

Original Paper

# Computerized Block Games for Automated Cognitive Assessment: Development and Evaluation Study

Xiangyi Cheng<sup>1</sup>, PhD; Grover C Gilmore<sup>2</sup>, PhD; Alan J Lerner<sup>3</sup>, MD; Kiju Lee<sup>4,5</sup>, PhD

<sup>1</sup>Dr Carl D and H Jane Clay Department of Mechanical Engineering, TJ Smull College of Engineering, Ohio Northern University, Ada, OH, United States

<sup>2</sup>Department of Psychological Sciences, Case Western Reserve University, Cleveland, OH, United States

<sup>3</sup>Department of Neurology, Case Western Reserve University, Cleveland, OH, United States

<sup>4</sup>Department of Engineering Technology and Industrial Distribution, Texas A&M University, College Station, TX, United States

<sup>5</sup>J Mike Walker '66 Department of Mechanical Engineering, Texas A&M University, College Station, TX, United States

**Corresponding Author:**

Kiju Lee, PhD

Department of Engineering Technology and Industrial Distribution

Texas A&M University

Fermier Hall, 3367 TAMU

466 Ross St

College Station, TX, 77843

United States

Phone: 1 979 458 6479

Email: [kiju.lee@tamu.edu](mailto:kiju.lee@tamu.edu)

## Abstract

**Background:** Cognitive assessment using tangible objects can measure fine motor and hand-eye coordination skills along with other cognitive domains. Administering such tests is often expensive, labor-intensive, and error prone owing to manual recording and potential subjectivity. Automating the administration and scoring processes can address these difficulties while reducing time and cost. e-Cube is a new vision-based, computerized cognitive assessment tool that integrates computational measures of play complexity and item generators to enable automated and adaptive testing. The e-Cube games use a set of cubes, and the system tracks the movements and locations of these cubes as manipulated by the player.

**Objective:** The primary objectives of the study were to validate the play complexity measures that form the basis of developing the adaptive assessment system and evaluate the preliminary utility and usability of the e-Cube system as an automated cognitive assessment tool.

**Methods:** This study used 6 e-Cube games, namely, Assembly, Shape-Matching, Sequence-Memory, Spatial-Memory, Path-Tracking, and Maze, each targeting different cognitive domains. In total, 2 versions of the games, the fixed version with predetermined sets of items and the adaptive version using the autonomous item generators, were prepared for comparative evaluation. Enrolled participants (N=80; aged 18-60 years) were divided into 2 groups: 48% (38/80) of the participants in the *fixed group* and 52% (42/80) in the *adaptive group*. Each was administered the 6 e-Cube games; 3 subtests of the Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV; Block Design, Digit Span, and Matrix Reasoning); and the System Usability Scale (SUS). Statistical analyses at the 95% significance level were applied.

**Results:** The play complexity values were correlated with the performance indicators (ie, correctness and completion time). The adaptive e-Cube games were correlated with the WAIS-IV subtests ( $r=0.49$ , 95% CI 0.21-0.70;  $P<.001$  for Assembly and Block Design;  $r=0.34$ , 95% CI 0.03-0.59;  $P=.03$  for Shape-Matching and Matrix Reasoning;  $r=0.51$ , 95% CI 0.24-0.72;  $P<.001$  for Spatial-Memory and Digit Span;  $r=0.45$ , 95% CI 0.16-0.67;  $P=.003$  for Path-Tracking and Block Design; and  $r=0.45$ , 95% CI 0.16-0.67;  $P=.003$  for Path-Tracking and Matrix Reasoning). The fixed version showed weaker correlations with the WAIS-IV subtests. The e-Cube system showed a low false detection rate (6/5990, 0.1%) and was determined to be usable, with an average SUS score of 86.01 (SD 8.75).

**Conclusions:** The correlations between the play complexity values and performance indicators supported the validity of the play complexity measures. Correlations between the adaptive e-Cube games and the WAIS-IV subtests demonstrated the potential

utility of the e-Cube games for cognitive assessment, but a further validation study is needed to confirm this. The low false detection rate and high SUS scores indicated that e-Cube is technically reliable and usable.

(*JMIR Serious Games* 2023;11:e40931) doi: [10.2196/40931](https://doi.org/10.2196/40931)

## KEYWORDS

cognitive assessment; computerized block games; adaptive serious games; computerized cognitive assessment

## Introduction

### Background

Cognitive assessment aims to measure multiple domains of cognition, including visuospatial abilities, working memory, language, attention, executive function, fine motor skills, and orientation [1]. One's cognitive abilities affect learning outcomes, physical and mental health, social behavior, and interaction with the environment [2-4]. Identifying impairment in any of these domains, diagnosing the cause, specifying the severity, and tracking the progression of the symptoms are the common purposes of cognitive assessment in clinical settings [5]. This paper presents an innovative technology called *e-Cube* for adaptive, automated cognitive testing and reports the evaluation results in terms of preliminary utility and usability.

There are standardized instruments widely used for cognitive assessment. The Wechsler Adult Intelligence Scale (WAIS) has been broadly adopted in clinical, research, and educational settings and is often referred to as a *gold* standard [6]. The WAIS Fourth Edition (WAIS-IV) is normed for the ages of 16 to 90 years. It comprehensively assesses cognitive abilities using 15 subtests that target various cognitive domains [7]. This instrument is administered and scored by a qualified psychologist, taking approximately 60 to 90 minutes. This process is labor-intensive and costly [8]. The Stanford-Binet Intelligence Scales, Fifth Edition, is another standardized instrument commonly used in both clinical and research settings [9]. Several WAIS and Stanford-Binet Intelligence Scales subtests rely heavily on a person's verbal skills and, therefore, show limitations when administered using a non-native language version [10]. There are also nonverbal instruments aiming to eliminate cultural and language biases in the assessment. For example, the Raven Progressive Matrices consist of 60 items measuring the basic cognitive functioning of individuals, each of which is a visual geometric design with a missing piece [11,12].

The advancements in digital technologies have enabled researchers to explore computer-based methods for cognitive assessment. Computer-based methods can reduce the administrative burden, automate the scoring process, reduce cheating, and standardize test conditions once successfully validated [13]. A straightforward application is to convert a paper-and-pencil test into a computerized version while retaining the contents and formats. Q-interactive is a digital system initially developed for the WAIS-IV that uses 2 iPads, one for the administrator and the other for the test taker [14]. This digital version reduces labor-intensity but takes approximately the same time for a trained professional. Moreover, it can only automate some types of tests. In particular, one of the subtests, Block Design (BD), requires the examinee to assemble physical

blocks to match the top surface with a given image displayed on an iPad. The administrator then has to check the correctness and input the results manually. In addition to the computerization of existing instruments, an increasing body of research has adopted the concept of computer- or tablet-based *serious games* to make the experience more engaging [15-17]. Some serious games use dynamic difficulty adjustment to achieve adaptive testing by tuning item difficulty autonomously [18,19]. However, most of the previously developed games for cognitive assessment do not include measurements of fine motor and hand-eye coordination skills.

### Cognitive Assessment Using Tangible Objects

Cognitive assessment sometimes uses tangible objects to measure one's cognitive skills together with fine motor and hand-eye coordination skills. These skills are closely linked to many neurological diseases and brain injuries [20]. Existing research also suggests that the deterioration of fine motor control and coordination characterizes sensorimotor deficiencies in mild cognitive impairment and Alzheimer disease [15,21-23]. The BD subtest in the WAIS [5] and the Kohs Block Design test [24] use a set of cubes and require an examinee to place and assemble the top surfaces of the blocks to match the given image. Unlike the simple multiple-choice questions used in many other assessment instruments, the administrator has to inspect the correctness of the block manipulation visually while timing in these tests. This is labor-intensive and error prone owing to manual recording and subjectivity, possibly affecting the assessment results.

Automating the assessment using physical objects also involves additional challenges and requires technological innovations beyond what is expected for computerized tests. For example, a platform called ETAN supports the use of tangible user interfaces and physical objects for evaluating visuospatial cognition by implementing the Baking Tray Task [25]. Cognitive Cubes were designed to assess spatial and constructive abilities by asking users to build 3D shapes with the cubes. A pilot study involving 16 participants demonstrated that the Cognitive Cubes were sensitive to differences in cognitive ability [20].

SIG-Blocks and TAG-Game, developed for the automated assessment of cognitive and fine motor skills, were the previous research of this work [26-28]. Each SIG-Block, covered with simple black-and-white geometric shapes, can sense physical motions applied to it, detect adjacent blocks, and send sensor data to a local host computer in real time. TAG-Games are computerized games that use SIG-Blocks as a means of game control. In total, 3 types of TAG-Games, namely, Assembly, Shape-Matching, and Memory, were designed and tested. These games are all nonverbal and require hand manipulation of physical blocks. The TAG-Game technology is one of the few

systems that can automate the administration and data collection of tasks involving physical object manipulation. However, despite its demonstrated potential, several challenges were identified in our previous research. Specifically, hardware costs, occasional technical failure, and high maintenance make the system unsuitable for broad and long-term adoption and use.

### e-Cube Games for Automated Assessment of Fine Motor and Cognitive Skills

e-Cube is our latest technical innovation that converts the original TAG-Game system into a computer vision-based system using a set of plastic cubes and a webcam. The e-Cube system reduces the device cost from US \$1500 to approximately US \$50 (excluding a computing device needed for any computerized assessment), decreases potential technical errors, and nearly eliminates the maintenance burden. The entire system is fully autonomous and easy to use. In addition to these benefits, a new *adaptive* test environment was established based on the embedded algorithms for measuring play complexity and generating adaptive test items autonomously. These features enable personalized assessment based on an individual's real-time performance. e-Cube consists of 6 types of games: (1) Assembly, (2) Shape-Matching, (3) Sequence-Memory, (4) Spatial-Memory, (5) Path-Tracking, and (6) Maze. The first 3 were directly adopted and converted from TAG-Games, and the other 3 were newly created. New computational measures of play complexity were defined and implemented for each game.

The evaluation focused on testing 2 objectives. *Objective 1* was to validate the proposed play complexity measures that form

the algorithmic basis of the adaptive games. Correlation analyses were performed between the developed play complexity measures and 2 performance indicators, mean correctness and mean completion time. *Objective 2* was to understand the preliminary utility and usability of the e-Cube system as an automated cognitive assessment tool. The non-age-corrected raw scores of 3 WAIS-IV subtests—BD, Digit Span (DS), and Matrix Reasoning (MR)—were adopted to compare their results with the e-Cube game scores. The WAIS-IV is a well-established instrument, and the 3 selected subtests measure the target cognitive domains of the e-Cube games. Specifically, the Assembly game was conjectured to be related to BD as both require the assembly of block surfaces to match a given pattern. The Shape-Matching game requires the participant to find a shape that completes a pattern, so it was expected to tap the same cognitive abilities as the MR subtest. Sequence-Memory and Spatial-Memory were expected to be related to DS as they all target working memory skills. The remaining games, Path-Tracking and Maze, are timed games asking participants to give the shortest trajectory by reasoning, so they were both hypothesized to show a relationship with BD and MR. The *hypothesized relationships* between the e-Cube games and the WAIS subtests are summarized in Table 1. The false detection rate of the system determines whether it produces reliable and accurate data. Usability was evaluated by administering the System Usability Scale (SUS) to all participants upon the completion of the assessment session. The SUS is a 10-item questionnaire measuring usability with high validity and reliability and, thus, used as a measure of perceived usability [29-31].

