

Play-Centric Design and Usability Evaluation of a Block-Based Programming Serious Game for Low Vision Children

Huang Sheng^{ID}, Chau Kien Tsong*^{ID}, Wan Ahmad Jaafar Wan Yahaya^{ID}
Universiti Sains Malaysia, Malaysia

Abstract—Serious games are effective tools for learning beyond entertainment, yet children with low vision face obstacles when playing block-based programming games because these environments rely on small graphical elements, low colour contrast, and drag-and-drop interaction, a barrier specific to block-based programming that does not arise in other game genres. This study presents the design and usability evaluation of a block-based programming serious game for teaching sequences, loops, and conditionals to low-vision children aged seven years old. Accessibility elements were implemented in the prototype, namely screen design, multimedia, keyboard-based interaction replacing drag-and-drop, and language. Usability testing and a pre-test and post-test were conducted with sixteen children in matched experimental and control groups of eight, the control group using a feature-reduced baseline version, with results reported descriptively. The accessible version obtained a Good overall usability score of 3.80, against a Weak score of 1.65 for the baseline. Only 25 per cent of control participants independently reached all three target concepts, compared with 100 per cent in the accessible version. Because only baseline participants who reached the target concepts were compared, the learning gain difference is a conservative estimate of the accessibility effect.

Keywords—*Serious game; block-based programming; low vision; accessibility; usability testing; computational thinking; assistive technology*

I. INTRODUCTION

Nowadays, information and communication technologies (ICT) are used in education. We can find a lot of educational applications that can help young learners learn. Several ICT applications are on the market, such as e-books, multimedia courseware, and games [1][2]. Gaming is a very popular activity enjoyed by children. Thus, a serious game is the most promising tool for helping children learn. A serious game is used beyond entertainment [1]. Serious games are powerful tools for building language, science, math, and computing knowledge and skills. This educational value has been demonstrated across contexts including early childhood education [22] and STEM learning more broadly [24]. Despite a growing number of studies on using serious games for learning, children with visual impairments have obstacles when playing the games. There are two categories of visual impairment: blindness and low vision. Low vision children have a visual impairments that can be assisted with assistive technology. The World Health Organisation defines low vision as visual acuity between 20/70 and 20/400, with a possible correction of 20 degrees or less [31]. Low vision children have obstacles accessing graphical

elements in games. Prior studies indicate that low vision children face several accessibility issues, such as visual, animation, audio, and navigation obstacles [9], [11]. This is consistent with broader evidence that low vision affects reading behaviour [17] and the interpretation of visual information more generally [18]. They need to look at graphical elements from a very close distance because the size of the text and graphics is small. The colour contrast between the graphics and background is often low. They also have difficulty navigating the menus and buttons in games, and fast animation movements affect their vision when playing games.

Among educational applications, block-based programming environments such as Scratch have become the most popular tool for introducing young children to computational thinking, because they allow children to build instructions by assembling visual blocks instead of typing code [4], [21]. The use of such tools to build computational thinking and programming skills has been demonstrated across K-12 educational settings [25], [26]. Unlike most other educational applications, however, block-based programming environments depend on an interaction method that other serious game genres do not require: drag-and-drop, where children must click, hold, and drag a block precisely into place to construct a program. This dependence on drag-and-drop is a barrier specific to block-based programming. This is consistent with accessibility issues documented across block-based programming environments more broadly [20]. A mathematics or language serious game can be made usable for low vision children through screen design and multimedia adaptations alone [3], but a block-based programming game remains unusable if children cannot physically perform the drag-and-drop action, regardless of how large or high-contrast the blocks are made. Thus, to ensure low-vision children can play the game, a block-based programming serious game for learning sequences, loops, and conditionals is developed for low-vision children aged seven years old. The game consists of three levels of programming tasks, based on introductory computational thinking content appropriate for Year 1 pupils, namely ordering instructions (sequences), repeating instructions (loops), and if-then decision making (conditionals). These constructs are consistent with the core computational-thinking skills assessed in prior block-based programming research [14].

Serious games should be easily accessible to low vision children. The game design should be flexible, with an interface adapted to the accessibility requirements of low vision children.

*Corresponding author.

To enhance the game's accessibility for low vision children, accessibility elements have been implemented in the serious game prototype. Those elements are screen design (buttons, menus, and navigation), multimedia (text, graphics, and audio), interaction method, and language. The children can set the appropriate background colour and contrast level based on their vision level. The text size is large and adjustable. The background colour has a high colour contrast, and bright colours and large graphics are implemented in the game. Background audio and navigation cues are provided so children can identify their position on the screen and understand the game's instructions without relying on vision alone. Most importantly, the conventional drag-and-drop interaction used to assemble programming blocks is replaced with a keyboard-based selection method: children navigate the interface using arrow keys, and press Enter to select and place a block, removing the need for precise cursor control. This design choice is consistent with accessible input approaches explored for other populations with fine motor or visual constraints [19]. A feature-reduced baseline version of the same game, without these accessibility elements, is also developed to serve as a comparison during usability testing.

Upon completion of the serious game, usability testing is conducted to identify the accessibility of the serious game for low-vision children based on the usability level, and a pre-test/post-test is conducted to identify whether the game helps low-vision children understand the target programming concepts.

Thus, this study presents the design and usability testing of a block-based programming serious game prototype developed specifically for low vision children aged seven years old. The serious game is developed based on the accessibility requirements of low vision children, refined through pilot testing and expert heuristic evaluation prior to user testing. The effectiveness, efficiency, and user satisfaction of the serious game are tested in the usability testing, and the learning outcome of the game is tested using a pre-test and post-test benchmarked against a feature-reduced baseline version. The observation technique is used for analysing the serious game. The usability testing aims to identify the serious game's usability level and learning effectiveness for low-vision children.

This study is organised as follows: Section I presents the introduction, while Section II describes the related work. Section III describes the methodology, and Section IV and Section V discuss the results and discussion, respectively, while Section VI presents the conclusion and plan for future work.

II. RELATED WORK

The International Organization for Standardization (ISO) 9241-11 (1998) defines usability as the ability of a product to be used by users to accomplish certain goals effectively, efficiently, and with user satisfaction [30]. Usability also refers to whether a product is easy to use. Based on the ISO 9241-11 definition, usability consists of the following components:

- Effectiveness: The accuracy and completeness with which a user can achieve a certain goal in a certain environment.

- Efficiency: The effort the user gives to complete a task and achieve the objective of using an application or system.
- User Satisfaction: User comfort and acceptance of the system.

Usability testing is conducted to identify usability problems before releasing the serious game in its real context. There is some overlap between the concepts of usability and accessibility: accessibility enables users to access ICT applications independently without accounting for their disability, while usability refers to the ability of the ICT application to carry out the intended function effectively, efficiently, and with satisfaction when used by users in general [30]. Therefore, accessibility specifically refers to disabled users, while usability refers to the general population. Usability testing is also considered suitable for assessing ICT applications for accessibility [7].

