3rd International Conference of Bunga Bangsa Cirebon (ICOBBA)
PROCEEDINGS ICOBBA
ISSN : 3032-5641, DOI: 10.47453
Web Journal : https://journal.ljpi.bbc.ac.id/icobba/index

Developing an Interaction-Based Digital Puzzle Game To Foster Critical Thinking in Children

Suzana[1]✉, Julaeha[2], Supriyatun[3]

Universitas Islam Bunga Bangsa Cirebon
Email : suzana311000@gmail.com

Received: 2025-09-13; Accepted: 2025-10-10; Published: 2025-10-31

ABSTRACT

This research is motivated by the lack of development and utilization of learning media on the aspect of digital-based critical thinking, so this study aims to develop digital puzzle educational game media based on Interacty for early childhood in Kindergartens in Cirebon Regency. Research and development (R&D) method with the ADDIE Model. The research procedure goes through the stages of Problem Needs Analysis and literature review, data collection, Product Design, Design Validation, Design Revision, Small-scale Trial, Product revision, large-scale trial, product revision and production. The results of small-scale testing show that the effect size of the difference between the pretest and posttest is moderate to large (around 0.87 to 0.89), which shows a practically meaningful difference, not just statistics and the results of large-scale testing show Cohen's d = -1.672 indicates a very large effect (because the value far exceeds 0.8), and the negative sign indicates that the posttest value is higher than the pretest (because of the Pretest - Posttest sequence). The Hedges' correction value = -1.653 also confirms a very large effect, with minor adjustments to address sample size bias. The 95% confidence intervals for both do not cross zero and are all negative, strengthening the evidence that there is a significant and large effect of the treatment given. In conclusion, the Interacty-based digital puzzle educational game is feasible to use and effective in improving critical thinking skills of early childhood in Kindergartens in Cirebon Regency.

Keywords: *Educational Game, Digital Puzzle, Interacty, Early Childhood, Critical Thinking, ADDIE*

INTRODUCTION

Technological developments have not been optimally utilized in early childhood learning. Children aged 5–6 years are in a golden age that requires fun and meaningful cognitive stimulation. One important skill that must be developed is critical thinking, such as the ability to differentiate, organize

patterns, and make decisions. This research developed an Interactive-based educational game to facilitate interactive learning that can stimulate children's critical thinking skills. Early childhood is an individual with unique characteristics and characteristics that differ according to their developmental stage. The age period (0–6) is the golden age, during which all aspects of development play a crucial role in subsequent development (Uswatun et al., 2022).

Children aged 0–6 years experience very rapid growth, and development is crucial for subsequent development. This is known as the golden age. This age is crucial for determining how a child will develop in the future (Shofia & Dadan, 2021). Adolescence is the best time for a person to develop a strong character that will last into adulthood, making it difficult for them to be influenced by things that don't align with their basic character (Hasanah & Fajri, 2022).

From the above description, early childhood is defined as a period when children begin crucial or fundamental developments that will support their future development. To ensure strong, character-driven early childhood growth, stimulation is necessary, tailored to the child's individual characteristics. The government pays close attention to this, and therefore, numerous regulations and basic benchmarks are established for early childhood learning activities.

Based on the researcher's observations at the research site, several problems were faced by the school: conventional learning methods, suboptimal use of digital media, the absence of teacher-created digital media, and children's preference for audio-visual/digital learning. Therefore, the researcher will develop a digital-based educational game.

a. Educational Games

Games are both a necessity and a problem for computer users, as to run them comfortably, a computer must have higher specifications than those required for everyday use. Therefore, according to Purnomo, a game is a program designed to meet human needs (Purnomo, 2020). Learning games are a type of multimedia-based learning media developed with the aim of increasing motivation and learning effectiveness. Learning games are typically designed to combine game elements and learning materials so that students can learn while playing (Safitri et al., 2023-70). According to Prensky, as described by Karseno (2021), educational games are games designed to teach specific subjects and specific skills (Karseno et al., 2021). Educational games are also games used as a means of entertainment for children and contain educational content to facilitate the delivery of learning materials, with the goal of making children more responsible, intelligent, and skilled (Pradana & Nita, 2019). According to Sutarman et al. (2022), educational games are games specifically designed to

provide understanding of a topic, broaden knowledge, expand concepts, introduce historical or cultural events, and practice specific skills through enjoyable play experiences. Educational games are games that incorporate elements of education and learning. Furthermore, students can learn various things and feel entertained while playing them (Hendrawan & Marlina, 2022).