**Table 1.** The 6 e-Cube games with their associated task descriptions and the expected associations with the Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV), subtests (Block Design [BD], Digit Span [DS], and Matrix Reasoning [MR]).

e-Cube game	Task	WAIS-IV		
		BD	DS	MR
Assembly	Assemble multiple cubes to match the top assembly configuration	✓		
Shape-Matching	Manipulate 1 cube to complete the pattern with 1 missing piece			✓
Sequence-Memory	Memorize a sequence of geometric shapes and reconstruct it using 1 cube		✓	
Spatial-Memory	Memorize a spatial assembly of geometric shapes and reconstruct it using cubes		✓	
Path-Tracking	Trace a connected path between 2 blue dots using a single cube	✓		✓
Maze	Navigate through a maze to reach a goal point from a starting point using a single cube	✓		✓

## Methods

### e-Cube Games

#### System Overview

The e-Cube system consists of a set of 9 cubes with 1.2-inch-length edges, a place mat with a brown rectangular region in the center, a computing device with a display, and a webcam with a custom-designed stand (Figure 1). The cube's 6 faces are distinctive black-and-white geometric shapes, including squares, strips, and triangles representing 4-, 2-, and 1-fold rotational symmetry (Figure 2). The cubes preserve the same design as the SIG-Blocks [28]. When the system turns on, the camera automatically detects the corners of the brown

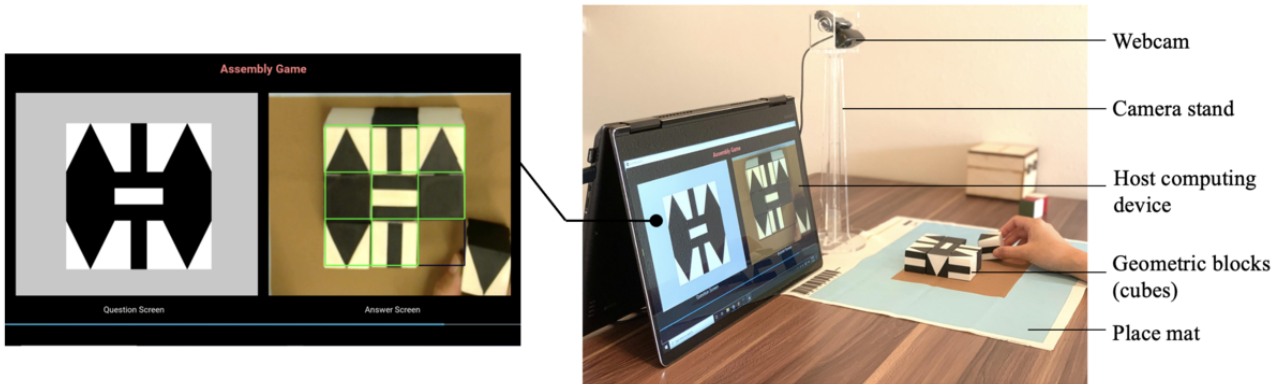
rectangular area on the place mat. This area is called the *play area*, where the cubes are expected to be placed and manipulated. The laptop with the connected webcam displays the cubes in the play area after perspective transformation projecting the original camera view onto a 2D plane and tracks their movements in real time [32]. This autonomous transformation offers flexibility in the camera location.

The e-Cube system requires the accurate identification of the top-surface images of the cubes. Individual box-shaped regions are assigned for placing the cubes (Figure 1), wherein the geometric shape and orientation detection algorithm is executed. The embedded algorithm first detects the black-and-white regions within each box to check whether a cube exists. It then identifies a polygon using the Ramer-Douglas-Peucker algorithm

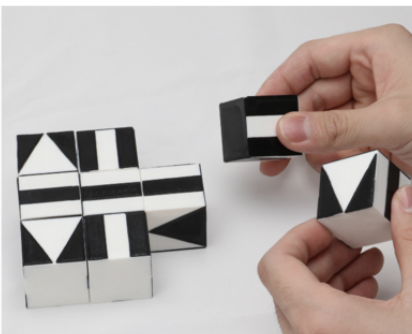
[33]. Finally, the specific shape and orientation of the detected polygon are determined. This simple strategy makes the system robust and reliable under different illumination conditions and

limited computing capabilities. The e-Cube system can run reliably on a relatively low-end computing device, such as Intel Core i5-7200U (2.5 GHz, 3 M cache, dual core, and 4 threads).

**Figure 1.** The hardware of e-Cube, consisting of the cubes, a webcam with a stand, a place mat, and a host computing device running the Assembly game.



**Figure 2.** A total of 9 geometric cubes and 14 distinctive surface shapes with their IDs formed by rotating the images on the 6 surfaces of a cube by 0°, 90°, 180°, and 270°.



Index	1	2	3	4	5	6	7
IDs (geometry, color, orientation)	(0,0,0)	(0,1,0)	(1,0,0)	(1,0,1)	(1,1,0)	(1,1,1)	(2,0,0)
Shape							
Index	8	9	10	11	12	13	14
IDs (geometry, color, orientation)	(2,0,1)	(2,0,2)	(2,0,3)	(2,1,0)	(2,1,1)	(2,1,2)	(2,1,3)
Shape							

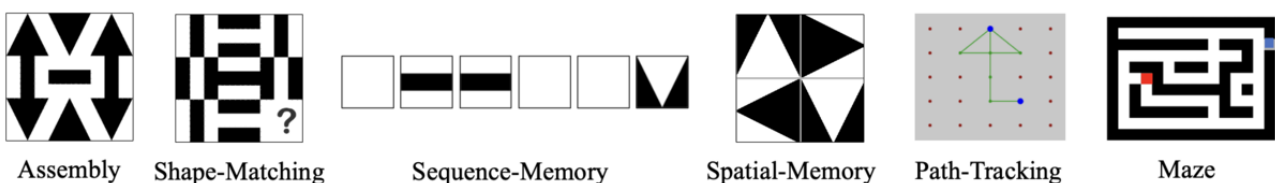
### Game Design

We developed 6 e-Cube games: Assembly, Shape-Matching, Sequence-Memory, Spatial-Memory, Path-Tracking, and Maze. We directly converted the 3 TAG-Games (Assembly, Shape-Matching, and Memory) into the vision-based e-Cube versions (Assembly, Shape-Matching, and Sequence-Memory) [26]. Spatial-Memory, Path-Tracking, and Maze were newly added.

Table 1 presents the tasks associated with each game, and Figure 3 shows an example item for each e-Cube game. Assembly asks the player to match the given assembly figure displayed on the screen using 4 or 9 cubes, similar to the BD subtest of the WAIS-IV. Shape-Matching involves items with assembly patterns, each missing 1 piece, and the player completes the pattern using a single cube. Sequence-Memory and Spatial-Memory require the player to memorize a sequence or

an assembly of geometric shapes. In Sequence-Memory, each shape is displayed for 1 second and then disappears. In Spatial-Memory, an assembly pattern of 2, 3, or 4 geometric shapes is displayed for 5 seconds. The items in the Spatial-Memory game are similar to those in Assembly, whereas visible outlines around individual shapes are added, as shown in Figure 3, to assist perceptual segmentation of the pattern [34]. Path-Tracking and Maze use only 1 cube with its white square facing up. In these 2 games, the vision algorithm detects the center of the white square and tracks it continuously on the screen; no assigned box-shaped regions are shown on the screen. Path-Tracking displays a green connected path between 2 blue dots, and the player must trace the path by moving the cube from one blue dot to the other on a 5 × 5 grid via the shortest path. The Maze game asks the player to find the shortest path of mazes shown on the screen by moving a cube from the start (blue) to the end (red).

**Figure 3.** Sample items for the 6 games.





## Computational Measures of Play Complexity

### Overview

The e-Cube system aims to dynamically adapt to individual differences in cognitive skills by generating test items autonomously based on one's real-time performance. To do so, a computational method to measure the difficulty of each item is required. The previously defined measures of play complexity presented in the studies by Lee et al [26,35] and Jeong et al [28] were highly correlated with the participants' performances measured using completion time or accuracy. These measures

**Figure 4.** Items formed by the same geometric shapes but with different play complexity (with identical compositional complexity but different configurational complexity).



To address this limitation, we defined new complexity measures for the 6 e-Cube games. Two mathematical concepts were applied: (1) the Shannon entropy and (2) the gray-level co-occurrence matrix (GLCM). The Shannon entropy measures the uncertainty, randomness, or disorder existing in the data [36] and is calculated as

$$H = - \sum_{i=1}^n p_i \log_2 p_i$$

where  $p_i$  is the probability of the  $i$ th event. When the probabilities are evenly distributed, the Shannon entropy is calculated as  $H = \log_2 n$ . The GLCM was originally proposed to classify image texture in grayscale [37,38]. For an image with an  $m \times n$  dimension and  $L$  gray level, the GLCM of the image ( $f$ ) is defined as an  $L \times L$  square matrix such that

$$f(i, j) = \sum_{x=1}^n \sum_{y=1}^m \begin{cases} 1, & \text{if } J(x, y) = i \text{ \& } J(x + \Delta x, y + \Delta y) = j \\ 0, & \text{otherwise} \end{cases} \quad (1)$$

where  $\Delta x$  and  $\Delta y$  are typically defined as the horizontal, vertical, or diagonal position differences between the 2 adjacent pixels [37]. Horizontally adjacent pixels can be paired along  $0^\circ$  or  $180^\circ$ ; vertically adjacent pixels can be paired along  $90^\circ$  or  $270^\circ$ ; and diagonally adjacent pixels can be paired along  $45^\circ$ ,  $135^\circ$ ,  $225^\circ$ , or  $315^\circ$ . On the basis of the Shannon entropy and the GLCM, the computational measures of play complexity for the 6 e-Cube games are defined in the following sections.