Prior research has proposed a serious-game accessibility design model consisting of three components, namely accessibility, game design, and pedagogy [11], and implemented accessibility elements including screen design, multimedia, object motion, and language in a mathematics serious game for low vision children, refined using a play-centric design approach through a low-fidelity prototype and pilot feedback from low vision children [8]. Usability testing of this game, conducted through direct observation with fifteen low vision children, reported good usability scores across effectiveness (4.74), efficiency (4.55), and user satisfaction (4.92), with an overall usability score of 4.76 [3]. These findings show that accessibility elements such as screen design, multimedia, language, and object motion are important in increasing the usability of serious games for low vision children. However, this study focused on a mathematics game, in which children select and tap answers on screen; it does not address the additional interaction barrier present in block-based programming, where children are required to drag and connect blocks to construct a sequence of instructions, nor does it assess whether use of the game was associated with a measurable gain in the target academic content.

Block-based programming languages have become the dominant approach to introductory programming education because they lower the syntactic barrier to entry for young learners [4]. Prior efforts to make block-based programming accessible have largely targeted children who are blind: Blocks4All, an iPad-based block-based programming environment, replaces the visual programming canvas with touchscreen exploration and VoiceOver audio feedback, and was evaluated with children who are blind or have visual impairments controlling a physical robot [5]. This and related approaches are designed around the assumption that the user has no usable vision, and therefore do not directly address the residual-vision use case of low vision children, for whom enlarged and high-contrast visual elements, combined with reduced dependence on fine motor precision rather than a fully non-visual interface, are more appropriate.

A representative sample of users must conduct usability testing because users are different and have different problems [6]. Suitable usability testing methods should be considered when planning testing with children. Usability testing with children consists of introspection, direct observation, thinking aloud, and interaction. Direct observation is used to observe users as they perform their tasks, and this technique has been recognised as one of the best techniques for usability testing with young children, since it does not depend on children's ability to verbalise their experience [6], [15]. Satisfaction with children is commonly measured using simplified emoticon-based scales, such as like, neutral, and dislike, to observe the user's reaction to the game [3].

Accessibility research more broadly has also been reported within this journal, for example in evaluations of web accessibility barriers [9] and usability evaluations of interfaces for specific user populations [12][13], reflecting the value of usability testing as a method for surfacing accessibility problems that are not apparent from design review alone. While prior studies have evaluated the usability of serious games for low vision children, usability testing alone does not establish whether the target educational content was learned. Coupling usability testing with a pre-test/post-test learning outcome assessment allows a study to determine not only whether low vision children can use a serious game, but whether using it is associated with a gain in the knowledge the game is intended to teach.

III. METHODOLOGY

The usability testing method used in this study is direct observation. The observation method is conducted to evaluate the user experience of the product by observing users carrying out the tasks provided. Observations are conducted based on effectiveness, efficiency, and user satisfaction. User satisfaction

is subjective; thus, interviews are also conducted with low-vision children to obtain feedback regarding their satisfaction level with the tested serious game. During the interview, emoticons (like, neutral, dislike) are prepared to observe the user's reaction to the game. In addition to usability testing, a pre-test and post-test are conducted to identify whether the game is associated with a gain in understanding of sequences, loops, and conditionals. The overall study follows a three-phase process: 1) design and iterative refinement of the game prototype, incorporating pilot testing with children and expert heuristic evaluation; 2) recruitment and assignment of participants; and 3) usability and learning-outcome testing.

A. Game Design and Prototype Refinement

The game is designed following a three-component accessibility model, namely accessibility, game design, and pedagogy [11]. A play-centric design approach is adopted, beginning with a low-fidelity prototype refined through pilot feedback from three to five low-vision children, consistent with the play-centric prototyping approach used in prior serious-game design research for this population [8].

Table I shows the accessibility elements implemented in the accessible version of the game, and the corresponding elements omitted in the feature-reduced baseline version used for comparison during testing.

Both versions shared the same level content, learning objectives, and progression structure. They differed only in the presence or absence of the full set of accessibility elements summarised in Table I (screen design, multimedia, interaction method, and text). Consequently, any systematic differences observed between the experimental and control groups can be attributed to the accessibility package, rather than to the keyboard-based interaction method in isolation.

TABLE I. ACCESSIBILITY ELEMENTS IN THE ACCESSIBLE AND BASELINE VERSIONS

Element	Accessible version	Baseline version
Screen design	Large, adjustable buttons and menus	Standard, fixed-size buttons and menus
Multimedia	High-contrast graphics, adjustable colour scheme, audio feedback	Standard-contrast graphics, no audio feedback
Interaction	Keyboard-based selection (arrow keys + Enter)	Mouse-based drag-and-drop
Text	Large, adjustable font size	Standard, fixed font size

Note: The accessible version differs from the baseline version by the full set of accessibility adaptations listed above. Differences in outcomes between the two versions are therefore attributable to the combined accessibility package rather than to any single element in isolation.

The baseline version is intended as a credible feature-reduced comparator rather than a deliberately inaccessible interface: it retains the same learning content, level structure, narrative theme, and progression sequence as the accessible version, and systematically omits only the accessibility elements listed in Table I (large, adjustable buttons and menus; high-contrast graphics; adjustable colour schemes; audio feedback; keyboard-based selection; and large, adjustable text). This design allows any observed differences in usability scores, concept-exposure rates, and learning gains between the two groups to be attributed to the presence or absence of the accessibility package, rather than to differences in pedagogical content or task difficulty.

Following pilot testing, the refined prototype is submitted to an expert heuristic evaluation prior to full user testing. Four

experts in human-computer interaction, special education, and game design independently evaluate the prototype using two heuristic sets: game heuristics (challenge, reward, goals, storyline) and pedagogical heuristics (learning objectives, learning theory, content alignment), and complete a structured evaluation questionnaire with improvement recommendations. This heuristic evaluation approach follows methods developed specifically for assessing usability and user experience in children's digital tools [16]. Recommendations arising from this evaluation are consolidated and used to guide a second round of prototype refinement before the prototype is used in the user testing phase.

The serious game user interface, incorporating the accessibility elements described above, is depicted in Fig. 1.

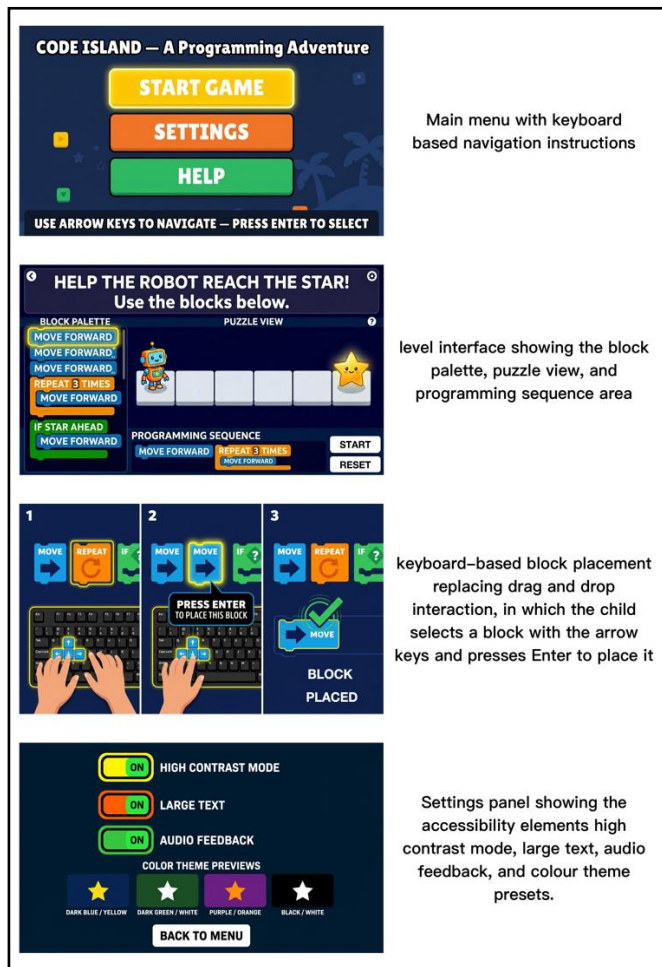


Fig. 1. Serious game user interface.