From the expert explanations above, it can be concluded that educational games can help develop various aspects of a child's development, such as cognitive, social, and emotional skills. They can also transform a child's potential energy into various abilities. Furthermore, the use of educational games in accordance with the Early Childhood Education guidelines for Playing While Learning and learning while playing. Educational games can also help teachers develop all basic AUD skills in a way that is not boring. One educational game that can be developed is the Digital Puzzle game.

b. Digital Puzzle

The Big Indonesian Dictionary defines a puzzle as "teka rebus." A puzzle is a game that requires perseverance and patience from children to solve. Puzzles are small pieces consisting of two or more pieces made of cardboard or wood. One factor that encourages children to try new things is the satisfaction they gain from completing a puzzle (Widyanarti, 2013). According to Bahar & Soegiarto (2019:3) in Nuryanti et al., 2022, puzzles are play tools that train students' psychomotor skills and improve hand-eye coordination and assembly concepts in the form of two- or three-person disassembly games. Puzzle games are games played by one or more people consisting of pieces or pieces of a picture that can be assembled (Nuryanti et al., 2022). According to Musthofa (in Sumani et al., 2022), puzzles are pieces of pictures with difficulty levels tailored to the child's development. Puzzles are a type of game that requires precision and trains children to concentrate when assembling puzzle pieces to form a complete and complete picture. (Sumani et al., 2022).

According to Kozachenko and Davydova (2019), maze puzzle games can improve children's intellectual abilities. Going through maze-shaped puzzles has a beneficial effect on children's intelligence, developing spatial thinking, memory, and teaching children how to find solutions to problems at a subconscious level, as well as discipline and developing concentration. (Kozachenko & Davydova, 2019). This article was taken from the iPusnas application with the title "Guidebook for making learning instruments with a puzzle maker" on January 24, 2025 at 10:46 PM WIB. Author: Bunga Ayu Dibtasari, Achmad Chusairi, Galih Yuningtyas, Slamet Arifin. Publisher: PT Literasi Nusantara Abadi Group. Year of Publication: 2023 Pages: 36. (Ayu Bunga Dibtasari et al., 2023). Situmorang (2012) in Yulinda & Saifuddin (2022) stated

that puzzles are activities involving assembling pieces of a puzzle that can motivate students to solve the puzzle correctly and quickly. Learning through puzzles provides a visual activity for students, where they must assemble the puzzles in a fun way, allowing them to learn while playing (Yulinda & Saifuddin, 2022).

Designing an interactive game is not easy. Teachers must adjust the difficulty level, material, objectives, values to be conveyed, and the game design. This can certainly take up a significant amount of teachers' time. Therefore, a website called interacty.me was created to assist teachers in designing interactive games.

Interacty.me is a digital platform that provides interactive content creation services. Teachers, as designers, have a variety of game templates available that can be customized to suit their needs and tailored to the material being taught. Interacty.me offers simple game creation proven to capture students' attention and foster a competitive spirit. Examples of available game templates that can be edited by teachers include crosswords, treasure hunts, coin flips, and many other simple games. (San et al., 2024)

So, based on the explanation above, puzzles are games that require hand-eye coordination. They consist of pieces from a picture that can train creativity, organization, and concentration, while also enhancing children's intellectual abilities. One website that helps teachers create engaging games with numerous examples is interacty.me. According to experts, the following indicators should be present in a digital puzzle:

Digital Puzzle Indicator Table

No.	Expert	Indicator
1.	Nurfadilah et al., 2021	Indicators for selecting digital puzzle educational games include: Purpose of using the educational game, Target children who will play, Safety and security, Type of game media, Skills to be developed.
2.	Rahmadhea, 2024	Indicators for selecting educational games for teachers: Educational games must be designed to be not only entertaining but also educational. They must align with the applicable curriculum and educational standards, consider student needs and characteristics. They must consider the

		psychological and social aspects of using educational games. Finally, the game must be able to increase student motivation and engagement.
3.	Khadijah, 2015	1. Learning Objectives, 2. Effectiveness, 3. Students/children, 4. Availability, 5. Technical Quality, 6. Cost, 7. Flexibility, 8. User Ability, 9. Available Time Allocation.

The indicators above demonstrate that all of these contribute to achieving learning objectives in the cognitive domain, particularly critical thinking skills.

c. Cognitive/Critical Thinking

1. Cognitive Development Achievements in Children Aged 5-6 Years

According to Law Number 20 of 2003 concerning the Indonesian National Education System, Early Childhood Education (PAUD) is an education that is part of every person's learning experience throughout their life, from early childhood to adulthood (Nur Kholisoh et al., 2022). In PAUD, children aged 5-6 are taught a variety of basic knowledge and advanced skills that help them continue their education to elementary school (Ayuni et al., 2022). According to Jean Piaget, in Yus et al. (2023), between 1896 and 1980, children's cognitive development is divided into four stages: (1) the sensorimotor stage, which is between 0 and 2 years of age. (2) the pre-operational stage, which is between 2 and 7 years of age. (3) Concrete operational stage, between the ages of 7 to 12 years. (4). Formal-functional operational phase, namely the development of cognitive systems that occur between the ages of 12 and adulthood.