### Play Complexities of Assembly, Sequence-Memory, Spatial-Memory, and Shape-Matching

The play complexity of the items in Assembly, Sequence-Memory, and Spatial-Memory is computed using

$$C_{play} = k \cdot C_{compos} + C_{config} \quad (2)$$

where  $C_{compos}$  represents the compositional complexity associated with individual shapes (ie, the number of shapes and their rotational symmetry),  $C_{config}$  captures the configurational complexity associated with the orientation and color differences among the shapes in the way that they are arranged, and  $k$  is a sigmoid function defined as

captured the complexities associated with individual geometric shapes without considering the spatial complexity of the assembly patterns. For example, the 3 assembly patterns shown in Figure 4 had the same complexity value using the previously defined measures. As our previous study used a handcrafted set of items, we could select the items where their difficulties could be clearly differentiated using the previously defined measures. However, for generating adaptive test items, the complexity measures must capture the difficulties associated with both the individual shapes and the assembly patterns.

$$k = \frac{1}{1 + e^{-1.5 C_{config}}}$$

If  $C_{config}$  is small, a small  $k$  leads to a lower impact of  $C_{compos}$  on  $C_{play}$ . For example, if an item is formed only by identical triangles (large  $C_{compos}$  and small  $C_{config}$ ),  $C_{play}$  will still be small owing to  $k$ .

The Shannon entropy forms the basis of  $C_{compos}$  such that

$$C_{compos} = \sum_{i=1}^Q (\log_2 m_i + \log_2 r_i) \quad (3)$$

where  $Q$  is the total number of shapes in the item,  $m_i$  is the number of available distinctive shapes among the 6 faces of a cube ( $m_i=6$  if all faces of a cube are different), and  $r_i$  is the number of distinctive orientations obtained by rotating this shape  $90^\circ$  ( $r_i=1$  for squares, 2 for strips, or 4 for triangles). The 3 images in Figure 4 have the same  $C_{compos}$  value.

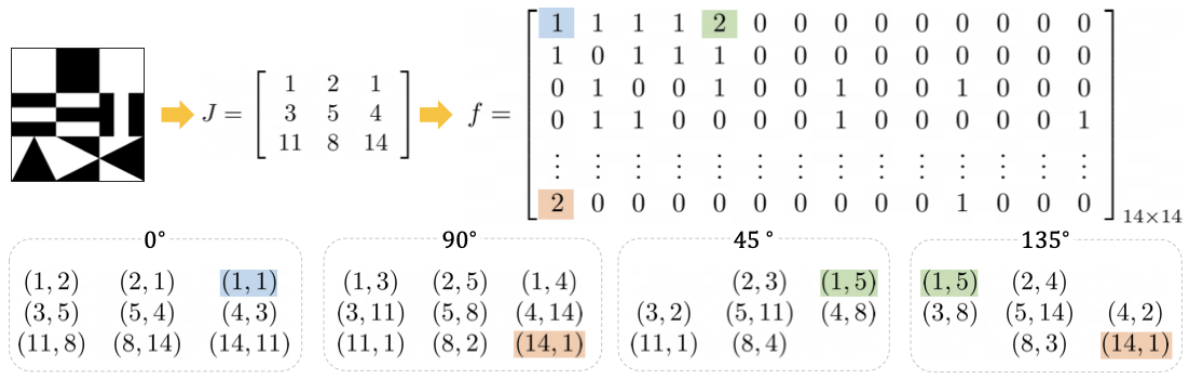
The GLCM was adopted for capturing the configurational disorder ( $C_{config}$ ) [39]. Figure 5 illustrates how it is obtained for an Assembly item. First, all geometric shapes used in each item are represented as  $J$  with the indexes corresponding to each shape defined in Figure 2 and its location. Second, all adjacent pairs along the  $0^\circ$ ,  $45^\circ$ ,  $90^\circ$ , and  $135^\circ$  directions on  $J$  are extracted. For example, the second row in  $J$  is (3, 5, 4), and the ordered pairs along  $0^\circ$ , including the circulant pair, are (3, 5), (5, 4), and (4, 3). Once all pairs are obtained, the number of each pair is imposed on the location in a 14-by-14 matrix  $f$ , that is, the GLCM, following equation 1. As shown in Figure 5, there are two (1, 5) pairs that correspond to 2 in the (1, 5) coordinate in  $f$  and one (1, 1) pair that corresponds to 1 in the (1, 1) coordinate. The weighted entropy [40] based on  $f$  is then calculated using

$$C_{config} = - \sum_{i=1}^{14} \sum_{j=1}^{14} w_{i,j} p_{i,j} \log_2 p_{i,j} \quad (4)$$

where

$$p_{i,j} = \frac{f(i, j)}{\sum_{i=1}^{14} \sum_{j=1}^{14} f(i, j)}$$

Figure 5. Gray-level co-occurrence matrix computation of a given 3 × 3 Assembly item.



The weight  $w_{i,j}$  estimates the configurational complexity of 2 adjacent elements based on their colors and orientations. To compare the differences among these 14 distinctive shapes, 3 IDs were assigned to each shape to categorize its geometric shape (square, strip, or triangle), color, and orientation (Figure 2). Regarding the IDs, 2 was assigned to the weight if the 2 adjacent shapes had different colors and orientations, and 1 was assigned otherwise.

In Shape-Matching, as the player was asked to find a single shape that best completed the pattern, more shapes used in the pattern do not necessarily indicate greater difficulty in the pattern. Therefore, we only used the configurational complexity to estimate the item difficulty such that  $C_{play} = C_{config}$ , where  $C_{config}$  is defined as the summation of the weighted entropies based on the 3 GLCMs estimating how frequently a pair occurs horizontally, vertically, and diagonally.

**Play Complexities of Path-Tracking and Maze**

Path-Tracking and Maze do not use the geometric shapes of the cubes and, instead, use a single cube for creating a path. Therefore, the aforementioned method is not applicable. The play complexity of Path-Tracking adopts the network complexity based on the Shannon entropy [41], given by

$$C_{play} = \sum_{i=1}^V a_i \log_2 a_i \quad (5)$$

where  $V$  is the number of vertices and  $a_i$  is the associated vertex degree. For Maze, the play complexity is defined as

$$C_{play} = 0.4\{C_m + 10(C_s + C_l)\} \quad (6)$$

where  $C_m$  is the maze complexity using equation 5 and  $C_s$  and  $C_l$  are calculated using

$$C_s = \sum_{i=1}^L s_i \log_2 s_i; \quad C_l = \log_2 L \quad (7)$$

$C_m$  reflects the complexity of the maze itself, but the complexity of solving a maze should also consider the start and end locations. The solution logarithmic complexity ( $C_s$ ) in equation 7 represents the complexity caused by the vertex degrees, where  $L$  is the total length of the shortest path solved by the A\* algorithm [42] and  $s_i$  is the degree of each grid in this solution. The solution length complexity  $C_l$  in equation 7 captures the length of the shortest path. In equation 6, the 0.4 value is

multiplied to make the complexity values comparable with those of other e-Cube games. In addition,  $C_s + C_l$  is multiplied by 10 to balance with the range of  $C_m$ .

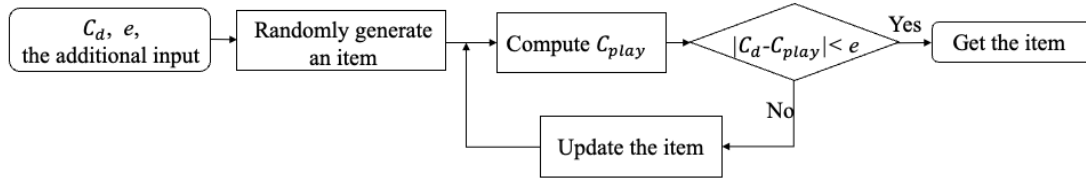
**Adaptive Game Generators**

The computational measures of play complexity form the basis of the adaptive algorithms, which can automatically generate test items. On the basis of the concept of dynamic difficulty adjustment, we created an adaptive e-Cube system that can adjust the item difficulty based on a player’s performance measured using correctness.

The game begins with an item with a predefined low complexity. If the player answers the first item correctly, it proceeds to the next item with a higher complexity; otherwise, a new item with the same complexity is generated. If 2 consecutive incorrect answers are received, the complexity reverts to the midpoint between the latest correctly answered item complexity and the current incorrectly answered item complexity. The difference between the current and the next complexity value is referred to as a step size that can be either positive, 0, or negative. The game ends at a predefined highest complexity level or when the absolute value of the step size becomes sufficiently small.

The item generators for all games except for Shape-Matching follow a similar process, shown in Figure 6. The system takes a desired play complexity value  $C_d$  and a small tolerance  $e$  as input and generates a new item with a complexity  $C_{play}$ , where  $|C_d - C_{play}| \leq e$ . Apart from  $C_d$  and  $e$ , additional input is needed in these games except for Maze. This input is the dimension of the pattern (eg, 2 × 2 or 3 × 3) in Assembly and Spatial-Memory, the number of images to be displayed in Sequence-Memory, and the number of dots to be connected (referred to as nodes) in Path-Tracking. The following steps generate items: (1) the system randomly generates an item based on the inputs; (2) the absolute difference between  $C_d$  and  $C_{play}$  is computed, where  $C_{play}$  is the complexity of the current item computed using the proposed measure; (3a) if the absolute difference is smaller than  $e$ , the system outputs the current item and ends the process; and (3b) if the absolute difference is not smaller than  $e$ , the system updates one feature of the current item to make the item easier or harder and then goes back to step 2. The features of the item can be geometric shapes in Assembly, Sequence-Memory, and Spatial-Memory; the paths connected by nodes in Path-Tracking; or the position of the end point in Maze.

**Figure 6.** The flowchart of the item generators for Assembly, Sequence-Memory, Spatial-Memory, Path-Tracking, and Maze.



Shape-Matching uses assembly configurations with embedded patterns where the types of patterns are predefined in the item generator, such as symmetry and rotation. Shape-Matching generates items from a predefined pool. For example, the easiest pattern in the predefined pool is formed by 4 identical shapes, in which one of the shapes will be hidden from players and treated as the missing piece. The item generator for Shape-Matching randomly selects a shape to form the easiest pattern, which leads to different items with the same play complexity.