B. Participants

This study uses a selected study sample recruited from a special education school. Low vision is defined according to World Health Organisation criteria as visual acuity between 20/70 and 20/400, with a possible correction of 20 degrees or less [31], verified through participants' school medical records. This testing involves low vision children aged seven years old, assigned to an experimental group using the accessible version of the game and a control group using the feature-reduced baseline version, matched by age, prior programming or game experience, and level of visual acuity. This study was conducted in accordance with the principles of the Declaration of Helsinki. Ethical approval for this multi-centre study was obtained from the institutional review boards corresponding to the authors' respective affiliations: the Jawatankuasa Etika Penyelidikan Manusia, Universiti Sains Malaysia (No. USM/JEPeM/PP/26070645), for all authors from Universiti Sains Malaysia, and the South China Normal University Ethics Approval Certificate for Research Involving Human Subjects (No. SCNU-2026-039), for the author also affiliated with South China Normal University. To conduct this testing, ethics that the researcher must follow are observed: consent from parents or guardians is required before testing to ensure the study meets research guidelines and ethics, the purpose of the research is explained to the teachers involved, and children are informed of

their right to withdraw at any time. Both groups receive the same fixed session duration and the same level of researcher presence throughout testing, so that any difference observed between groups can be attributed to the accessibility adaptations rather than to differences in attention received or to the basic novelty of encountering an unfamiliar digital tool.

C. Usability Testing Procedures

The activities involved in usability testing include determining the testing objective, preparing the instruments and testing tools, identifying the task scenario, preparing the testing checklist, and measuring usability. The testing procedure is further explained in the following sections.

1) *Determination of testing objective:* This usability testing is conducted to verify the accessibility of the serious game to low-vision children. The evaluation is conducted based on the ISO 9241-11 definition, which covers effectiveness, efficiency, and user satisfaction concerning the serious game [30].

2) *Testing instruments and testing tools preparation:* The testing instruments consist of the accessible and baseline serious game prototypes, a usability testing task list, an observation checklist, and audio and screen recording tools.

3) *Identify task scenarios:* Usability testing is conducted by giving the user a task scenario while the user plays the game, prepared according to the game's screens, such as the start screen, the level screens, and the task screen within each level.

4) *Testing checklist preparation:* A checklist is prepared based on the usability constructs of effectiveness, efficiency, and user satisfaction [30], covering the accessibility elements shown in Table I.

5) *Measurement of usability testing:* The measurement of effectiveness, efficiency, and user satisfaction is conducted using a usability score, following the scoring approach used in prior serious-game usability research [3].

6) *Effectiveness measurement:* If a task is completed successfully, a score of 5.0 is marked on the effectiveness checklist item. If the task is half completed, a score of 2.5 is marked; if the task is not completed, a score of 0 is marked. The usability score for effectiveness, X , is formulated as:

$$X = \frac{\sum \text{score for each effectiveness checklist item}}{\text{total effectiveness checklist items}} \quad (1)$$

7) *Efficiency measurement:* Efficiency is measured based on the time taken to complete the task relative to the average task completion time. The average task completion time used as the scoring benchmark is established separately for each version (accessible and baseline) from the pilot study described in Section III-C6, to account for the difference in interaction speed between the keyboard-based and mouse-based interaction methods; each level's benchmark is computed as the mean completion time recorded for that level across the pilot sample. If the time taken is equal to or faster than the average, a score of 5.0 is marked; if slower, a score of 2.5 is marked; if the task fails, a score of 0 is marked. The usability score for efficiency, X , is formulated as:

$$X = \frac{\sum \text{score for each efficiency checklist item}}{\text{total efficiency checklist items}} \quad (2)$$

8) *User satisfaction measurement*: A score of 5.0 is marked if a child's reaction is "like"; 2.5 if "neutral"; and 0 if "dislike". The usability score for user satisfaction, X, is formulated as:

$$X = \frac{\sum \text{score for each satisfaction checklist item}}{\text{total satisfaction checklist items}} \quad (3)$$

The average value of each usability category is considered the score for that category, and the overall usability score is the average of the three category scores. This scoring approach is consistent with psychometric usability and acceptability scales developed for children and adolescent users [23]. Table II shows the interpretation of the overall usability score, following the interpretation bands used in prior research [3].

TABLE II. USABILITY SCORE INTERPRETATION

Weak	Moderately Good	Good
$0 \leq x \leq 2.49$	$2.5 \leq x \leq 3.49$	$3.5 \leq x \leq 5.0$

9) *Implementation of usability test*: A pilot study is conducted on the instrument to test its reliability before it is used in the actual testing, involving three to five low-vision children, distinct from the pilot testing conducted during prototype refinement. During usability testing, children play the game with minimal researcher presence; no instructional assistance is provided during the session, so that the effectiveness, efficiency, and satisfaction scores reflect the accessibility of the game itself rather than researcher intervention. Because young children may not spontaneously ask to stop an activity even when struggling, the session is not left to run purely on the child's initiative to withdraw: the researcher proactively pauses the session and checks in with the child if either of the following is observed: no measurable progress for three consecutive minutes, or five consecutive failed attempts at the same step without progression, at which point the child is asked whether they would like to continue or take a break, and a break or early termination is offered accordingly. This check-in is in addition to, not instead of, immediate termination upon overt signs of distress. Control group participants are offered the opportunity to try the accessible version of the game after their session has ended. A screen recorder and camera are used to record testing sessions, and researchers observe and record children's actions on the provided task scenario checklist, which is then transferred to the usability checklist. The data collected are analysed using a descriptive analysis method.

D. Learning Outcome Assessment

In addition to usability testing, a pre-test and post-test are administered to identify whether use of the game is associated with a gain in understanding of sequences, loops, and conditionals.

1) *Instrument*: Each pre-test and post-test consisted of nine items organised into three equal sections corresponding to the three target concepts: three items on sequences, in which

children arranged or identified the correct order of a short series of action steps to complete a simple path; three items on loops, in which children identified or constructed the correct repeating pattern needed to complete a repeated action; and three items on conditionals, in which children selected the appropriate if-then rule to make a character respond differently under two conditions. Each correct response was awarded one point, with no partial credit for individual items, yielding a maximum possible score of 9 on both the pre-test and the post-test. The pre-test and post-test are administered one-on-one, in oral format, accompanied by large, high-contrast visual aids, consistent with the accessibility elements applied to the game itself. Equivalent item sets are used for the pre-test and post-test, differing only in surface-level details such as character names, object colours, and path layouts, to reduce practice effects. The item set is reviewed by a special education teacher and a computer science educator for alignment with the target concepts, age-appropriateness, and clarity for oral administration.