Children have the ability to coordinate two types of cognitive skills simultaneously and sequentially, namely: 1) the ability to use hypotheses; 2) the ability to apply abstract principles. (Daehler & Bukatko, 1985; Best, 1989; Anderson, 1990). (Yus et al., 2023). In addition to experts from the government also determine critical thinking as one of the developmental achievements for early childhood. Early childhood development achievements are stated in Permendikbudristek no. 12 of 2024 which explains that there is a Curriculum Structure in Early Childhood Education or other equivalent forms, for the basis of cognitive development achievements are in: elements of Basic Literacy,

Mathematics, Science, Technology, Engineering, and Art. One of the indicators of the level of cognitive development achievement for early childhood is (1) learning and problem solving, which includes the ability to solve simple problems in everyday life in a flexible and socially acceptable way and apply knowledge or experience in new contexts, (2) logical, critical and creative thinking, including the ability to connect, similarities, differences, classification, patterns, take initiative, plan, and recognize cause and effect, (3) symbolic thinking, including the ability to recognize, name, and use the concept of numbers or letters, and be able to present various objects and imagination in the form of works. (Mardianingsih & Nuris, 2022)

According to Wolfolk (1995:21 in Salsabila Samsyah & Dheasari, 2022), the ability to gather and use knowledge for easy problem-solving and adaptability to the environment. Children's cognitive abilities depend on their cognitive development. This includes knowing and recognizing colors, numbers, and the names of objects, using objects according to their function, understanding their environment, and solving simple problems in everyday life. Santrock (2010:190) also stated that the ZPD is a term used by Vygotsky to describe tasks that are too difficult for children to master alone but can be learned with the guidance and assistance of adults or more skilled children. The second element of Vygotsky's theory is scaffolding. Scaffolding means changing the level of support. When a child is learning, a teacher or parent should adapt the material to the child's current performance. When a child learns a new concept, adults (teachers, parents) can be directly involved to help the child learn to master the new concept (Sutisna & Laiya, 2020). From From the explanation above, it can be concluded that cognitive is one of the developmental achievements of early childhood which must be stimulated with several stages and several levels of achievement.

2. Characteristics of Critical Thinking in Cognitive Development

Zakaria et al. wrote several expert opinions on critical thinking. Experts who evaluate critical thinking taxonomies explain that critical thinking consists of mental processes, decision-making, and learning new concepts (Robert & Price, 1986). Ennis conducted research on critical thinking techniques and the representations people use to solve problems. The result was a philosophical taxonomy that combines educational traditions, philosophy, and critical thinking. This taxonomy stems from a combination of dispositions toward critical thinking with various critical thinking abilities (Ennis, 1985:23). The taxonomy (Sternberg, 1985) divides critical idea-generation skills into three types: meta-components, performance components, and lower-level knowledge. Next are the components of understanding, knowledge acquisition, application, analysis, and synthesis, which are prioritized in Bloom's educational taxonomy (Bloom, 1957).

Chance was the next learning expert to research critical thinking skills. He wrote a book on the importance of teaching critical thinking. Each chapter discusses different ways to teach students to think in schools. Assumptions, objectives, methods and materials, target audiences, teacher expertise, advantages, and specific issues are key questions considered during the discussion (Zakaria et al., 2021). John Dewey (1859-1952) was the first thinker (in Sihotang, 2018). According to Dewey, critical thinking is the active and careful consideration of beliefs or types of knowledge that are taken for granted. Beliefs or types of knowledge are learned by seeking reasons that support conclusions. This article was taken from the iPusnas application with the title "Critical Thinking: Life Skills in the Digital Era" on January 24, 2025, at 11:20 PM WIB. Author: Kasdin Sihotang. Publisher: Kanisius PT. Year of Publication: 2018. Pages: 37. (Sihotang, 2018).

From the description above, the characteristics of critical thinking include problem-solving, different approaches to teaching in schools, and critical thinking as an active and careful consideration of a belief or form of accepted knowledge by seeking supporting reasons.

3. Critical Thinking Indicators

The indicators of critical thinking according to experts can be described in the following table:

Table of Critical Thinking Achievement Indicators

No.	Expert	Achievement
1.	Mal Leicester and Denise Taylor from Brookfield in Mardianingsih & Nuris, 2022,	Critical thinking is reviewed through six indicator items: (1) finding and showing differences in objects, (2) arranging patterns, (3) grouping objects by category, (4) restoring an arrangement after being separated, (5) placing objects according to rank, and (6) making decisions.
2.	Hanscomb, 2017 in Yulianti & Susiana, 2023	the process of identifying reasons and determining whether the reasons support the conclusion drawn.
3.	According to the National Council for Excellence in	There is an active and skilled process of knowledge discipline in conceptualizing to analyze, synthesize,