**Evaluation of e-Cube**

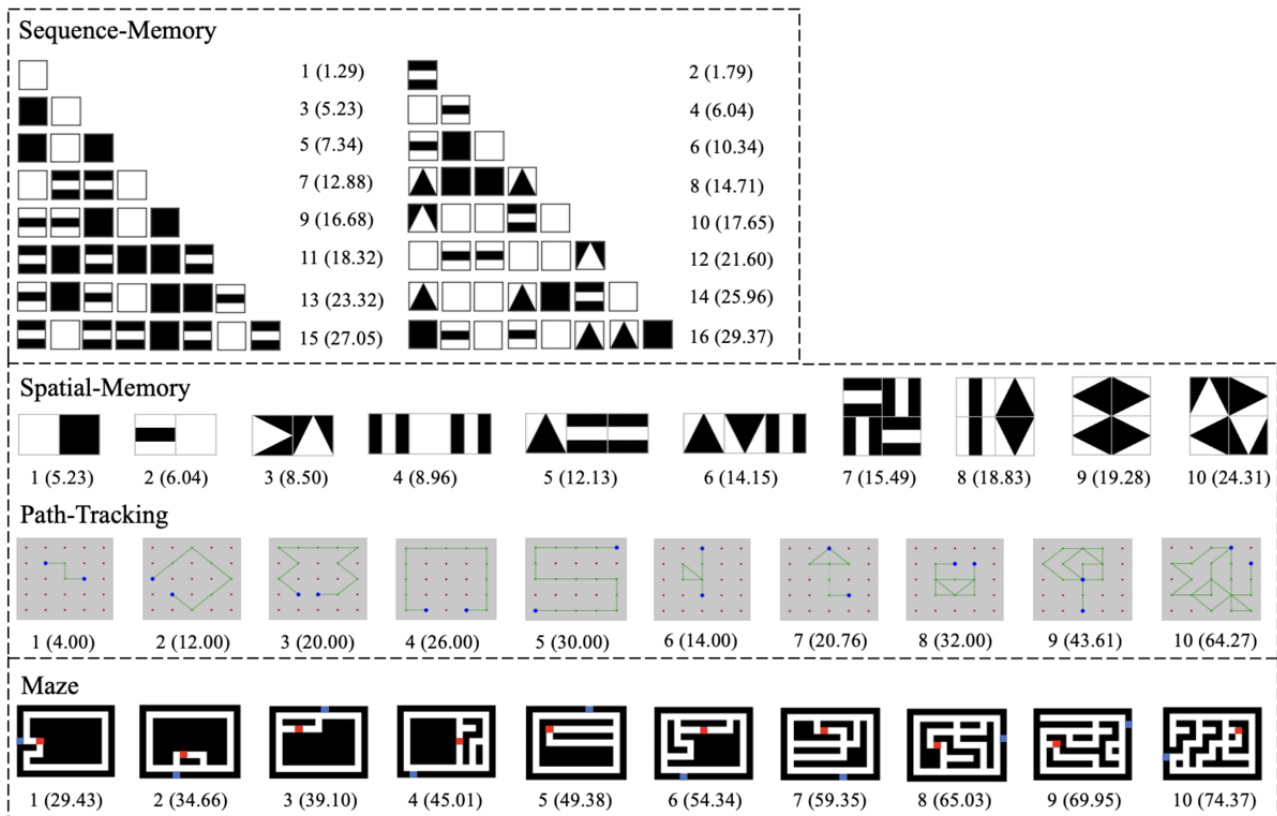
The evaluation study focused on the preliminary validation of (1) the proposed play complexity measures that form a basis for developing adaptive games (objective 1) and (2) the

preliminary utility and usability of the e-Cube system as an automated cognitive assessment tool (objective 2).

**Materials and Methods**

The study used 2 versions of e-Cube games: e-Cube with the item generators (called *adaptive e-Cube*) and e-Cube with fixed items (called *fixed e-Cube*). Each participant was assigned to 1 of the 2 groups to experience the adaptive or fixed e-Cube games (ie, adaptive group and fixed group). The fixed games provided the same items for each player, whereas the adaptive games offered different items and different numbers of items based on the players’ performance. The fixed versions of Assembly and Shape-Matching used the same items as in the study by Lee et al [26]. The fixed items of the rest of the games are shown in Figure 7.

**Figure 7.** Items with their ordering numbers and play complexity values in the fixed version of Sequence-Memory, Spatial-Memory, Path-Tracking, and Maze.



Objective 1 was tested by performing correlation analyses between the play complexity measures and performance indicators, including the mean correctness and mean completion time obtained by the participants for individual items in the fixed group. For objective 2, the correlations between the raw scores of 3 WAIS-IV subtests (BD, DS, and MR) and the 6 e-Cube games were analyzed to understand their relationships.

We also investigated the technical reliability of the system using the false detection rate and usability based on the SUS results.

**Protocol and Recruitment**

This human participant study took place at Texas A&M University (TAMU). Bulk recruitment emails were sent to TAMU communities, and flyers were placed in buildings within

the university for recruiting healthy participants aged 18 to 64 years. Once potential participants contacted the research team, a prescreening survey was sent via email to self-identify their eligibility before scheduling a visit. The prescreening survey consisted of 4 questions on age, date of birth, sex, and health conditions. Individuals who were beyond the target age range or had any of the following health conditions were excluded: stroke, other neurological diseases, low vision or blindness with aid, hearing loss or deafness with aid, or difficulties in arm or hand movements for manipulating small objects.

The sample size of a main trial is usually determined through a power analysis, where the variance is known from previous or pilot studies [43]. However, for this preliminary study, we applied the simplest method—sample size rules of thumb, which recommended samples of a minimum of 70 (35 per group) in pilot studies [44,45]. In our study, 80 participants ( $n=47$ , 59% male) were recruited and screened. All (80/80, 100%) were eligible and, thus, enrolled in the study. Informed consent and background information (ie, age and sex) were obtained from each participant. Most of the participants were randomly assigned to either the fixed or adaptive group, whereas efforts were made to balance the sex and age distribution between the

2 groups when we placed the participants in the groups toward the end. The fixed group included 48% (38/80) of the participants (23/38, 61% male), and the adaptive group included 52% (42/80) of the participants (24/42, 57% male). Owing to the convenience of recruitment and proximity to the study location, most participants were students from various departments and programs across the TAMU College Station campus, whereas several faculty and staff members, alumni of the university, and a few residents also participated. As a result, 82% (66/80) of the participants were aged between 18 and 30 years, 9% (7/80) were aged between 31 and 40 years, 2% (2/80) were aged between 41 and 50 years, and 6% (5/80) were aged between 51 and 60 years. There were no participants aged >60 years. Age mean, SD, and IQR; age distribution; and sex distribution are summarized in Table 2. We applied a chi-square test at a 95% confidence level to determine if there were differences in sex and age distribution between the 2 groups. The results showed no difference in the proportions of male, female, and intersex participants in the groups ( $\chi^2_2=0.1$ ,  $P=.76$ ) and no difference in the proportion of age in the groups ( $\chi^2_4=2.4$ ,  $P=.50$ ).

**Table 2.** Participant demographic data (N=80).

Characteristic	Participants	
	Fixed group (n=38)	Adaptive group (n=42)
Age (years), mean (SD; IQR)	26.71 (9.24; 22.00-28.00)	25.74 (8.50; 20.00-27.25)
<b>Age range (years), n (%)</b>		
18-30	31 (39)	35 (44)
31-40	4 (5)	3 (4)
41-50	0 (0)	2 (2)
51-60	3 (4)	2 (2)
>60	0 (0)	0 (0)
<b>Sex, n (%)</b>		
Male	23 (29)	24 (30)
Female	15 (19)	18 (22)
Intersex	0 (0)	0 (0)

The administration order between WAIS-IV and e-Cube was randomized. The order of the 3 subtests of the WAIS-IV followed the standardized protocol (BD, DS, and MR), whereas the order of the 6 e-Cube games was randomized. Upon the completion of both tests, the SUS was administered to each participant. The entire session took approximately 90 minutes: 50 minutes for e-Cube, 25 minutes for the WAIS-IV subtests, 5 minutes for the SUS, and a 10-minute break between e-Cube and the WAIS-IV subtests. Each participant was given a US \$10 gift card upon the completion of participation.

### Scoring System

The e-Cube games have not been standardized yet, and therefore, scoring methods are not finalized at this stage. We benchmarked the scoring methods used for the WAIS-IV subtests and our previous study [26] and modified them to suit the e-Cube games.

The scoring of Assembly considers correctness, item size, and completion time—a  $2 \times 2$  item that is correctly completed within 15 seconds or between 15 and 30 seconds yields 3 or 2 points, respectively; a  $3 \times 3$  item correctly completed within 30 seconds, between 30 and 40 seconds, or between 40 and 60 seconds results in 4, 3, or 2 points, respectively. Shape-Matching, Sequence-Memory, and Spatial-Memory use correctness only as the scoring criteria—2 points for each correct answer and 0 for an incorrect answer. Scoring methods for Path-Tracking and Maze are based on correctness, completion time, and whether the path taken is the shortest. For Path-Tracking, the shortest path finished within 20 seconds, between 20 and 40 seconds, or between 40 and 80 seconds yields 4, 2, or 1 points, respectively; a correct path, but not the shortest, completed within 20 seconds or between 20 and 40 seconds yields 2 or 1 points, respectively. For Maze, the shortest path completed



within 10 seconds, between 10 and 20 seconds, or between 20 and 40 seconds yields 4, 2, or 1 points, respectively; a correct path, but not the shortest, completed within 10 seconds or between 10 and 20 seconds yields 2 or 1 points, respectively. Others not satisfying the aforementioned conditions result in 0 points.

The adaptive e-Cube games require some additional considerations for scoring. If an item is generated with the same play complexity as the previous one answered incorrectly, the score for the correct answer is 1 point less than the score used in the fixed version. A total of 2 consecutive incorrect answers result in the system generating an easier item, and in this case, a correct answer for that newly generated item yields only 1 point.

### Statistical Analysis

Correlations were computed to determine the relationships between the computed complexity values and participants' performance, the connections between the WAIS subtests and the e-Cube games, and the relationships among the 6 e-Cube games. We used the Spearman correlation to measure the monotonic association among them. The correlation is interpreted as "weak," "moderate," and "strong/high" when the coefficient is  $<0.36$ , between  $0.36$  and  $0.67$ , and  $>0.67$ , respectively [46]. We used 2-tailed  $t$  tests to identify the mean differences in the game or subtest scores and the SUS scores between the 2 groups.

### Ethics Approval and Informed Consent

This human participant study was reviewed and approved by the TAMU Institutional Review Board (IRB2019-1079D; approval date: December 22, 2020). Informed consent was obtained from all participants before taking part in this study.

## Results

All enrolled participants (80/80, 100%) completed the entire session without withdrawal. The results and findings for objectives 1 and 2 are presented in the following sections.

### Objective 1: Evaluation of the Measures of Play Complexity

The preliminary validity of the proposed play complexity measure ( $C_{play}$ ) was evaluated by analyzing the correlations between the  $C_{play}$  values and the performance indicators from the fixed group participants. If the defined complexity measures properly reflected the difficulty associated with the individual items, participants would perform worse on the items with higher complexity values. Two performance indicators were used to evaluate the play complexity measures: (1) mean correctness and (2) mean completion time obtained for each item from the fixed group participants. The correlation analyses were performed at a 95% confidence level between the  $C_{play}$  values and all the mean values. The correlation coefficients  $r$  with  $P$  values and 95% CIs are shown in Table 3.

The  $C_{play}$  values showed strong positive correlations with the mean completion time in all e-Cube games, indicating that the items with higher  $C_{play}$  yielded a longer time to answer. Negative correlations between the  $C_{play}$  values and the mean correctness were found in Assembly, Shape-Matching, and Sequence-Memory, indicating that higher  $C_{play}$  items yielded lower accuracies. In Spatial-Memory, we found no substantial correlation between the mean correctness and  $C_{play}$ , mainly because of items 8 and 9 (Figure 7). The symmetry in these items seemed to make them easy to memorize, whereas it was not taken into account for the defined complexity measures. Without these 2 items, a correlation was found as  $r_s = -0.67$  (95% CI  $-0.93$  to  $0.066$ ;  $P = .06$ ). A few participants correctly answered all the items in Path-Tracking (12/38, 32%;  $P = .58$ ) and Maze (31/38, 82%;  $P = .07$ ), and therefore, correctness did not yield any significant correlation with  $C_{play}$ .