2) *Procedure*: Both the experimental group and the control group complete the pre-test before their assigned game session and the post-test immediately afterwards. Learning gain is calculated as the difference between the post-test and pre-test scores.

3) *Concept-exposure criterion and analysis*: For the experimental group, all participants are expected to reach the levels covering all three core concepts within the session and are included in the learning gain analysis. For the control group, only participants who independently reach these levels are included; participants who do not reach concept exposure are excluded from the learning gain comparison, but their outcome-how many control participants did and did not reach concept exposure, and at which level they stopped is retained and reported as usability evidence in its own right, since it directly quantifies the severity of the barriers present in the baseline environment. If the number of control group participants who independently reach concept exposure is fewer than three, group-level means are not reported for this subgroup; instead, the individual pre-test and post-test scores of these participants are reported as case data. Because the participants who reach concept exposure independently are not representative of the control group as assigned, they are more likely to be those with relatively milder visual impairment, better fine motor control, or prior exposure to similar interfaces; any comparison between this subgroup and the experimental group is interpreted as a conservative estimate of the accessibility effect: the true gap between the accessible and baseline conditions is likely larger than the reported comparison suggests. Given the exploratory nature of this study and the anticipated small sample size, learning gain and usability scores are reported descriptively (means, standard deviations, frequencies); no inferential statistical tests are performed. A non-parametric procedure such as the Wilcoxon signed-rank test was considered as an exploratory measure for the matched-pair learning-gain comparison; however, it is not feasible here because six of the

eight matched pairs are incomplete for this comparison, as the corresponding control-group participant did not reach concept exposure (Section IV-D). A matched-pairs analysis restricted to the two complete pairs would be too small to yield an interpretable test statistic, even as an exploratory measure, and is therefore not reported.

IV. RESULTS

This section reports the results of the usability testing and the pre-test/post-test learning outcome assessment. Consistent with the exploratory design and small sample size described in Section III-D3, all usability scores, learning gains, and event

counts are reported descriptively using means, standard deviations, and frequencies; no inferential statistical tests are performed. The results are presented in the order of the participant characteristics, the expert heuristic evaluation conducted prior to user testing, the usability scores, the concept-exposure outcomes for the control group, the learning outcome results, the termination-criterion events, and the inter-rater reliability of the learning-outcome scoring.

A. Participants

Table III shows the demographic and matching characteristics of the participants recruited for this study.

TABLE III. PARTICIPANT CHARACTERISTICS AND GROUP ASSIGNMENT

ID	Group	Matched pair	Age (months)	Visual acuity (decimal)	Prior experience	Sex
P01	Experimental	M01	83	0.10	Minimal	M
P02	Control	M01	84	0.12	Minimal	M
P03	Experimental	M02	87	0.08	None	F
P04	Control	M02	86	0.10	None	F
P05	Experimental	M03	80	0.15	Some	M
P06	Control	M03	81	0.13	Some	M
P07	Experimental	M04	89	0.05	Minimal	F
P08	Control	M04	88	0.07	Minimal	F
P09	Experimental	M05	84	0.11	None	M
P10	Control	M05	83	0.09	None	M
P11	Experimental	M06	86	0.14	Minimal	F
P12	Control	M06	87	0.12	Minimal	F
P13	Experimental	M07	79	0.10	Some	M
P14	Control	M07	80	0.11	Some	M
P15	Experimental	M08	90	0.06	None	F
P16	Control	M08	89	0.08	None	F
Total (Experimental, n=8)		8 pairs	Range 79–90 (M≈84.8)	0.05–0.15, all WHO-confirmed low vision	None=3, Minimal=3, Some=2, (see note)	M=4, F=4
Total (Control, n=8)			Range 80–89 (M≈84.8)	0.07–0.13, all WHO-confirmed low vision	None=3, Minimal=3, Some=2, (see note)	M=4, F=4

Note: All 16 participants were assigned in eight age-, sex-, prior-experience-, and visual-acuity-matched pairs (one experimental, one control participant per pair), consistent with the matching procedure described in Section III-B. Prior experience distribution across the full sample (both groups combined) was None=6, Minimal=6, Some=4. Visual acuity was confirmed against school medical records for all participants and fell within the WHO low vision range in every case.

B. Expert Heuristic Evaluation

Table IV shows the results of the expert heuristic evaluation conducted prior to user testing, covering game heuristics and pedagogical heuristics, together with the resulting prototype refinements.

As shown in Table IV, all seven heuristic dimensions received mean expert ratings between 3.8 and 4.5 on the 5-point scale, indicating that the prototype was rated favourably across both the game and pedagogical heuristics prior to user testing. The lowest-rated dimension was Challenge (3.8), reflecting

expert concern about the difficulty increase between the Loops and Conditionals concepts, which was addressed by adding an intermediate practice level before the Conditionals concept. The Goals and Learning objectives dimensions received the highest ratings (4.5 each), and the associated recommendations, together with those for the remaining dimensions, were implemented as concrete refinements to the prototype before the user testing phase. Because each recommendation was translated into a specific design change, the version subsequently evaluated in the usability testing already incorporated the improvements identified during the heuristic evaluation.

TABLE IV. EXPERT HEURISTIC EVALUATION RESULTS

Heuristic dimension	Mean expert rating	Key recommendation	Refinement implemented
Challenge	3.8 (0.5)	Reduce the difficulty jump from loops to conditionals	Added an intermediate practice level before conditionals
Reward	4.3 (0.5)	Make success feedback more salient for low vision users	Added a distinct success chime and large green checkmark animation
Goals	4.5 (0.6)	Clarify the learning goal of each level more explicitly	Added a short audio and large-text goal statement at the start of every level
Storyline	4.0 (0.8)	Strengthen narrative continuity across the three concepts	Linked levels with a simple character-progress story and a consistent visual theme
Learning objectives	4.5 (0.6)	Make the three target concepts more explicitly visible	Display the current concept name (Sequences, Loops, Conditionals) in large high-contrast text on each level screen
Learning theory alignment	4.2 (0.5)	Better support progressive scaffolding of computational thinking constructs.	Inserted guided practice tasks that reuse previous concept blocks before introducing the next
Content alignment	4.3 (0.5)	Ensure block vocabulary matches Year-1 curriculum language	Simplified block labels and added optional audio read-aloud of each block name

Note: Each dimension was rated by the four experts on a 5-point scale (1 = poor, 5 = excellent); values are reported as mean (SD) across experts. Game heuristics and pedagogical heuristics were evaluated using two separate structured checklists, as described in Section III-A. Each key recommendation was addressed through the corresponding refinement prior to the user testing phase.

C. Usability Results

Table V shows the effectiveness, efficiency, and user satisfaction scores for the experimental and control groups,

calculated using Eq. (1)-(3) and interpreted using the bands in Table II, reported per level.