	Critical Thinking (Yuliati & Susianna, 2023)	and evaluate various information gathered through observation, experience, reflection, reasoning, or communication, which is used as a guide for beliefs and actions.
4.	According to Anderson and Krathwohl (Fatra & Maryati, 2018)	Critical thinking skills according to Anderson and Krathwohl include the ability to analyze, evaluate, and create.
5.	According to Ennis (Nufus & Kusaeri, 2020),	(1) understand the problem, (2) provide reasons based on relevant evidence or facts, (3) draw correct conclusions, (4) find answers according to the context of the problem, (5) provide explanations for the conclusions drawn and provide explanations if there are any terms in answering questions, and (6) recheck the answers.
6.	According to John Dewey's Theory In critical thinking	Critical thinking is an important component of critical thinking. This disposition can include curiosity, open-mindedness, self-awareness, empathy, and persistence.

Relevant Research Results

Relevant research is research that has been conducted that is nearly relevant or related to the research to be conducted. This research is useful to avoid repetition of the same research. (Tiaswani, 2020)

The purpose of relevant research is to identify similarities and differences between the research to be conducted. The search for relevant research will focus on development actions using educational game learning media. Many types of educational games are used in research, many of which have been used in previous studies, including:

1. Research by Saputri, N. E., Novianti, R., & Febrialismanto, F. (2022). Development of Educational Prayer Puzzle Media to Improve Prayer Skills in Children Aged 5-6 Years. Journal of Education Research, 2(1), 26–36. https://doi.org/10.37985/jer.v2i1.43 (Saputri et al., 2022) Similarities:

Research and Development (R&D) research, children aged 5-6 years. Differences: Data collection technique using a questionnaire.

2. Research by Amalia, S., & Patiung, D. "Development of Puzzle Media to Develop Early Childhood Children's Latin Letter Recognition Skills." NANAEKE: Indonesian Journal of Early Childhood Education, 4(1), 2021 (Amalia & Patiung, 2021) Similarities: Research and Development/R&D, Latin letter puzzle media. Differences: Improving children's letter recognition skills.
3. Made Intan Asri Devi, N. (2020). Development of Number Puzzle Learning Media to Improve Number Symbol Recognition Skills. Scientific Journal of Teacher Professional Education, 3(3), 417–428. http://dx.doi.org/10.23887/jippg.v3i3 (Made Intan Asri Devi, 2020). Similarities: puzzle learning media, This development research uses the ADDIE model. Differences: improving the ability to recognize number symbols.
4. Surentu, O. S., Sumilat, J. M., & Tarusu, D. T. (2023). Development of Puzzle Media to Increase Interest in Learning Civics in Grade III Elementary School Students. Jurnal Review Pendidikan Dan Pengajar (JRPP), 6(4), 959–963. https://journal.universitaspahlawan.ac.id/index.php/jrpp/article/view/20614/14844 (Surentu et al., 2023). Similarities: development of learning media in the form of Puzzles, Borg and Gall development model. Differences: Improving Civics Learning Scores for Third Grade Elementary School Students
5. Defi, R. A., Abidin, Z., & Susilaningsih, S. (2021). Development of Puzzle Game Media on Force Material for Fourth Grade Elementary School. JKTP: Journal of Educational Technology Studies, 4(4), 329–338. https://doi.org/10.17977/um038v4i42021p329 (Defi et al., 2021) Similarities: Puzzle Educational Media Differences: Development of the Lee & Owens method, Fourth Grade Science Learning, Theme 8.

METHOD

In this study, the researcher chose to use the Research and Development method. This is related to the research objectives, which include describing the concept of an interactive digital puzzle educational game development model. Research and Development (R&D) is a method used to produce specific products and test their effectiveness.

Research and Development (R&D) is a term commonly used to describe activities undertaken by companies and other entities, such as individual entrepreneurs, to create new or improved products and processes. The term's

broadest meaning encompasses activities ranging from basic scientific research conducted in universities and laboratories to testing and refining products before they are sold or used commercially. The performance, intensity, and contribution of R&D are topics widely studied in management, economics, and other social science disciplines. Total expenditure on R&D activities is also one of the most widely used indicators of the innovative performance of companies, industries, and countries (Haryati, 2012).

The ADDIE research model, developed by Dick and Carry in 1996, was chosen by researchers to develop a digital puzzle educational game focused on improving critical thinking skills in early childhood aged 5-6 years. The procedural stages are analysis, design, development, implementation, and evaluation.