**Table 3.** Correlations (Spearman  $r$ , 2-tailed  $P$  value, and 95% CIs) between the  $C_{play}$  values and the mean correctness and mean completion time for each item from the fixed group participants.

Game ( <i>df</i> )	Mean completion time	Mean correctness
<b>Assembly (20)</b>		
<i>r</i>	<i>0.86</i> <sup>a</sup>	<i>-0.50</i>
<i>P</i> value	<i>&lt;.001</i>	<i>.02</i>
95% CI	<i>0.67 to 0.94</i>	<i>-0.46 to 0.43</i>
<b>Shape-Matching (10)</b>		
<i>r</i>	<i>0.95</i>	<i>-0.75</i>
<i>P</i> value	<i>&lt;.001</i>	<i>.009</i>
95% CI	<i>0.80 to 0.99</i>	<i>-0.94 to -0.23</i>
<b>Sequence-Memory (16)</b>		
<i>r</i>	<i>0.98</i>	<i>-0.95</i>
<i>P</i> value	<i>&lt;.001</i>	<i>&lt;.001</i>
95% CI	<i>0.94 to 0.99</i>	<i>-0.98 to -0.86</i>
<b>Spatial-Memory (10)</b>		
<i>r</i>	<i>0.94</i>	<i>-0.30</i>
<i>P</i> value	<i>&lt;.001</i>	<i>.30</i>
95% CI	<i>0.76 to 0.99</i>	<i>-0.78 to 0.41</i>
<b>Path-Tracking (10)</b>		
<i>r</i>	<i>0.82</i>	<i>-0.20</i>
<i>P</i> value	<i>.002</i>	<i>.58</i>
95% CI	<i>0.39 to 0.96</i>	<i>-0.74 to 0.49</i>
<b>Maze (10)</b>		
<i>r</i>	<i>0.72</i>	<i>-0.59</i>
<i>P</i> value	<i>.01</i>	<i>.07</i>
95% CI	<i>0.17 to 0.93</i>	<i>-0.89 to 0.06</i>

<sup>a</sup>Italics indicate that a correlation existed.

## Objective 2: Evaluation of Preliminary Utility and Usability of e-Cube Games for Cognitive Assessment

### Overview

The mean, SDs, and IQR values obtained from participants in each group for the WAIS-IV subtests (raw scores) and e-Cube games are summarized in Table 4. We also conducted a 2-tailed

$t$  test with equal variance (Cronbach  $\alpha=.05$ ) comparing the test scores from the adaptive and fixed groups to determine whether significant differences existed in mean scores between the 2 groups. The  $t$  test showed no significant differences in the mean scores of the 3 WAIS subtests—BD ( $P=.37$ ), MR ( $P=.06$ ), and DS ( $P=.18$ )—between the 2 groups. Group differences were found in Shape-Matching and Sequence-Memory, but not in other e-Cube games.

**Table 4.** Score statistics from the Wechsler Adult Intelligence Scale, Fourth Edition (WAIS-IV), subtests and e-Cube games.

	Fixed group, mean (SD; IQR)	Adaptive group, mean (SD; IQR)	2-tailed <i>t</i> test ( <i>df</i> )	<i>P</i> value
<b>WAIS-IV raw score</b>				
BD <sup>a</sup>	50.95 (11.26; 41.00-60.00)	53.02 (9.35; 47.75-60.00)	-0.90 (78)	.37
DS <sup>b</sup>	28.34 (4.86; 24.75-32.00)	29.98 (5.83; 26.00-34.00)	-1.36 (78)	.18
MR <sup>c</sup>	21.55 (2.45; 11.00-14.00)	22.62 (2.49; 21.00-24.00)	-1.93 (78)	.06
<b>e-Cube score</b>				
Assembly	54.37 (11.10; 48.75-62.00)	58.52 (12.52; 51.75-67.25)	-1.56 (78)	.12
Shape-Matching	<i>16.53 (1.84; 16.00-18.00)<sup>d</sup></i>	<i>14.67 (2.86; 13.00- 16.25)</i>	<i>3.42 (78)</i>	<i>.001</i>
Sequence-Memory	<i>20.63 (4.63; 16.00-24.00)</i>	<i>17.98 (3.83; 15.00-20.25)</i>	<i>2.81 (78)</i>	<i>.006</i>
Spatial-Memory	17.21 (2.16; 16.00-18.50)	16.88 (2.93; 15.00-19.00)	0.57 (78)	.57
Path-Tracking	26.95 (5.83; 24.00-31.00)	26.21 (7.54; 22.00-31.25)	0.48 (78)	.63
Maze	22.47 (6.27; 18.75-26.25)	22.12 (5.18; 17.00-26.00)	0.28 (78)	.78

<sup>a</sup>BD: Block Design.

<sup>b</sup>DS: Digit Span.

<sup>c</sup>MR: Matrix Reasoning.

<sup>d</sup>Italics indicate that a difference existed.

### **Relationship Between e-Cube Games and WAIS-IV Subtests**

We presented the expected relationships between the e-Cube games and the WAIS-IV subtests in [Table 1](#). The evaluation results are shown in [Tables 5](#) and [6](#), which list the correlations between the e-Cube scores and WAIS-IV subtest scores in the fixed and adaptive groups, respectively. The 2 groups showed somewhat different trends in results. In both groups, Assembly and BD were moderately correlated, as expected in [Table 1](#). Shape-Matching was expected to be correlated with MR, and the results from the adaptive group agreed with this. The Shape-Matching results from the fixed group showed a weak correlation with BD but no correlation with MR. Sequence-Memory and Spatial-Memory were expected to tap working memory as assessed by DS, but only the adaptive version of Spatial-Memory was moderately correlated with DS.

The adaptive version of Sequence-Memory and BD also showed a weak correlation. Path-Tracking and Maze were expected to be related to BD and MR, and only the adaptive version of Path-Tracking yielded the expected results. However, no correlations were found between both versions of Maze and any WAIS subtest. Another notable finding was that both versions of Sequence-Memory showed no significant correlation with DS. Overall, the results suggested that the adaptive versions better tap into the cognitive abilities assessed by the 3 WAIS-IV subtests.

We further analyzed the intercorrelations among the e-Cube games ([Multimedia Appendix 1](#) for fixed games and [Multimedia Appendix 2](#) for adaptive games). The adaptive e-Cube showed fewer intercorrelations than the fixed version. In the fixed version, most of the games were somewhat correlated except for Shape-Matching. Both versions of Path-Tracking and Maze were correlated with Assembly.

**Table 5.** Correlations (Spearman  $r_{38}$ , 2-tailed  $P$  value, and 95% CIs) between the e-Cube scores and raw scores of the Wechsler Adult Intelligence Scale, Fourth Edition, subtests in the fixed group.

Fixed game	BD <sup>a</sup>	DS <sup>b</sup>	MR <sup>c</sup>
<b>Assembly</b>			
<i>r</i>	<i>0.49</i> <sup>d,e</sup>	-0.20	0.10
<i>P</i> value	<i>.002</i> <sup>d</sup>	.23	.57
95% CI	<i>0.20 to 0.70</i> <sup>d</sup>	-0.49 to 0.13	-0.23 to 0.41
<b>Shape-Matching</b>			
<i>r</i>	<i>0.33</i>	-0.30	0.09 <sup>d</sup>
<i>P</i> value	<i>.04</i>	.07	.57 <sup>d</sup>
95% CI	<i>0.01 to 0.59</i>	-0.57 to 0.02	-0.24 to 0.40 <sup>d</sup>
<b>Sequence-Memory</b>			
<i>r</i>	0.31	0.14 <sup>d</sup>	0.26
<i>P</i> value	.06	.40 <sup>d</sup>	.11
95% CI	-0.01 to 0.57	-0.19 to 0.44 <sup>d</sup>	-0.07 to 0.54
<b>Spatial-Memory</b>			
<i>r</i>	0.26	0.23 <sup>d</sup>	0.19
<i>P</i> value	.12	.16 <sup>d</sup>	.25
95% CI	-0.07 to 0.54	-0.10 to 0.51 <sup>d</sup>	-0.14 to 0.48
<b>Path-Tracking</b>			
<i>r</i>	<i>0.43</i> <sup>d</sup>	0.04	0.04 <sup>d</sup>
<i>P</i> value	<i>.007</i> <sup>d</sup>	.81	.80 <sup>d</sup>
95% CI	<i>0.13 to 0.66</i> <sup>d</sup>	-0.28 to 0.36	-0.28 to 0.36 <sup>d</sup>
<b>Maze</b>			
<i>r</i>	0.30 <sup>d</sup>	-0.05	0.17 <sup>d</sup>
<i>P</i> value	.06 <sup>d</sup>	.78	.29 <sup>d</sup>
95% CI	-0.02 to 0.57 <sup>d</sup>	-0.36 to 0.27	-0.16 to 0.46 <sup>d</sup>

<sup>a</sup>BD: Block Design.<sup>b</sup>DS: Digit Span.<sup>c</sup>MR: Matrix Reasoning.<sup>d</sup>Indicates that the 2 were expected to be correlated in [Table 1](#).<sup>e</sup>Italics indicate that a correlation existed.



**Table 6.** Correlations (Spearman  $r_{42}$ , 2-tailed  $P$  value, and 95% CI) between the e-Cube scores and raw scores of the Wechsler Adult Intelligence Scale, Fourth Edition, subtests in the adaptive group.

Adaptive game	BD <sup>a</sup>	DS <sup>b</sup>	MR <sup>c</sup>
<b>Assembly</b>			
<i>r</i>	0.49 <sup>d,e</sup>	0.18	0.31
<i>P</i> value	<.001 <sup>d</sup>	.25	.05
95% CI	0.21 to 0.70 <sup>d</sup>	-0.14 to 0.47	-0.00 to 0.57
<b>Shape-Matching</b>			
<i>r</i>	0.26	0.22	0.34 <sup>d</sup>
<i>P</i> value	.09	.16	.03 <sup>d</sup>
95% CI	-0.06 to 0.53	-0.10 to 0.50	0.03 to 0.59 <sup>d</sup>
<b>Sequence-Memory</b>			
<i>r</i>	0.34	0.21 <sup>d</sup>	0.11
<i>P</i> value	.03	.19 <sup>d</sup>	.50
95% CI	0.03 to 0.59	-0.11 to 0.49 <sup>d</sup>	-0.21 to 0.41
<b>Spatial-Memory</b>			
<i>r</i>	0.09	0.51 <sup>d</sup>	0.17
<i>P</i> value	.59	<.001 <sup>d</sup>	.29
95% CI	-0.23 to 0.39	0.24 to 0.71 <sup>d</sup>	-0.15 to 0.46
<b>Path-Tracking</b>			
<i>r</i>	0.45 <sup>d</sup>	-0.01	0.45 <sup>d</sup>
<i>P</i> value	.003 <sup>d</sup>	.93	.003 <sup>d</sup>
95% CI	0.16 to 0.67 <sup>d</sup>	-0.32 to 0.30	0.16 to 0.67 <sup>d</sup>
<b>Maze</b>			
<i>r</i>	0.30 <sup>d</sup>	0.07	0.11 <sup>d</sup>
<i>P</i> value	.05 <sup>d</sup>	.62	.49 <sup>d</sup>
95% CI	-0.01 to 0.56 <sup>d</sup>	-0.25 to 0.37	-0.21 to 0.41 <sup>d</sup>

<sup>a</sup>BD: Block Design.