TABLE V. EFFECTIVENESS, EFFICIENCY, AND SATISFACTION SCORES BY GROUP AND LEVEL

Group	Level	Effectiveness (SD)	Efficiency (SD)	Satisfaction (SD)	Overall score	Interpretation
Experimental	1 – Sequences	4.54 (0.81)	4.22 (0.99)	4.54 (0.81)	4.43	Good
Experimental	2 – Loops	3.90 (1.11)	3.59 (1.19)	3.90 (1.11)	3.80	Good
Experimental	3 – Conditionals	3.29 (1.23)	2.97 (1.27)	3.29 (1.23)	3.18	Moderately Good
Experimental	Overall	3.91 (1.12)	3.60 (1.18)	3.91 (1.12)	3.80	Good
Control	1 – Sequences	3.12 (1.25)	2.50 (1.31)	2.81 (1.28)	2.81	Moderately Good
Control	2 – Loops	1.88 (1.25)	1.25 (1.16)	1.56 (1.21)	1.56	Weak
Control	3 – Conditionals	0.78 (0.99)	0.47 (0.81)	0.47 (0.81)	0.57	Weak
Control	Overall	1.93 (1.34)	1.41 (1.24)	1.62 (1.29)	1.65	Weak

Note: Efficiency scores are calculated using effective gameplay time, not total session duration. Sub-level scores were recorded for two task instances per concept; the values above are the mean of the two sub-level scores per concept per group. Control-group scores are computed across the full control group (n=8), following the effectiveness/efficiency/satisfaction scoring rule in Section III-C5 in which a non-completed task is scored 0; these scores therefore reflect both the performance of participants who progressed through a level and the zero scores of participants who did not reach it, and are the appropriate usability evidence for the group as assigned. The n=2 control-independent subgroup (participants who reached all three concepts unaided) is reported separately in Section IV-E for the learning-outcome comparison and is not the basis for this table. In the control group, the number of participants completing each level was 6 at Level 1, 4 at Level 2, 3 at Level 3, and 2 at each of Levels 4, 5, and 6, consistent with the per-participant stopping levels reported in Table VI.

As shown in Table V, the experimental group obtained an overall usability score of 3.80, which falls within the good band, whereas the control group obtained an overall score of 1.65, which falls within the Weak band. The experimental group scored in the good band for the Sequences (4.43) and Loops (3.80) concepts, declining to the Moderately Good band for the Conditionals concept (3.18). The control group, in contrast, scored in the Moderately Good band only for Sequences (2.81) and fell into the Weak band for both Loops (1.56) and Conditionals (0.57). This pattern indicates that usability declined with increasing programming-concept complexity in both groups, but that the decline was substantially steeper for the control group, which used the feature-reduced baseline version without the accessibility elements listed in Table I.

The same downward trend is observed across all three usability components. For the experimental group, effectiveness declined from 4.54 at Sequences to 3.29 at Conditionals, efficiency from 4.22 to 2.97, and satisfaction from 4.54 to 3.29. For the control group, the corresponding declines were steeper,

with effectiveness falling from 3.12 to 0.78, efficiency from 2.50 to 0.47, and satisfaction from 2.81 to 0.47. Across all three components and all three concepts, the experimental group scored higher than the control group, consistent with the accessibility elements contributing to usability for low-vision children playing a block-based programming game.

Effective gameplay time, excluding pauses and check-ins, averaged 24 min 15 s for the experimental group and 16 min 52 s for the control group, out of an equal total session duration of approximately 28 min 40 s for both groups. The two groups therefore received effectively equal session time, but the control group spent proportionally more of that time in pauses, check-ins, and early termination, with active gameplay accounting for 58.9% of the session for the control group compared with 84.5% for the experimental group. Because the efficiency scores in Table V are computed from effective gameplay time rather than total session duration, this difference is reflected in the efficiency component but does not arise from unequal time allocation between groups.

D. Concept Exposure in the Control Group

Table VI shows, for the control group, whether each participant independently reached the levels covering all three core concepts within the session; this outcome is reported as a usability finding, not merely as a precondition for inclusion in the learning-gain analysis.

As shown in Table VI, only two of the eight control-group participants (25%) independently reached the levels covering all three core programming concepts within the session. The remaining six participants stopped at an earlier level: two stopped at Level 1, two at Level 2, one at Level 3, and one at Level 4, with no participant stopping at Level 5. The attrition was concentrated within the Sequences concept itself, where four of the eight participants stopped at Level 1 or Level 2 without progressing to the Loops concept. This concept-exposure outcome is reported here as a usability finding, since the proportion of control participants unable to reach a given concept directly quantifies the severity of the barriers present in the unadopted baseline environment and does so in a way that the group-mean usability scores in Table V, which average completed and non-completed tasks together, do not fully convey.

TABLE VI. CONTROL GROUP CONCEPT-EXPOSURE OUTCOMES

ID	Reached all 3 concepts?	Last level reached	Reason for non-completion
P06	Yes	Level 6 (Conditionals)	—
P14	Yes	Level 6 (Conditionals)	—
P02	No	Level 1 (Sequences)	Unable to place the first block using mouse drag-and-drop after multiple attempts; stopped interacting..
P04	No	Level 1 (Sequences)	Could not distinguish the low-contrast blocks from the background; repeated wrong selections
P08	No	Level 2 (Sequences)	Lost track of the block sequence without audio cues; five consecutive failed attempts, session ended early after check-in
P10	No	Level 2 (Sequences)	Unable to complete drag-and-drop placement of the second block; disengaged after repeated failed attempts
P12	No	Level 3 (Loops)	Could not select and connect the loop block with the mouse; five consecutive failed attempts; session ended early after check-in
P16	No	Level 4 (Loops)	Lost track of the required loop repetitions without audio feedback; stopped progressing after repeated incorrect placements
Total reaching full exposure: 2 / 8 (25%)			

Note: The last level reached, and the reason for non-completion for each participant were recorded from the session observation checklists and screen recordings referenced in Section III-C6. Reasons are reported as observed interaction behaviours (for example, inability to complete a drag-and-drop placement, or repeated wrong selections of low-contrast blocks) rather than inferred internal states.

E. Learning Outcome Results

Table VII reports individual-level pre-test and post-test scores for all experimental group participants and for control group participants who independently reached concept exposure.

As shown in Table VII, all eight experimental-group participants recorded a positive learning gain. For the experimental group, the pre-test mean was 1.75 and the post-test mean was 7.44, giving a mean gain of 5.69 (SD=0.53), with individual gains ranging from +5.0 to +6.5. No experimental-group participant recorded a gain below +5.0, indicating a consistent improvement across the group.

Because the control-independent subgroup comprises only two participants, fewer than the pre-registered threshold of three specified in Section III-D3, the corresponding group-level learning-gain comparison is omitted, and only the individual case data are reported for this subgroup. Participants P06 and P14 recorded gains of +2.5 and +2.0, respectively. Both gains are smaller than that of any experimental-group participant, and both participants began from a pre-test range (1.5 to 2.0) comparable to that of several experimental-group participants, such as P01 and P11 (pre-test 2.5).

TABLE VII. INDIVIDUAL PRE-TEST/POST-TEST SCORES

ID	Group	Pre-test	Post-test	Gain
P01	Experimental	2.5	8.0	+5.5
P03	Experimental	1.0	7.5	+6.5
P05	Experimental	3.0	9.0	+6.0
P07	Experimental	2.0	7.0	+5.0
P09	Experimental	1.5	6.5	+5.0
P11	Experimental	2.5	8.5	+6.0
P13	Experimental	0.5	6.0	+5.5
P15	Experimental	1.0	7.0	+6.0
P06	Control-independent	2.0	4.5	+2.5
P14	Control-independent	1.5	3.5	+2.0

Note: The control-independent subgroup is not representative of the control group as originally assigned, since it comprises only the two participants (of eight) who were able to independently reach concept exposure in the baseline environment. Any comparison drawn between this subgroup and the experimental group is therefore interpreted as a conservative estimate of the accessibility effect (Section III-D3): the six control-group participants who did not reach concept exposure are absent from this comparison entirely, and their exclusion is itself evidence of a barrier severity that a group-mean comparison would not capture.