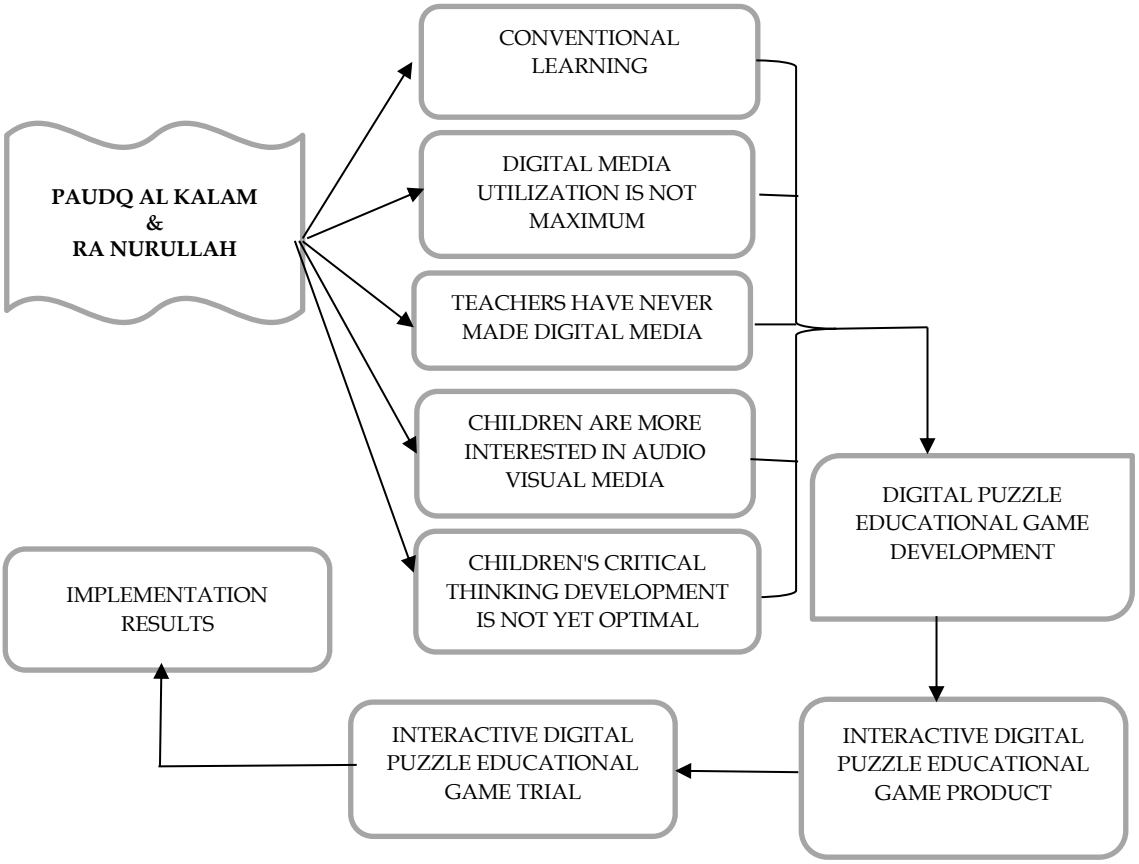
A. Framework/Conceptual Framework

A framework, also known as a conceptual framework, is used when conducting research to understand the relationships between facts, observations, and literature. It serves as a basis for their thinking. Researchers use a framework to analyze, plan, and argue, and develop assumptions that will be discussed. Quantitative research ultimately determines whether the research hypothesis is accepted or rejected, while research based on statements or narratives begins with data and uses theory to explain it, ending with an updated statement or hypothesis (Syahputri et al., 2023).

Widayat and Amirullah (2002) state that a framework, also known as a conceptual framework, is a model of how theory relates to various factors identified as the main problem. Furthermore, a framework provides a tentative explanation of the phenomena being researched. A framework useful for developing hypotheses is built on a foundation of thought patterns based on previous theories and empirical experience. Thus, a framework is the basis for developing hypotheses.

The framework of thought is based on the background of the problem and the learning conditions that exist at the PAUDQ Al Kalam and RA Nurullah Institutions. To make it clearer, the framework of thought will be made with the following diagram.

Conceptual Framework Table



B. Thinking Framework

1. Needs Analysis

The needs analysis will be implemented based on the background of the problem: conventional learning is still being implemented, the lack of digital media utilization, and teachers have never created their own digital learning media. This motivated the researcher to develop an Interacty.me-based digital puzzle educational game. In product development, the researcher used standard indicators from experts, including:

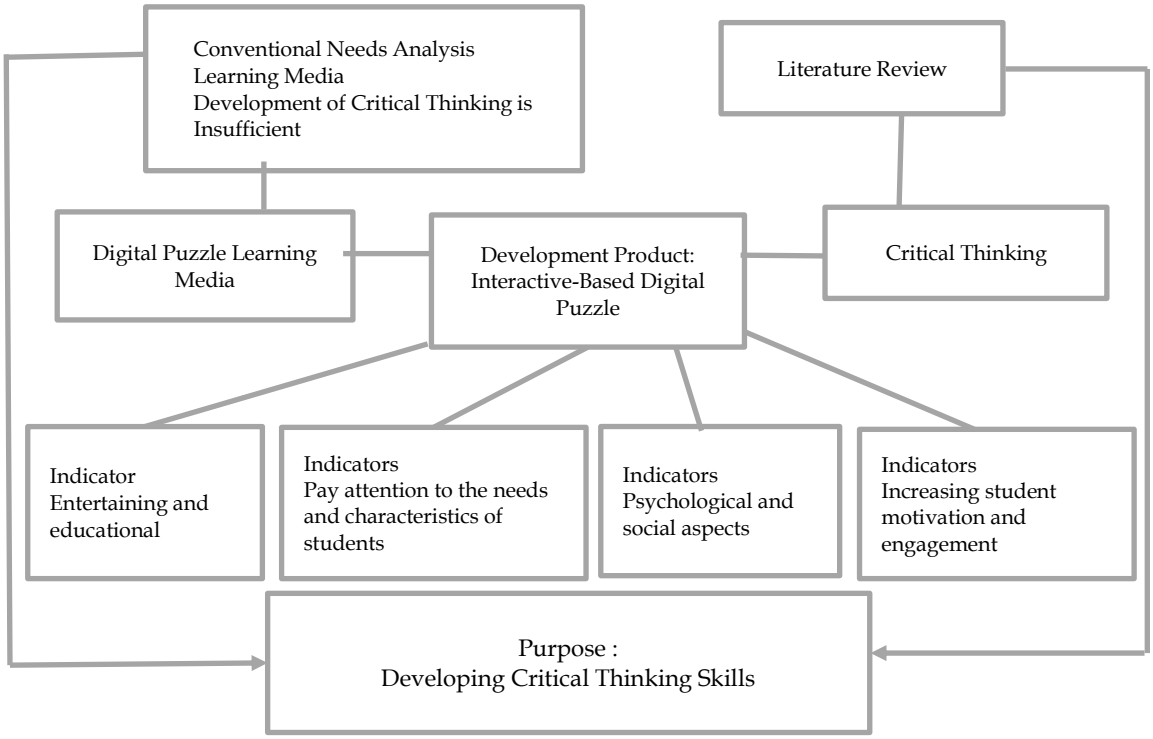
a. Digital puzzles are games that not only develop critical thinking skills in children aged 5-6 years but are also engaging and fun.
b. The educational game design is designed to make children more engaged and engaged.
c. The educational game design is designed to be as safe and flexible as possible for children aged 5-6 years.

2. Literature Review

The literature review will focus on cognitive theory and the critical thinking skills of children aged 5-6 years. The research will focus on indicators according to Mal Leicester and Denise Taylor, namely: finding or pointing out differences,

arranging patterns, grouping objects, restoring order after separation, placing objects according to rank, and making decisions. To achieve critical thinking skills in children aged 5-6 years with these indicators, researchers developed an Interacty-based educational game that can improve critical thinking skills. A framework for thinking or conceptual framework is the basis for research synthesized from facts, observations, and literature reviews.

