look at computers in childhood. ERIC.
CS Unplugged (2019). Principles. Retrieved March 24, 2022, from https://csunplugged.org/en/principles/
Dagliene, V., Fuchs, G., Koivisto, J., & Stupuriënė, G. (2017). The card game of Bebras-like tasks for introducing informatics concepts. In ISSEP 2017 Online Proceedings. Helsinki, 13–15 November 2017.
Fornell, C., & Larcker, D. F. (1981). Evaluating Structural Equation Models with unobservable variables and measurement error. Journal of Marketing Research, 18(1), 39–50. https://doi.org/10.2307/3151312
Furber, S. (2012). Shut down or restart? The way forward for computing in UK Schools. The Royal Society. Retrieved from https://royalsociety.org/topics-policy/projects/computing-in-schools/report
Gadanidis, G., Cendros, R., Floyd, L., & Namukasa, I. (2017). Computational thinking in mathematics education. Contemporary Issues in Technology and Teacher Education, 17(4), 458–477. Retrieved May 19, 2021 from https://www.learntechlib.org/primary/p/173103/
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
Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2010). Multivariate Data Analysis. 7th Edition, Pearson.
Hickmott, D., Prieto-Rodriguez, E., & Holmes, K. A. (2018). Scoping review of studies on computational thinking in K–12 mathematics classrooms. Digital Experiences in Mathematics Education, 4, 48–69. https://doi.org/10.1007/s40751-017-0038-8
Higgins, H. J., & Wiest, L. R. (2006). Individual interviews as insight into children’s computational thinking. Australian Primary Mathematics Classroom, 11(1), 25–29. Retrieved May 19, 2021 from http://www.aamt.edu.au/Professional-learning/Journals/Journals-Index/Australian-Primary-Mathematics-Classroom2/APMC-11-1-25
Howard, S. J., & Melhuish, E. (2015). An Early Years Toolbox (EYT) for assessing early executive function, language, self-regulation, and social development: Validity, reliability, and preliminary norms. Journal of Psychoeducational Assessment, 33(3), 255–275. https://doi.org/10.1177/0734282916633009
Ishii, H., & Ulmer, B. (1997). Tangible bits: Towards seamless interfaces between people, bits and atoms. Proceedings of CHI’97 ACM March 1997.
Kelley, T. R., & Knowles, J. G. (2016). A conceptual framework for integrated STEM education. International Journal of STEM Education, 3(1), 11. https://doi.org/10.1186/s40594-016-0046-z
Kennedy, T. J., & Odell, M. R. L. (2014). Engaging students in STEM education. Science Education International, 25(3), 246–258. EJ1044508.pdf (ed.gov).
Kermani, H., & Aldemir, J. (2015). Preparing children for success: integrating science, math, and technology in early childhood classroom. Early Child Development and Care, 185(9), 1504–1527. https://doi.org/10.1080/03004430.2015.1007371
Kırçali, A., & Özdener, N. (2022). A comparison of plugged and unplugged tools in teaching engineering to the K-12 level for computational thinking skills. Technology Knowledge and Learning, 1–29. https://doi.org/10.1007/s10758-021-09585-4
Korea Ministry of Education (2015). 2015 대한 학교 과학 교육 개정 학교과목실험 교육 [The 2015 Revised Science Curriculum]. Report no. 2015–74. Korea Ministry of Education.
Lavigne, H. J., Lewis-Presser, A., & Rosenfeld, D. (2020). An exploratory approach for investigating the integration of computational thinking and mathematics for preschool children. Journal of Digital Learning in Teacher Education, 36(1), 63–77. https://doi.org/10.1080/21532974.2019.1693940
Marinus, E., Powell, Z., Thornton, R., McArthur, G., Crain, S., & Conference on International Computing Education. (2018). Unravelling the cognition of coding in 3-to-6-year olds: the development of an assessment tool and the relation between coding ability and cognitive compiling of syntax in natural language. Proceedings of the 2018 ACM Research - ICER ’18, 133–141. https://doi.org/10.1145/3230977.3230984
Mioduser, D., & Levy, S. T. (2010). Making sense by building sense: kindergarten children’s construction and understanding of adaptive robot behaviors. International Journal of Computers for Mathematical Learning, 15(2), 99–127. https://doi.org/10.1007/s10758-010-9163-9