F. Termination-Criterion Events

Table VIII summarises how often the proactive termination criteria described in Section III-C6 were triggered during testing, and their outcomes.

TABLE VIII. TERMINATION-CRITERION EVENTS BY GROUP

Group	"3 min no progress" triggers	"5 consecutive failed attempts" triggers	Continued after check-in	Took a break	Session ended early
Experimental	3	2	4	1	0
Control	11	9	2	3	9

As shown in Table VIII, the control group triggered a proactive check-in substantially more often than the experimental group, with 20 triggers in total compared with 5. The outcome of these check-ins also differed between the two groups. In the experimental group, most check-ins led to the session continuing: of the five check-in occasions, four led to continuation and one led to a break, and no experimental-group session ended early. In the control group, by contrast, the session ended early on 9 of the 14 check-in occasions. All trigger events were logged with a timestamp during testing; for example, participant P04 triggered the “3-minute no-progress” criterion at 00:08:42 into the session. These timestamps are retained in the raw observation records but are not reported individually.

G. Inter-Rater Reliability

Before formal scoring, both the primary rater and the second rater underwent a short calibration session: the raters independently scored a small set of practice responses drawn from the pilot testing phase (not included in the main analysis) and then compared their scores, discussing discrepancies until consensus was reached on the application of the scoring rubric described in Section III-D (one point per item, nine items in total). After this calibration, to assess scoring reliability, 30% of all pre-test and post-test responses (stratified by participant and by test occasion) were independently re-scored by the second rater, who remained blinded to group assignment and to the primary rater’s scores. Inter-rater agreement calculated with Cohen’s κ was 0.82, indicating almost perfect agreement (Landis & Koch, 1977), supporting the reliability of the scoring approach used for the primary rater’s results reported in Table VII.

V. DISCUSSION

A. Usability of the Accessible Version

The experimental group achieved a Good overall usability rating (3.80) for the accessible version, consistent with the range reported for accessible serious games for low vision children in prior work [3]. However, the score was not uniform across concepts: usability was Good for Sequences (4.43) and Loops (3.80) but declined to Moderately Good for Conditionals (3.18), and the same downward pattern was present in effectiveness, efficiency, and satisfaction scores alike. This suggests that the accessibility elements applied large, high-contrast screen design, adjustable colour and text, and audio feedback supported usability effectively for lower-complexity tasks, but that usability nonetheless degrades as programming-concept complexity increases, even when accessibility barriers related to screen design and multimedia are addressed. This is consistent with prior findings that screen design elements such as menus and buttons contribute directly to interface usability [10]. Because the accessibility package (Table I) remained constant across all levels while only task complexity varied, the within-group decline is more plausibly attributable to the cognitive demands of conditional logic itself than to a breakdown in the accessibility adaptations at higher complexity. This is a favourable finding relative to the interaction barrier reported for drag-and-drop block assembly in prior block-based programming accessibility work [5]: the accessible version, which incorporates a keyboard-based selection method alongside its other accessibility adaptations, appears to have

kept the interaction itself usable across all three concepts, with the residual difficulty concentrated in the conceptual content rather than the input method. Because the present design varies screen design, multimedia, interaction method, and text simultaneously, this observation cannot isolate the specific contribution of the keyboard-based interaction component from the accessibility package.

B. Severity of Baseline Barriers and the Value of the Control Condition

Only two of the eight control-group participants (25%) reached concept exposure to all three programming concepts independently in the feature-reduced baseline version (Table VI); the remaining six stopped at some point between Level 1 and Level 4, with the attrition concentrated within the Sequences concept itself, where four of the eight participants stopped at Level 1 or Level 2. This is a substantially more severe barrier than the usability scores alone would suggest: the control-group overall usability score fell from Moderately Good for Sequences (2.81) to Weak for both Loops (1.56) and Conditionals (0.57), but because these scores are computed across all eight participants, including zero-scored non-completions, the numeric decline understates the practical severity captured in Table VI: a quarter of participants, not a numeric average, is arguably the more informative statistic for judging whether the baseline environment is usable at all for this population. The reasons for non-completion recorded in Table VI point to the interaction method and the screen design as the two most frequently observed points of difficulty: four of the six non-completing participants were recorded as unable to complete mouse-based drag-and-drop block placement, and the remaining difficulties involved distinguishing low-contrast blocks and losing track of the task without audio cues. Since both versions differ only in the accessibility elements listed in Table I (screen design, multimedia, interaction method, and text), and the control-group attrition is concentrated at the earliest levels while the experimental group performed well on the same concepts, the pattern is consistent with the removed accessibility package as a whole rather than the increase in conceptual complexity alone being the primary driver of the gap, since conceptual complexity was held constant between conditions. Because the accessible and baseline versions differ in multiple accessibility elements simultaneously (Table I), the specific contribution of the interaction method relative to the other adaptations cannot be isolated from the present design.

C. Learning Outcomes

All eight experimental-group participants showed a positive learning gain, ranging from +5.0 to +6.5 points ($M=5.69$, $SD=0.53$), a consistent pattern with no participant falling below the group’s lower bound. The two control-independent participants who reached full concept exposure in the baseline version showed smaller gains (+2.5 and +2.0). As established in Section III-D3, this comparison must be read as a conservative estimate of the accessibility effect rather than a representative group comparison: it includes only the two control participants best positioned to learn from the baseline environment (Table VI), while the six control participants who did not reach concept exposure and who, by definition, had markedly lower learning opportunity are absent from the comparison entirely. The true

gap between the accessible and baseline conditions is therefore likely larger than the individual-level gain figures alone suggest.

Relating this to the usability findings in Section V-A, the experimental group's gains remained consistently high ($\geq +5.0$) even for Conditionals, the concept at which usability had declined to Moderately Good rather than Good. This suggests that, within the accessible version, the residual usability decline observed for more complex concepts did not translate into a proportional decline in learning outcome: a Moderately Good usability level for a given concept still supported a learning gain comparable to concepts rated Good. This pattern would benefit from confirmation with a larger and more complexity-differentiated sample, given the descriptive nature of this dataset.

D. Design Tensions Between Playability and Accessibility

The termination-criterion event data provide indirect evidence of one such tension: keyboard navigation speed relative to task completion time. The experimental group's efficiency scores declined from 4.22 (Sequences) to 2.97 (Conditionals) even though effectiveness (task success) remained comparatively higher (4.54 to 3.29), a gap plausibly consistent with the keyboard-based selection method requiring more discrete key presses to construct increasingly complex block arrangements, trading interaction precision and accessibility for the raw speed of a direct drag-and-drop action; as with the other component-level observations in this study, this specific mechanism cannot be isolated from the accessibility package as a whole (Table I). This did not, however, translate into disengagement: experimental-group participants who triggered a proactive check-in continued the session in four of five instances, versus nine of fourteen check-in instances in the control group ending in early termination, suggesting that the accessible version's slower interaction pace was tolerated by children rather than experienced as frustrating enough to prompt withdrawal. This extends the accessibility model of [11] originally validated for a tap-to-select mathematics game to the specifically higher-friction case of block-based programming, indicating that the accessible version's design, including its keyboard-based interaction method, can preserve engagement even where it does not match the raw efficiency of unrestricted pointer-based interaction. A full account of the trade-off between audio feedback and game immersion, and between visual simplification and game attractiveness, requires the qualitative observation notes referenced in Section III-C6 and is not addressed by the quantitative data reported here.