<sup>b</sup>DS: Digit Span.

<sup>c</sup>MR: Matrix Reasoning.

<sup>d</sup>Indicates that the 2 were expected to be correlated in Table 1.

<sup>e</sup>Italics indicate that a correlation existed.

### Technical Reliability and Usability of e-Cube

The e-Cube technology operated smoothly without any substantial technical issues identified during the study. The false detection rate, defined as the percentage ratio of incorrect detections to the total number of detections, was approximately 0.1% (6/5990). We note that all the analyses and computations mentioned previously were based on the corrected data. To analyze the results of the SUS, the scores from the 10 items were converted into a scale of 0 to 100 [47]. The overall mean SUS score was 83.40 (SD 11.52). There was a significant group difference in the results. The mean of the SUS scores from the

fixed group participants was 80.79 (SD 13.23), whereas the mean score from the adaptive group participants was 86.01 (SD 8.75). The 2-sample, 2-tailed  $t$  test with Cronbach  $\alpha=.05$  showed  $t_{78}=-2.10$  ( $P=.04$ ). The result from the adaptive group showed a significantly higher mean SUS score with a smaller SD than the fixed group. On the basis of the industry standard [48], the usability of both fixed and adaptive e-Cube games is considered grade A (ie, the games are acceptable).

## Discussion

### Principal Findings

We presented the design, development, and evaluation of the e-Cube system for automated cognitive assessment. e-Cube is a vision-based system converted from TAG-Games, a computerized system using a set of highly instrumented blocks [28]. e-Cube adopted a set of plastic cubes and a webcam instead, costing only approximately US \$50. e-Cube also reduced the labor burden by generating adaptive items, detecting answers and behavior, and scoring autonomously. A total of 6 games—Assembly, Shape-Matching, Sequence-Memory, Spatial-Memory, Path-Tracking, and Maze—were designed using the proposed play complexity measures and adaptive item generators. The e-Cube technology and the adaptive games were evaluated by testing the 2 objectives. This human participant study was conducted on the TAMU campus, and thus, most of our study participants (66/80, 82%) were college students aged between 18 and 30 years, with only 18% (14/80) aged between 31 and 60 years. Therefore, the results must be interpreted considering the skewed age distribution and demographic characteristics.

*Objective 1* was supported by the correlation analyses performed between the  $C_{play}$  values and the 2 performance indicators—the mean correctness and mean completion time obtained from the fixed group. We found that each game was correlated with at least one performance indicator. The  $C_{play}$  values of Assembly, Shape-Matching, and Sequence-Memory showed high correlations with both means. No correlations were found using the mean correctness in Spatial-Memory, Path-Tracking, and Maze. As discussed previously, correctness was not the dominant factor that widened the performance difference in Path-Tracking and Maze; thus, no correlations with correctness were found. For Spatial-Memory, 2 items involved symmetric arrangements of the geometric shapes—which made them easy to memorize regardless of the geometric complexity of the shapes. We used the same play complexity measure for both Spatial-Memory and Assembly, which appeared not to ideally reflect the difficulty associated with such memory tasks despite a high correlation in the mean completion time. This problem can be avoided at the software level by adjusting the algorithm for the item generator. Nevertheless, such symmetric images were rarely created in the adaptive version and, thus, are expected to have minimal effect on the assessment outcome.

To test the preliminary utility of the adaptive e-Cube games for cognitive assessment (*objective 2*), correlation analyses were performed between the scores from the e-Cube games and the WAIS-IV subtests. The adaptive games yielded more significant correlations with the WAIS-IV subtests than the fixed ones. This implies the potential utility of the adaptive feature of the e-Cube games based on the complexity measures. The adaptive version used a discontinuation rule (ie, the substantially small step size leading to the termination of the game), which possibly reduced the number of items in each game, fatigue, and unintended correct answers. For example, given a fixed number of items sorted by increasing difficulties, one may fail to answer correctly in the early items but can unintentionally provide

correct answers in the later items. The automatic item generator in the adaptive games adjusts the item complexity based on real-time performance, enabling the system to generate a more appropriate assessment for everyone. Note that the administration of the WAIS-IV also applies the discontinuation rule in the subtests to minimize time [49]. The subtest is terminated when a participant fails to answer a certain number of consecutive items, which differs for each subtest. Intercorrelation analyses also showed that the games in the adaptive version were more independent of one another.

There were some other notable findings from the *objective 2* evaluation study. The mean scores of Shape-Matching and Sequence-Memory in the fixed group were higher than those in the adaptive group (Table 4). In the fixed group, we found that most of the participants (26/38, 68%) correctly answered items 1 to 7 and 9, whereas only 50% (19/38) answered item 8 correctly and 24% (9/38) answered item 10 correctly in Shape-Matching. This inconsistency resulted in a higher mean score in the fixed group, but the results were not correlated with MR. In contrast, the adaptive group showed a significant correlation between Shape-Matching and MR. Sequence-Memory and Spatial-Memory were evaluated to understand which game has a monotonic relationship with DS, but a correlation was only found between the adaptive version of Spatial-Memory and DS. DS measures verbal working memory, which relies on auditory recall of numbers, sequences, and orders. However, Sequence-Memory was performed through the visual recall of geometric images, and Spatial-Memory used visual-spatial images. This fundamental design difference may have led to a lack of correlation. In addition, the language differences in participants and how they differently affect DS scores were not analyzed in this study as we did not collect such background data. Some nonnative speakers mentioned slight difficulty in memorizing the numbers said in English during the DS subtest. This feedback was collected only informally. The Path-Tracking and Maze games were correlated with Assembly, implying that their game settings or measured cognitive outcomes were similar to those of Assembly. Furthermore, Maze was not correlated with any WAIS subtest.

We further analyzed the correlation between the composite scores of the 6 e-Cube games and those of the 3 WAIS subtests. The results were  $r_{38}=0.51$  (95% CI 0.23-0.71;  $P=.001$ ) for the fixed group and  $r_{42}=0.53$  (95% CI 0.26-0.72;  $P<.001$ ) for the adaptive group. When only 4 e-Cube games (ie, Assembly, Shape-Matching, Sequence-Memory, and Spatial-Memory) were considered, the results were  $r_{38}=0.50$  (95% CI 0.21-0.71;  $P=.001$ ) for the fixed group and  $r_{42}=0.59$  (95% CI 0.34-0.76;  $P<.001$ ) for the adaptive group. Path-Tracking and Maze did not result in any meaningful relationship with the WAIS, and thus, their potential utility in cognitive assessment requires further exploration.

The low false detection rate (0.1%) demonstrated the technical functionality of the e-Cube system. Regarding the usability evaluation (*objective 2*), the average SUS scores from participants in both the fixed and adaptive groups were acceptable based on the industry standard [48]. The adaptive games resulted in a considerably higher mean SUS score with

a smaller SD than that of the fixed games. To understand the feedback for individual items, we combined the results from both groups and computed the average rate for each item. The results for the individual SUS items were uniformly positive. For the 5 even-numbered questions that were in a negative tone, such as “I found the e-Cube games unnecessarily complex,” all rates were between 1 (*strongly disagree*) and 2 (*disagree*). For the odd-numbered questions that were in a positive tone, the rates were between 4 (*agree*) and 5 (*strongly agree*) except for the following question—“I think that I would like to use the e-Cube games frequently”—with an average rate of 3.8. This was mainly due to the e-Cube games taking relatively long (approximately 50 minutes) to complete at this preliminary stage. Most of our participants (73/80, 91%) were aged <40 years, so using the game frequently to track cognitive decline was unnecessary for them. We received the highest evaluation of 4.5 on the following item: “I found the various functions in the e-Cube games were well integrated among all questions.”

### Limitations

Most participants (66/80, 82%) were TAMU students aged between 18 and 30 years. The data from participants who were young, educated, and motivated do not represent the general population well. This may also explain why none of the participants withdrew from the study. Furthermore, additional demographic information such as education level, socioeconomic status, race, and ethnicity was not collected in this preliminary evaluation study. A larger-scale validation study will be needed to involve participants from various communities with diverse backgrounds.

The administration order of the 6 e-Cube games was randomized to control for an order effect. The test order can influence the test results and bring about different levels of fatigue [50], so a well-developed cognitive assessment usually requires a standardized administration order. Although order and fatigue effects were not found in some standardized tests [50,51], the impact of the administration order of e-Cube on the scores was not investigated.

The WAIS-IV DS includes Forward, Backward, and Sequencing, which measure auditory working memory and attention with information reordering. However, Sequence-Memory and Spatial-Memory rely on visual recall and do not require any manipulation of information. Therefore, DS may not be an ideal choice for validating these 2 games. Another measure, such as Spatial Span Forward in the Wechsler Memory Scale, Fourth Edition, may be selected to compare the results with those of Sequence-Memory and Spatial-Memory in measuring relevant working memory skills.

### Future Work

The e-Cube technology was developed for fully autonomous administration and scoring of cognitive assessment targeting fine motor, hand-eye coordination, cognitive reasoning, and

working memory skills. Building on our prior work [26,28], the technology was converted into a much simpler, cheaper, and easy-to-use form, thus showing potential for use in larger-scale research studies. Our long-term objective is for the e-Cube games to serve as a routine self-assessment tool used by individuals who require continuous monitoring of their cognitive health, such as older adults and people with mild cognitive impairment. Once fully established, e-Cube can also be adapted in clinical settings, especially for remote assessment without requiring in-person interactions with an administrator. Our future work will be geared toward this long-term objective.

This extended human participant study will involve diverse participants (eg, age, sex, education, and socioeconomic status) to better represent the general population. In this future evaluation study, the existing instruments for comparison must be revisited and selected to ensure that the measures match the target cognitive domains of individual e-Cube games. Future work will also aim to establish reliability via test-retest evaluation and the validity of self- and remote administration functions via comprehensive and comparative evaluations. The study to understand the user experiences may also be extended by including an additional set of questionnaires to compare traditional instruments and the e-Cube games to gauge their preference if the e-Cube system is proven to replace some of these. To further improve the technical performance, additional vision processing methods may be added to improve this rate, such as hand detection algorithms to prevent hand motions from interfering with block detection.