Thinking Framework Table



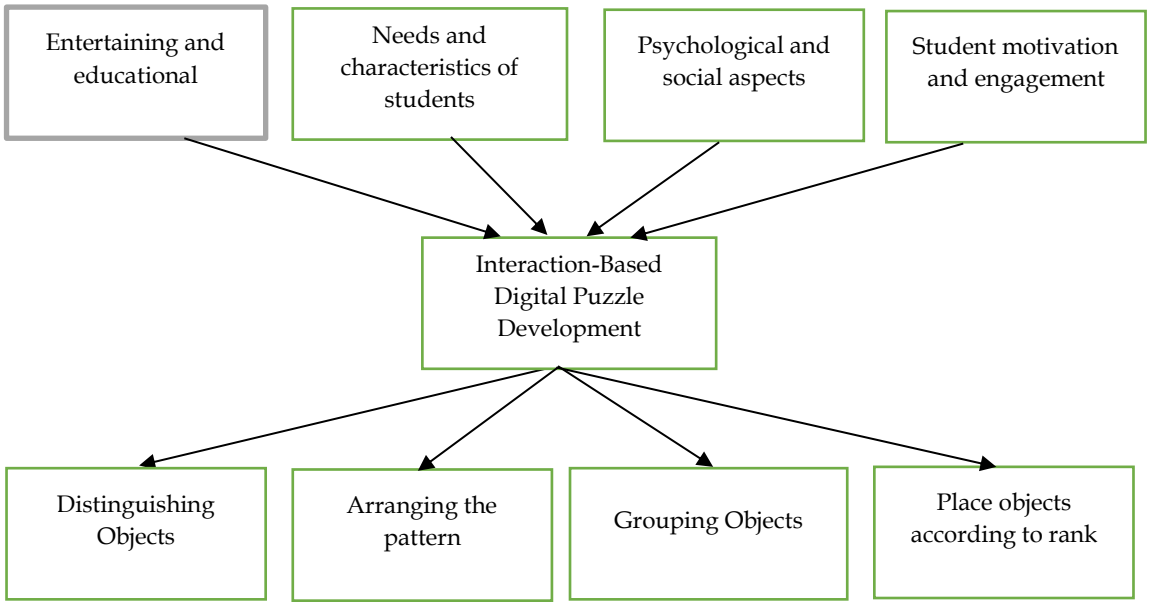
3. Physical Framework

Designing an interactive game is no easy task. Teachers must adjust the level of difficulty, material, objectives, desired values, and the desired game design. This will undoubtedly take up a significant amount of teachers' time. Therefore, a website called interacty.me was created to assist teachers in designing interactive games.

Interacty.me is a website that provides interactive content creation services. Teachers, as designers, have a variety of game templates that can be customized to suit their needs and adapt them to the material being taught. Interacty.me offers simple game creation proven to attract students' attention and foster a competitive spirit. Examples of available game templates that can be edited by teachers include crosswords, Treasure Hunt, Coin Flip, and many other simple games. (San, et al., 2024)

So, based on the explanation above, a puzzle is a game that requires hand-eye coordination. It consists of pieces of a picture that can train creativity, order, and concentration, while also enhancing children's intellectual abilities. One site that helps teachers create engaging games with numerous examples is interacty.me.

Physical Framework Table

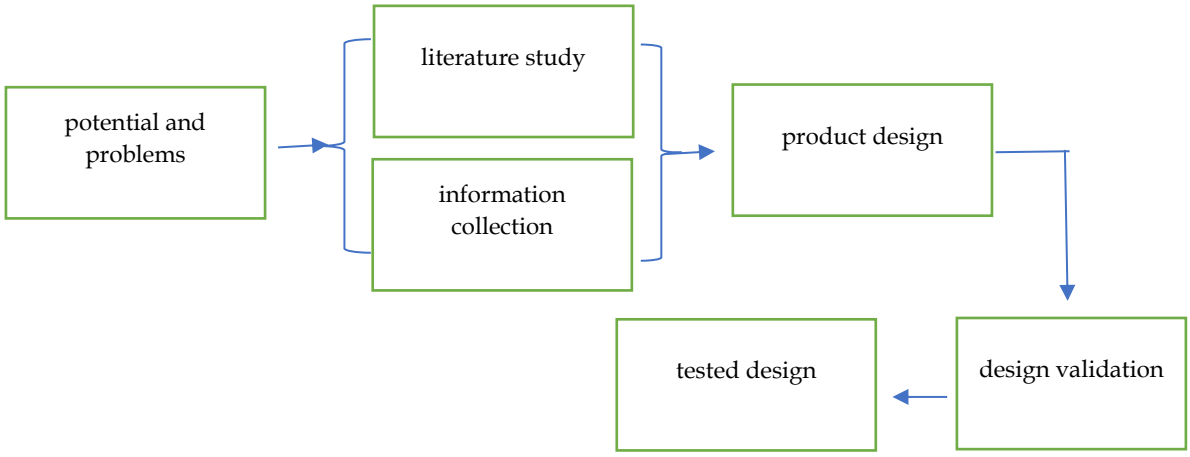


4. Preliminary Study

Based on the framework above, the preliminary study is the initial or preparatory stage for developing and gathering information. This stage involves two main activities: literature review and information gathering regarding the learning process in the field. During this stage, the researcher interviewed three teachers—two from PAUDQ Al Kalam and one from RA Nurullah—regarding the use of learning media to improve critical thinking skills, focusing on the theme of understanding environmental awareness among students. This stage also included reviewing literature and concepts relevant to the research problem.

Based on the interviews, which revealed that the learning conditions at the institution still rely heavily on lectures and less innovative media, often relying on worksheets (LKA), the researcher then drew the conclusion based on the field study that students are still receiving insufficient stimulation to develop their critical thinking skills due to a lack of learning challenges. The development plan will be based on the results of the field study. The conclusions from the analysis of the interviews and observations will lead to the development of an Interacty-based digital puzzle educational game to stimulate critical thinking skills in group B.

Procedural Framework Table



5. Model Development Stage

The ADDIE Development Research Model adopted is based on the book "The ADDIE Learning Model for PEDATI Integration at SMK PGR Karisma Bangsa as a Substitute for Field Work Practices during the Covid-19 Pandemic," written by Taufik Rusmayana, 2021 (Rusmayana, Taufik, 2021).

a. Analysis

In the ADDIE development research model, the first stage is analyzing the need for new product development (models, methods, media, teaching materials) and analyzing the feasibility and requirements for product development. Product development can be initiated by problems with existing/implemented products. Problems can arise and occur because the current or available product is no longer relevant to the target audience's needs, the learning environment, technology, student characteristics, and so on.