E. Comparison with Prior Work

The overall effectiveness (3.91), efficiency (3.60), and satisfaction (3.91) scores obtained for the accessible version in this study are all lower than the corresponding scores reported for an accessible mathematics serious game for the same population (4.74, 4.55, and 4.92, respectively) [3]. Because both studies applied a broadly similar accessibility model, screen design, multimedia, and language adaptations derived from [11], this gap is more plausibly attributable to the additional interaction barrier specific to block-based programming than to a difference in the underlying accessibility approach: a mathematics game in which children select and tap an answer does not require the sequential, multi-step block construction

demanding by sequences, loops, and conditionals, even when both interactions are mediated through an accessible, non-drag-and-drop input method. This is consistent with the distinction drawn in Section I between accessibility barriers that screen design and multimedia adaptations alone can resolve, and the interaction-specific barrier that block-based programming introduces regardless of those adaptations [20]. Relative to the fully non-visual, audio-based approach used for blind children in [5], the present results suggest that the accessible version's overall approach, preserving visual access through a keyboard-based interaction method combined with the other accessibility adaptations in Table I, rather than removing the visual channel entirely, is a viable middle path for the residual-vision case of low vision children, achieving Good overall usability despite the added interaction complexity of block-based programming. This is broadly consistent with assistive-technology approaches for mathematics learning among visually impaired populations, which similarly favour input methods that preserve residual vision where possible [28].

F. Limitations

This study has several limitations. First, the sample size ($n=16$, split into matched experimental and control groups of 8 participants each) is small; findings are reported descriptively and are not intended to support statistical generalisation. Second, participants were recruited from a single special education school; although experimental and control sessions were scheduled on separate days to minimise cross-group contamination, the sample reflects the specific educational and cultural context of that institution. Third, children with additional cognitive or motor impairments were excluded to reduce confounding, which limits the applicability of the findings to low vision children with co-occurring conditions. Fourth, although both groups received an equal fixed session duration, this does not imply an equal experience within that time: an experimental group participant who completed the core levels early spent the remainder of the session on non-core decorative activities, while a control group participant who struggled to progress may have spent most or all of the session attempting the first core level without success; this qualitative difference in what the time was spent on is a limitation of the equal-duration design. Fifth, the learning-gain comparison reported as individual case data in Table VII for the control-independent subgroup is subject to survivorship bias. Sixth, this study assessed learning gain immediately after the session; it does not assess retention of programming concepts over time. Seventh, beyond parental or guardian consent and the children's right to withdraw described in Section III-B, the study protocol did not include a separate, age-appropriate child assent procedure, nor did it specify formal data-protection arrangements for the screen and camera recordings collected during testing; both are noted here as procedural gaps that future replications of this design, particularly given the involvement of children with disabilities, should address explicitly.

VI. CONCLUSION AND FUTURE WORK

In this study, a block-based programming serious game was designed for low-vision children aged seven years old, incorporating accessibility elements such as screen design, multimedia, and a keyboard-based interaction method in place

of the drag-and-drop interaction typically used in block-based programming environments. The prototype was refined through pilot testing and expert heuristic evaluation before user testing. Usability testing and a pre-test/post-test learning outcome assessment were conducted, benchmarked against a feature-reduced baseline version of the game. The accessible version achieved a Good overall usability score (3.80), compared to a Weak overall score (1.65) for the feature-reduced baseline version when computed across the full control group. Only 25% of control-group participants reached independent exposure to all three target concepts in the baseline environment, compared with 100% of experimental-group participants in the accessible version. Experimental-group participants showed a consistent learning gain ($M=5.69$, $SD=0.53$, range $+5.0$ to $+6.5$), descriptively larger than the gains observed for the two control-group participants who reached concept exposure independently ($+2.5$ and $+2.0$), a comparison interpreted conservatively due to survivorship bias in the control-independent subgroup. Accessibility elements such as screen design, multimedia, and interaction method are important in increasing the usability of block-based programming serious games for low vision children. Future research could extend this design to a larger, multi-site sample of low vision children, incorporate a delayed post-test to assess retention of programming concepts, and explore extending the accessibility model to children with additional visual or motor impairments, building on related efforts to extend computational-thinking tools to younger children [27] and to children with other developmental disabilities such as autism spectrum conditions [29].

ACKNOWLEDGMENT

The authors would like to express their sincere gratitude to Universiti Sains Malaysia for the invaluable support provided throughout the course of this research. This study would not have been possible without their encouragement and contributions.

DECLARATION ON GENERATIVE AI-ASSISTED TOOLS

Generative AI tools were used for language editing, formatting assistance, and organisation of the manuscript during preparation. The authors critically reviewed, edited, and approved the study design, data collection, analysis, interpretation, references, and final manuscript, and take full responsibility for the scientific content.