The rich data from the e-Cube games on patterns, speed, and characteristics of physical movements applied to the cubes can also be explored to further explicate individual differences and cognitive and fine motor deficits. Such behavioral data may hold important information about individuals, especially those with cognitive deficits exacerbated by fine motor deficits or other behavioral symptoms such as hand tremors. Furthermore, the data provided by e-Cube have the potential to assess one's cognitive skills in a more objective way than in standard clinical settings. Upon validating its utility as a cognitive assessment tool, our future research may explore the e-Cube games for screening of early signs of neurological diseases. For the e-Cube games to be used as a routine assessment tool, we will consider shortening and gamifying the assessment to make it more fun and engaging. Enhanced graphics and sound and visual feedback mechanisms may be added to the game design. For example, we may benchmark the features of Music Blocks and iSIG-Blocks [52,53], allowing the users to customize audio, tactile, and visual sensory feedback during the cognitive assessment. The current system runs on a low-end laptop with a webcam, whereas further developments can make the algorithms executable on a tablet or cell phone using their built-in cameras. This may further reduce the cost, make it suitable for self- or remote assessment, and support long-term adoption and broader use of the technology.

### Acknowledgments

This work was supported by the National Science Foundation under grant 2002721. The authors thank Elisabeth S Ford for assisting with the human participant study and all study participants for their time and contribution to this project.

## Conflicts of Interest

None declared.

## Multimedia Appendix 1

Intercorrelations (Spearman  $r_{38}$ , 2-tailed  $P$  value, and 95% CIs) within the scores of fixed games.

[\[DOCX File , 26 KB-Multimedia Appendix 1\]](#)

## Multimedia Appendix 2

Intercorrelations (Spearman  $r_{42}$ , 2-tailed  $P$  value, and 95% CIs) within the scores of adaptive games.

[\[DOCX File , 25 KB-Multimedia Appendix 2\]](#)

## References

1. Iizuka A, Suzuki H, Ogawa S, Takahashi T, Murayama S, Kobayashi M, et al. Association between the frequency of daily intellectual activities and cognitive domains: a cross-sectional study in older adults with complaints of forgetfulness. *Brain Behav* 2021 Jan;11(1):e01923 [FREE Full text] [doi: [10.1002/brb3.1923](https://doi.org/10.1002/brb3.1923)] [Medline: [33145970](https://pubmed.ncbi.nlm.nih.gov/33145970/)]
2. Mackey AP, Park AT, Robinson ST, Gabrieli JD. A pilot study of classroom-based cognitive skill instruction: effects on cognition and academic performance. *Mind Brain Educ* 2017 Apr 11;11(2):85-95 [FREE Full text] [doi: [10.1111/mbe.12138](https://doi.org/10.1111/mbe.12138)]
3. Lipnicki DM, Makkar SR, Crawford JD, Thalamuthu A, Kochan NA, Lima-Costa MF, for Cohort Studies of Memory in an International Consortium (COSMIC). Determinants of cognitive performance and decline in 20 diverse ethno-regional groups: a COSMIC collaboration cohort study. *PLoS Med* 2019 Jul;16(7):e1002853 [FREE Full text] [doi: [10.1371/journal.pmed.1002853](https://doi.org/10.1371/journal.pmed.1002853)] [Medline: [31335910](https://pubmed.ncbi.nlm.nih.gov/31335910/)]
4. Cauchoix M, Chow PK, van Horik JO, Atance CM, Barbeau EJ, Barragan-Jason G, et al. The repeatability of cognitive performance: a meta-analysis. *Philos Trans R Soc Lond B Biol Sci* 2018 Sep 26;373(1756):20170281 [FREE Full text] [doi: [10.1098/rstb.2017.0281](https://doi.org/10.1098/rstb.2017.0281)] [Medline: [30104426](https://pubmed.ncbi.nlm.nih.gov/30104426/)]
5. Woodford HJ, George J. Cognitive assessment in the elderly: a review of clinical methods. *QJM* 2007 Aug;100(8):469-484. [doi: [10.1093/qjmed/hcm051](https://doi.org/10.1093/qjmed/hcm051)] [Medline: [17566006](https://pubmed.ncbi.nlm.nih.gov/17566006/)]
6. Hartman DE. Wechsler Adult Intelligence Scale IV (WAIS IV): return of the gold standard. *Appl Neuropsychol* 2009;16(1):85-87. [doi: [10.1080/09084280802644466](https://doi.org/10.1080/09084280802644466)] [Medline: [19205953](https://pubmed.ncbi.nlm.nih.gov/19205953/)]
7. Ward L, Bergman MA, Hebert KR. WAIS-IV subtest covariance structure: conceptual and statistical considerations. *Psychol Assess* 2012 Jun;24(2):328-340. [doi: [10.1037/a0025614](https://doi.org/10.1037/a0025614)] [Medline: [21928907](https://pubmed.ncbi.nlm.nih.gov/21928907/)]
8. Stirk S, Field B, Black J. An independent investigation of the utility of the Learning Disability Screening Questionnaire (LDSQ) within a community learning disability team. *J Appl Res Intellect Disabil* 2018 Mar;31(2):e223-e228. [doi: [10.1111/jar.12316](https://doi.org/10.1111/jar.12316)] [Medline: [27990765](https://pubmed.ncbi.nlm.nih.gov/27990765/)]
9. Roid GH, Pomplun M. The Stanford-Binet intelligence scales, fifth edition. In: Flanagan DP, Ackerman PL, editors. *Contemporary Intellectual Assessment: Theories, Tests, and Issues*. 4th edition. New York, NY, USA: Guilford Press; 2012:249-268.
10. Duggan EC, Awakon LM, Loaiza CC, Garcia-Barrera MA. Contributing towards a cultural neuropsychology assessment decision-making framework: comparison of WAIS-IV norms from Colombia, Chile, Mexico, Spain, United States, and Canada. *Arch Clin Neuropsychol* 2019 Jul 26;34(5):657-681. [doi: [10.1093/arclin/acy074](https://doi.org/10.1093/arclin/acy074)] [Medline: [30295696](https://pubmed.ncbi.nlm.nih.gov/30295696/)]
11. Raven J. The Raven's progressive matrices: change and stability over culture and time. *Cogn Psychol* 2000 Aug;41(1):1-48. [doi: [10.1006/cogp.1999.0735](https://doi.org/10.1006/cogp.1999.0735)] [Medline: [10945921](https://pubmed.ncbi.nlm.nih.gov/10945921/)]
12. Zhuo T, Kankanhalli M. Solving Raven's progressive matrices with neural networks. *arXiv* 2020 [FREE Full text] [doi: [10.48550/arXiv.2002.01646](https://doi.org/10.48550/arXiv.2002.01646)]
13. Bodmann SM, Robinson DH. Speed and performance differences among computer-based and paper-pencil tests. *J Educ Comput Res* 2004 Jul;31(1):51-60 [FREE Full text] [doi: [10.2190/GROQ-YT0F-7LKB-F033](https://doi.org/10.2190/GROQ-YT0F-7LKB-F033)]
14. Noland RM. Intelligence testing using a tablet computer: experiences with using Q-interactive. *Train Educ Prof Psychol* 2017;11(3):156-163 [FREE Full text] [doi: [10.1037/tep0000149](https://doi.org/10.1037/tep0000149)]
15. Delgado MT, Uribe PA, Alonso AA, Díaz RR. TENI: a comprehensive battery for cognitive assessment based on games and technology. *Child Neuropsychol* 2016;22(3):276-291. [doi: [10.1080/09297049.2014.977241](https://doi.org/10.1080/09297049.2014.977241)] [Medline: [25396766](https://pubmed.ncbi.nlm.nih.gov/25396766/)]
16. Tong T, Chignell M, Lam P, Tierney MC, Lee J. Designing serious games for cognitive assessment of the elderly. *Proc Int Symp Hum Factors Ergon Healthc* 2014 Jul 22;3(1):28-35 [FREE Full text] [doi: [10.1177/2327857914031004](https://doi.org/10.1177/2327857914031004)]
17. Leduc-McNiven K, White B, Zheng H, McLeod RD, Friesen MR. Serious games to assess mild cognitive impairment: 'the game is the assessment'. *Res Rev Insights* 2018;2(1):1-11 [FREE Full text] [doi: [10.2196/preprints.9354](https://doi.org/10.2196/preprints.9354)]
18. Zohaib M. Dynamic difficulty adjustment (DDA) in computer games: a review. *Adv Hum-Comput Interact* 2018 Jan 01;2018:5681652 [FREE Full text] [doi: [10.1155/2018/5681652](https://doi.org/10.1155/2018/5681652)]
19. de Andrade KO, Pasqual TB, Caurin GA, Crocomo MK. Dynamic difficulty adjustment with evolutionary algorithm in games for rehabilitation robotics. In: *Proceedings of the 2016 IEEE International Conference on Serious Games and*