Miyake, A., & Friedman, N. P. (2012). The nature and organization of individual differences in executive functions: Four general conclusions. Current Directions in Psychological Science, 21(1), 8–14. https://doi.org/10.1177/0963721411429458

Moore, T. J., Brophy, S. P., Tank, K. M., Lopez, R. D., Johnston, A. C., Hynes, M. M., & Gadjzik, E. (2020). Multiple representations in computational thinking tasks: a clinical study of second-grade students. Journal of Science Education and Technology, 29(1), 19–34. https://doi.org/10.1007/s10956-020-09812-0

National Research Council. (2011). Successful K-12 STEM education: Identifying effective approaches in science, technology, engineering, and mathematics. National Academies Press. https://doi.org/10.17226/13158

Nationale MINT (STEM) Forum. (2014). MINT-Bildung im Kontext ganzheitlicher Bildung [STEM-education in the context of holistic education]. Herbert Utz Verlag.

NGSS Lead States. (2013). Next generation science standards: For States, by States. The National Academies Press. https://doi.org/10.17226/18290

Papert, S. (1980). Mindstorms: children, computers, and powerful ideas. Basic Books. Retrieved from https://www.basicbooks.com/papert/mindstorms/

Relkin, E. (2018). Assessing young children’s computational thinking abilities (Master’s thesis). Retrieved from ProQuest Dissertations and Theses database. (UMI No. 10813994).

Relkin, E., de Ruiter, L., & Bers, M. U. (2020). TechCheck: Development and validation of an unplugged assessment of computational thinking in early childhood education. Journal of Science Education and Technology, 29, 482–498. https://doi.org/10.1007/s10956-020-09831-x

Rodriguez, B., Kennicut, S., Rader, C., & Cump, T. (2017). Assessing computational thinking in CS unplugged activities. In Proceedings of the 2017 ACM SIGCSE Technical Symposium on Computer Science Education, USA, pp 501–506. https://doi.org/10.1145/3017680.3017779

Royal Society (2017). After the reboot: Computing education in UK schools. Policy Report. Retrieved from http://royalsociety.org/computing-education

Selby, C. C., & Wollard, J. (2013). Computational thinking: The developing definitions. Proceedings of the 45th ACM technical symposium on computer science education. ACM.

Sentence, S. (2018). Bebras cards project: Report on distribution and evaluation of Bebras Cards. King’s College London.

Shute, V. J., Sun, C., & Asbell-Clarke, J. (2017). Demystifying computational thinking. Educational Research Review, 22, 142–158. https://doi.org/10.1016/j.edurev.2017.09.003

STEM Learning (2018). About us. https://www.stem.org.uk/. Accessed 29 Sep 2018

Stohlmann, M., Moore, T. J., & Roehrig, G. H. (2012). Considerations for teaching integrated STEM education. Journal of Pre-College Engineering Education Research (J-PEER), 2(1), 4.

Sung, W., Ahn, J., & Black, J. B. (2017). Introducing computational thinking to young learners: Practicing computational perspectives through embodiment in mathematics education. Technology Knowledge and Learning, 22, 443–463. https://doi.org/10.1007/s10758-017-9328-x

The Partnership for 21st Century Skills (2016). Framework for twenty-first century learning. Retrieved April 30, 2019 from http://www.battelleforkids.org/networks/p21/frameworks-resources

Vizner, M. Z. (2017). Big robots for little kids: investigating the role of Sale in early childhood robotics kits (Master’s thesis). Available from ProQuest Dissertations and Theses database. (UMI No. 10622097).

Voogt, J., Fisser, P., Good, J., Mishra, P., & Yadav, A. (2015). Computational thinking in compulsory education: Towards an agenda for research and practice. Education and Information Technologies, 20(4), 715–728. https://doi.org/10.1007/s10639-015-9412-6

Wang, D., Wang, T., & Liu, Z. (2014). A tangible programming tool for children to cultivate computational thinking. The Scientific World Journal, 2014. 1–10. https://doi.org/10.1155/2014/428080

Wing, J. M. (2006). Computational thinking. Communications of the ACM, 49(3), 33–35.

Wing, J. M. (2011). Computational thinking: What and why? The Link Magazine, 20–23.

Ximénez, C. (2009). Recovery of weak factor loadings in confirmatory factor analysis under conditions of model misspecification. Behavior Research Methods, 41(4), 1038–1052. https://doi.org/10.3758/BRM.41.4.1038

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