REFERENCES

- [1] Y. Zhonggen, "A Meta-Analysis of Use of Serious Games in Education over a Decade," *International Journal of Computer Games Technology*, vol. 2019, Article ID 4797032, 8 pages, 2019, doi: 10.1155/2019/4797032.
- [2] L. Sun and C. E. Pan, "Effects of the Application of Information Technology to E-Book Learning on Learning Motivation and Effectiveness," *Frontiers in Psychology*, vol. 12, Article 752303, 2021, doi: 10.3389/fpsyg.2021.752303.
- [3] N. I. Othman, H. Mohamed, and N. A. M. Zin, "Testing the Usability of Serious Game for Low Vision Children," *Int. J. Adv. Comput. Sci. Appl.*, vol. 14, no. 8, pp. 310–317, 2023, doi: 10.14569/IJACSA.2023.0140834.
- [4] J. Maloney, M. Resnick, N. Rusk, B. Silverman, and E. Eastmond, "The Scratch Programming Language and Environment," *ACM Trans. Comput. Educ.*, vol. 10, no. 4, pp. 1–15, 2010, doi: 10.1145/1868358.1868363.
- [5] L. R. Milne and R. E. Ladner, "Blocks4All: Overcoming Accessibility Barriers to Blocks Programming for Children with Visual Impairments," in *Proc. 2018 CHI Conf. Human Factors in Computing Systems*, 2018, pp. 1–10.
- [6] A. Donker and P. Reitsma, "Usability testing with young children," *Proc. 2004 Conf. Interact. Des. Child. Build. a Community, IDC 2004*, pp. 43–48, 2004, doi: 10.1145/1017833.1017839.
- [7] M. M. Ali-Shahid and S. Sulaiman, "A case study on reliability and usability testing of a Web portal," *2015 9th Malaysian Softw. Eng. Conf. MySEC 2015*, pp. 31–36, 2016, doi: 10.1109/MySEC.2015.7475191.
- [8] N. I. Othman, N. A. M. Zin, and H. Mohamed, "Play-Centric Designing of a Serious Game Prototype for Low Vision Children," *Int. J. Adv. Comput. Sci. Appl.*, vol. 11, no. 5, pp. 199–205, 2020, doi: 10.14569/IJACSA.2020.0110528.
- [9] H. Y. Abuaddous, M. Z. Jali, and N. Basir, "Web Accessibility Challenges," *Int. J. Adv. Comput. Sci. Appl.*, vol. 7, no. 10, 2016, doi: 10.14569/IJACSA.2016.071023.
- [10] K. K. Allah, N. A. Ismail, L. Hasan, and W. Y. Leng, "Usability Evaluation of Web Search User Interfaces from the Elderly Perspective," *Int. J. Adv. Comput. Sci. Appl.*, vol. 12, no. 12, pp. 647–657, 2021, doi: 10.14569/IJACSA.2021.0121281.
- [11] N. I. Othman, H. Mohamed, and N. A. M. Zin, "Serious Games Accessibility Design Model for Low-Vision Children," *Advances in Human-Computer Interaction*, vol. 2023, Article ID 9528294, 2023, doi: 10.1155/2023/9528294.
- [12] A. Stefik, W. Allee, G. Contreras, T. Kluthe, A. Hoffman, B. Blaser, and R. Ladner, "Accessible to Whom? Bringing Accessibility to Blocks," in *Proc. 55th ACM Tech. Symp. Comput. Sci. Educ.*, vol. 1, 2024, pp. 1286–1292, doi: 10.1145/3626252.3630770.
- [13] A. Mountapmbeme and S. Ludi, "Designing Accessible Block-Based Programming Environments for Persons with Visual Impairments," *Universal Access in the Information Society*, vol. 24, no. 3, pp. 2351–2382, 2025, doi: 10.1007/s10209-025-01200-y.
- [14] K. Teng and G. K. W. K. Chung, "Measuring Children's Computational Thinking and Problem-Solving in a Block-Based Programming Game," *Education Sciences*, vol. 15, no. 1, Article 51, 2025, doi: 10.3390/educsci15010051.
- [15] L. Wöbbeckind, K. Lorberg, and T. Mandl, "Emma Stop That, It's My Turn Now — Comparing Peer Tutoring and Thinking Aloud for Usability-Testing with Children in a School Setting," in *Proc. Mensch und Computer 2023*, 2023, pp. 442–447, doi: 10.1145/3603555.3608542.
- [16] F. Lehnert, S. Doublet, and G. Sim, "Designing Usability/User Experience Heuristics to Evaluate e-Assessments Administered to Children," *Computer Standards & Interfaces*, vol. 92, Article 103933, 2024, doi: 10.1016/j.csi.2024.103933.
- [17] R. Wang, L. Zeng, X. Zhang, S. Mondal, and Y. Zhao, "Understanding How Low Vision People Read Using Eye Tracking," in *Proc. 2023 CHI Conf. Human Factors in Computing Systems*, 2023, pp. 1–17, doi: 10.1145/3544548.3581213.
- [18] Y. Wang, Y. Zhao, and Y.-S. Kim, "How Do Low-Vision Individuals Experience Information Visualization?," in *Proc. 2024 CHI Conf. Human Factors in Computing Systems*, 2024, Article 43, doi: 10.1145/3613904.3642188.
- [19] R. Cossoovich, S. Hodges, J. Kang, and A. Girouard, "Co-designing New Keyboard and Mouse Solutions with People Living with Motor Impairments," in *Proc. 25th Int. ACM SIGACCESS Conf. Computers and Accessibility*, 2023, Article 98, doi: 10.1145/3597638.3614549.
- [20] R. H. Mollik, W. Aljedaani, and S. Ludi, "Understanding Design and Accessibility Issues in Block-Based Programming for People with Disabilities: A Literature Review," *Universal Access in the Information Society*, vol. 25, no. 3, Article 94, 2026, doi: 10.1007/s10209-026-01351-6.
- [21] K. Żyła, K. Chwaleba, and D. Choma, "Evaluating Usability and Accessibility of Visual Programming Tools for Novice Programmers—The Case of App Inventor, Scratch, and StarLogo," *Applied Sciences*, vol. 14, no. 21, Article 9887, 2024, doi: 10.3390/app14219887.
- [22] M. S. Alotaibi and M. B. Garcia, "Game-Based Learning in Early Childhood Education: A Systematic Review and Meta-Analysis," *Frontiers in Psychology*, vol. 15, Article 1307881, 2024, doi: 10.3389/fpsyg.2024.1307881.

- [23] M. Spinoso, M. Benassi, N. Mazzoni, M. Orsoni, L. Stefanutti, P. Anselmi, D. de Chiusole, A. Bacherini, I. Pierluigi, S. Garofalo, G. Balboni, and S. Giovagnoli, "Psychometric Properties of the Usability and Acceptability Scale (UAS) for Evaluating Digital Tools in Children and Adolescent Users," *Frontiers in Psychology*, vol. 17, Article 1702085, 2026, doi: 10.3389/fpsyg.2026.1702085.
- [24] T. Tene, D. F. Vique López, P. E. Valverde Aguirre, N. I. Cabezas Oviedo, C. Vacacela Gomez, and S. Bellucci, "A Systematic Review of Serious Games as Tools for STEM Education," *Frontiers in Education*, vol. 10, Article 1432982, 2025, doi: 10.3389/educ.2025.1432982.
- [25] S. W. Gundersen and G. Lampropoulos, "Using Serious Games and Digital Games to Improve Students' Computational Thinking and Programming Skills in K-12 Education: A Systematic Literature Review," *Technologies*, vol. 13, no. 3, Article 113, 2025, doi: 10.3390/technologies13030113.
- [26] X. Wang, M. Cheng, and X. Li, "Teaching and Learning Computational Thinking Through Game-Based Learning: A Systematic Review," *Journal of Educational Computing Research*, 2023, doi: 10.1177/07356331231180951.
- [27] G. P. Perez Valdes, O. Boude Figueredo, and A. D. Vargas Sanchez, "Integrating Computational Thinking in Children Aged 3 to 6: Challenges and Opportunities in Early Childhood Education," *Frontiers in Education*, vol. 10, Article 1535135, 2025, doi: 10.3389/educ.2025.1535135.
- [28] M. Shoaib, D. Fitzpatrick, and I. Pitt, "Assistive Technology-Based Solutions in Learning Mathematics for Visually-Impaired People: Exploring Issues, Challenges and Opportunities," *Multimedia Tools and Applications*, vol. 82, pp. 46153–46184, 2023, doi: 10.1007/s11042-023-17409-z.
- [29] K. F. Carhuaz Malpartida and K. Rios da Hora Rodrigues, "Building Serious Games to Exercise Computational Thinking: Initial Evaluation with Teachers of Children on the Autism Spectrum," *Journal on Interactive Systems*, vol. 16, no. 1, 2025, doi: 10.5753/jis.2025.4492.
- [30] ISO 9241-11:1998, *Ergonomic requirements for office work with visual display terminals (VDTs) — Part 11: Guidance on usability*, International Organization for Standardization, Geneva, 1998. (Superseded by ISO 9241-11:2018)
- [31] University of Pittsburgh Department of Ophthalmology, "What is Vision Impairment?" [Online]. Available: <https://www.ophtalmology.pitt.edu/vision-impairment/what-vision-impairment>. [Accessed: Sep. 10, 2026].