- Applications for Health. 2016 Presented at: SeGAH '16; May 11-13, 2016; Orlando, FL, USA p. 1-8 URL: <https://ieeexplore.ieee.org/abstract/document/7586277> [doi: [10.1109/segah.2016.7586277](https://doi.org/10.1109/segah.2016.7586277)]
20. Sharlin E, Itoh Y, Watson BA, Kitamura Y, Sutphen S, Liu L. Cognitive cubes: a tangible user interface for cognitive assessment. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2002 Presented at: CHI '02; April 20-25, 2002; Minneapolis, MN, USA p. 347-354. [doi: [10.1145/503376.503438](https://doi.org/10.1145/503376.503438)]
  21. Jimison H, Pavel M, Wild K, Williams D, McKanna J, Bissel A. Embedded assessment of cognitive performance with elders' use of computer games in a residential environment. In: Proceedings of the 2006 Workshop on the Cognitive Science of Games and Gaming. 2006 Presented at: CSGV '06; July 26-29, 2009; Vancouver, Canada URL: <https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=b5fe5e9609c0b8902d97c7567574e358f0d2dfb3>
  22. Tong T, Yeung J, Sandrakumar J, Chignell M, Tierney MC, Lee J. Improving the ergonomics of cognitive assessment with serious games. Proc Int Symp Hum Factors Ergon Healthc 2015 Jun 15;4(1):1-5 [FREE Full text] [doi: [10.1177/2327857915041001](https://doi.org/10.1177/2327857915041001)]
  23. Tong T, Chignell M, Tierney MC, Lee J. A serious game for clinical assessment of cognitive status: validation study. JMIR Serious Games 2016 May 27;4(1):e7 [FREE Full text] [doi: [10.2196/games.5006](https://doi.org/10.2196/games.5006)] [Medline: [27234145](https://pubmed.ncbi.nlm.nih.gov/27234145/)]
  24. Reid JM. Testing nonverbal intelligence of working-age visually impaired adults: evaluation of the adapted Kohs block design test. J Vis Impair Blind 2002;96(8):585-595 [FREE Full text] [doi: [10.1177/0145482x0209600805](https://doi.org/10.1177/0145482x0209600805)]
  25. Cerrato A, Ponticorvo M, Gigliotta O, Bartolomeo P, Miglino O. The assessment of visuospatial abilities with tangible interfaces and machine learning. In: Proceedings of the 8th International Work-Conference on the Interplay Between Natural and Artificial Computation. 2019 Presented at: IWINAC '19; June 3-7, 2019; Almeria, Spain p. 78-87. [doi: [10.1007/978-3-030-19591-5\\_9](https://doi.org/10.1007/978-3-030-19591-5_9)]
  26. Lee K, Jeong D, Schindler RC, Short EJ. SIG-Blocks: tangible game technology for automated cognitive assessment. Comput Human Behav 2016 Dec;65:163-175 [FREE Full text] [doi: [10.1016/j.chb.2016.08.023](https://doi.org/10.1016/j.chb.2016.08.023)]
  27. Jeong DH. Distributed wireless sensor network systems: theoretical framework, algorithms, and applications. Case Western Reserve University. 2015. URL: [https://etd.ohiolink.edu/apexprod/rws\\_olink/r/1501/10?clear=10&p10\\_accession\\_num=case1436541959](https://etd.ohiolink.edu/apexprod/rws_olink/r/1501/10?clear=10&p10_accession_num=case1436541959) [accessed 2023-04-17]
  28. Jeong D, Endri K, Lee K. TaG-games: tangible geometric games for assessing cognitive problem-solving skills and fine motor proficiency. In: Proceedings of the 2010 IEEE Conference on Multisensor Fusion and Integration. 2010 Presented at: MFI '10; September 5-7, 2010; Salt Lake City, UT, USA p. 32-37. [doi: [10.1109/mfi.2010.5604479](https://doi.org/10.1109/mfi.2010.5604479)]
  29. Brooke J. SUS: a retrospective. J Usability Stud 2013;8(2):29-40 [FREE Full text]
  30. Vlachogianni P, Tselios N. Perceived usability evaluation of educational technology using the System Usability Scale (SUS): a systematic review. J Res Technol Educ 2021 Jan 25;54(3):392-409. [doi: [10.1080/15391523.2020.1867938](https://doi.org/10.1080/15391523.2020.1867938)]
  31. Lewis JR. The system usability scale: past, present, and future. Int J Hum Comput Interact 2018;34(7):577-590. [doi: [10.1080/10447318.2018.1455307](https://doi.org/10.1080/10447318.2018.1455307)]
  32. Szeliski R. Computer Vision: Algorithms and Applications. Cham, Switzerland: Springer; 2010.
  33. Ramer U. An iterative procedure for the polygonal approximation of plane curves. Comput Graph Image Process 1972;1(3):244-256 [FREE Full text] [doi: [10.1016/S0146-664X\(72\)80017-0](https://doi.org/10.1016/S0146-664X(72)80017-0)]
  34. Royer FL, Gilmore GC, Gruhn JJ. Stimulus parameters that produce age differences in block design performance. J Clin Psychol 1984 Nov;40(6):1474-1485. [doi: [10.1002/1097-4679\(198411\)40:6<1474::aid-jclp2270400636>3.0.co;2-7](https://doi.org/10.1002/1097-4679(198411)40:6<1474::aid-jclp2270400636>3.0.co;2-7)] [Medline: [6511968](https://pubmed.ncbi.nlm.nih.gov/6511968/)]
  35. Lee K, Jeong D, Schindler RC, Hlavaty LE, Gross SI, Short EJ. Interactive block games for assessing children's cognitive skills: design and preliminary evaluation. Front Pediatr 2018 May 08;6:111 [FREE Full text] [doi: [10.3389/fped.2018.00111](https://doi.org/10.3389/fped.2018.00111)] [Medline: [29868520](https://pubmed.ncbi.nlm.nih.gov/29868520/)]
  36. Shannon C. A mathematical theory of communication. Mob Comput Commun Rev 2001;5(1):3-55. [doi: [10.1145/584091.584093](https://doi.org/10.1145/584091.584093)]
  37. Haralick RM. Statistical and structural approaches to texture. Proc IEEE 1979 May;67(5):786-804 [FREE Full text] [doi: [10.1109/proc.1979.11328](https://doi.org/10.1109/proc.1979.11328)]
  38. Partio M, Cramariuc B, Gabbouj M, Visa A. Rock texture retrieval using gray level co-occurrence matrix. In: Proceedings of the 5th Nordic Signal Processing Symposium. 2002 Presented at: NORSIG '02; October 4-7, 2002; Hurtigruten, Norway.
  39. Gao P, Li Z, Zhang H. Thermodynamics-based evaluation of various improved Shannon entropies for configurational information of gray-level images. Entropy (Basel) 2018 Jan 02;20(1):19 [FREE Full text] [doi: [10.3390/e20010019](https://doi.org/10.3390/e20010019)] [Medline: [33265110](https://pubmed.ncbi.nlm.nih.gov/33265110/)]
  40. Guiaşu S. Weighted entropy. Rep Math Phys 1971;2(3):165-179. [doi: [10.1016/0034-4877\(71\)90002-4](https://doi.org/10.1016/0034-4877(71)90002-4)]
  41. Bonchev D, Buck GA. Quantitative measures of network complexity. In: Rouvray DH, Bonchev D, editors. Complexity in Chemistry, Biology, and Ecology. New York, NY, USA: Springer; 2005:191-235.
  42. Hart PE, Nilsson NJ, Raphael B. A formal basis for the heuristic determination of minimum cost paths. IEEE Trans Syst Sci Cybern 1968;4(2):100-107. [doi: [10.1109/tssc.1968.300136](https://doi.org/10.1109/tssc.1968.300136)]
  43. Charan J, Biswas T. How to calculate sample size for different study designs in medical research? Indian J Psychol Med 2013 Apr;35(2):121-126 [FREE Full text] [doi: [10.4103/0253-7176.116232](https://doi.org/10.4103/0253-7176.116232)] [Medline: [24049221](https://pubmed.ncbi.nlm.nih.gov/24049221/)]

44. Whitehead AL, Julious SA, Cooper CL, Campbell MJ. Estimating the sample size for a pilot randomised trial to minimise the overall trial sample size for the external pilot and main trial for a continuous outcome variable. *Stat Methods Med Res* 2016 Jun;25(3):1057-1073 [[FREE Full text](#)] [doi: [10.1177/0962280215588241](https://doi.org/10.1177/0962280215588241)] [Medline: [26092476](https://pubmed.ncbi.nlm.nih.gov/26092476/)]
45. Teare M, Dimairo M, Shephard N, Hayman A, Whitehead A, Walters SJ. Sample size requirements to estimate key design parameters from external pilot randomised controlled trials: a simulation study. *Trials* 2014 Jul 03;15:264 [[FREE Full text](#)] [doi: [10.1186/1745-6215-15-264](https://doi.org/10.1186/1745-6215-15-264)] [Medline: [24993581](https://pubmed.ncbi.nlm.nih.gov/24993581/)]
46. Taylor R. Interpretation of the correlation coefficient: a basic review. *J Diagn Med Sonogr* 1990 Jan;6(1):35-39 [[FREE Full text](#)] [doi: [10.1177/875647939000600106](https://doi.org/10.1177/875647939000600106)]
47. Guerci J. Easily calculate SUS Score. *UX Planet*. 2020 Mar 27. URL: <https://uxplanet.org/easily-calculate-sus-score-a464d753e5aa> [accessed 2020-05-26]
48. Sauro J. 5 ways to interpret a SUS score. *Measuring U*. 2018 Sep 19. URL: <https://measuringu.com/interpret-sus-score/> [accessed 2023-04-17]
49. Abdelhamid GS, Bassiouni MG, Gómez-Benito J. Assessing cognitive abilities using the WAIS-IV: an item response theory approach. *Int J Environ Res Public Health* 2021 Jun 25;18(13):6835 [[FREE Full text](#)] [doi: [10.3390/ijerph18136835](https://doi.org/10.3390/ijerph18136835)] [Medline: [34202249](https://pubmed.ncbi.nlm.nih.gov/34202249/)]
50. Ryan JJ, Glass LA, Hinds RM, Brown CN. Administration order effects on the test of memory malingering. *Appl Neuropsychol* 2010 Oct;17(4):246-250. [doi: [10.1080/09084282.2010.499802](https://doi.org/10.1080/09084282.2010.499802)] [Medline: [21154037](https://pubmed.ncbi.nlm.nih.gov/21154037/)]
51. Tulskey DS, Zhu J. Could test length or order affect scores on letter number sequencing of the WAIS-III and WMS-III? Ruling out effects of fatigue. *Clin Neuropsychol* 2000 Nov;14(4):474-478. [doi: [10.1076/clin.14.4.474.7205](https://doi.org/10.1076/clin.14.4.474.7205)] [Medline: [11262717](https://pubmed.ncbi.nlm.nih.gov/11262717/)]
52. Miranda D, Lee K. Music blocks: audio-augmented block games for play-based cognitive assessment. In: Proceedings of the 2018 IEEE Games, Entertainment, Media Conference. 2018 Presented at: GEM '18; August 15-17, 2018; Galway, Ireland p. 375-381. [doi: [10.1109/GEM.2018.8516267](https://doi.org/10.1109/GEM.2018.8516267)]
53. Jeong D, Lee K. iSIG-Blocks: interactive creation blocks for tangible geometric games. *IEEE Trans Consum Electron* 2015 Nov;61(4):420-428. [doi: [10.1109/TCE.2015.7389795](https://doi.org/10.1109/TCE.2015.7389795)]

## Abbreviations

**BD:** Block Design

**DS:** Digit Span

**GLCM:** gray-level co-occurrence matrix

**MR:** Matrix Reasoning

**SUS:** System Usability Scale

**TAMU:** Texas A&M University

**WAIS:** Wechsler Adult Intelligence Scale

**WAIS-IV:** Wechsler Adult Intelligence Scale, Fourth Edition

*Edited by G Eysenbach; submitted 09.07.22; peer-reviewed by M Friesen, M Kapsetaki, A Bikic, K Wiley, L Campbell, A Lynham, Z Aghaei; comments to author 20.12.22; revised version received 17.02.23; accepted 13.03.23; published 16.05.23*

*Please cite as:*

*Cheng X, Gilmore GC, Lerner AJ, Lee K*

*Computerized Block Games for Automated Cognitive Assessment: Development and Evaluation Study*

*JMIR Serious Games* 2023;11:e40931

URL: <https://games.jmir.org/2023/1/e40931>

doi: [10.2196/40931](https://doi.org/10.2196/40931)

PMID:

©Xiangyi Cheng, Grover C Gilmore, Alan J Lerner, Kiju Lee. Originally published in JMIR Serious Games (<https://games.jmir.org>), 16.05.2023. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in JMIR Serious Games, is properly cited. The complete bibliographic information, a link to the original publication on <https://games.jmir.org>, as well as this copyright and license information must be included.