After analyzing the problem and the need for new product development, we also need to analyze the feasibility and requirements for product development. The analysis process can be conducted by asking several questions, for example: (1) Is the new product capable of addressing the learning problems encountered? (2) Does the new product have the necessary facilities for implementation? (3) Are lecturers or teachers capable of implementing the new product? Analysis of new products is necessary to determine their feasibility if implemented.

b. Design

The design activity in the ADDIE research and development model is a systematic process that begins with designing the concept and content of the product. A design is written for each product content. Instructions for

implementing the design or creating the product are written clearly and in detail. At this stage, the product design remains conceptual and will form the basis for the development process in subsequent stages.

c. Development

Development in the ADDIE research and development model involves the implementation of the previously created product design. In the previous stage, a conceptual framework for implementing the new product was developed. This conceptual framework is then transformed into a product ready for implementation. Instruments to measure product performance are also needed at this stage.

d. Implementation

The product implementation in the ADDIE research and development model is intended to obtain feedback on the product being created/developed. Initial feedback (the beginning of the evaluation) can be obtained by asking questions related to the product development objectives. Implementation is carried out based on the product design that has been created.

e. Evaluation

The evaluation stage in the ADDIE research and development model is conducted to provide feedback to product users, so that revisions can be made based on the evaluation results or needs that the product has not yet met. The ultimate goal of evaluation is to measure the achievement of development objectives.

ADDIE Model Process Flow Chart

(in Rusmayana Taufik, 2021)

1) Development Model (Product Design)

The next step was the preparation and creation of images and puzzle designs, as well as a learning model that included objectives, learning steps,

equipment used, and the teacher's role in learning. The researcher then consulted with the supervising lecturer before validation by a media expert.

2) Design Validation

After consulting with the supervising lecturer, the next step was validation by a media expert and an instrument for assessing children's critical thinking skills. This validation was considered crucial to ensure that the developed media was of high quality and worthy of trial. This media validation was conducted using expert judgment techniques, where experts, consisting of media experts and teachers, provided their views on the feasibility of the developed model through the distributed instruments.

At this stage, the model design was validated by digital media experts to improve cognitive comprehension skills in children aged 5-6 years, which is at level three (application/analysis) according to Bloom's theory. The results of the expert validation were then reviewed to refine the model design before trial. Before using an instrument in research, analysis is necessary to obtain a valid instrument.

3) Design Revision

Design revisions are conducted based on the results of the initial trial. This involves revising any deficiencies encountered during the trial. These deficiencies are then corrected and refined into a product for further trial.

4) Product Trial

Product trials are a crucial part of development research and are conducted after the product design is finalized. The purpose of a product trial is to determine whether the product is suitable for use. Furthermore, a product trial is used to assess the extent to which the product achieves its goals and objectives.

a) Trial Design

The product to be trialed with students is a digital educational game, a puzzle game, using an interactive application called INTERACTY.ME, using a laptop. The puzzle theme will be environmental and orderly, with themes such as neatness and others. The puzzles are designed in three levels of difficulty, with varying numbers of pieces: the easiest with 4 pieces, the medium with 9 pieces, and the most difficult with 16 pieces. In the initial trial phase, the simplest puzzle, a 4-piece puzzle with a slide-and-click technique, was used, with the theme of environmental love. Once the children have mastered the 4-piece puzzle, a more complex 9-piece version will be created. It is hoped that students will play the game with joy and enthusiasm. Link: https://interacty.me/projects/147fd4eb4ddabf32,
https://interacty.me/projects/990a4c0611021275

b) Trial Subjects

The sources of this research were students and teachers at the Early Childhood Education (PAUD)/TK B level at the PAUDQ AL KALAM Kedawung and RA NURALLAH Palimanan institutions. The reason for selecting these two institutions as subjects was the similarity in the problems faced at both institutions: the lack of use of digital media in learning. PAUDQ AL KALAM already has electronic devices for learning using audio, in the form of loudspeakers and electronic devices; most teachers have computers. The institutions also have projectors that can be used in learning. However, these electronic devices, such as audio devices, are used during group exercise sessions. Laptops/computers are used for administration and are rarely used for learning. Learning is more often conducted using printed textbooks.

c) Data Type

The data collected consisted of interviews with three teachers at PAUDQ Al Kalam (two teachers) and one at RA Nurullah (one teacher). (Interview texts are attached) and direct observations of media use at both institutions.

d) Data Collection Instruments

One of the objectives of this study was to test the effectiveness of an interactive digital puzzle educational game in stimulating children's critical thinking. Therefore, the appropriate method used in this study was Research and Development, as the products of research using the Research and Development method are not only learning materials such as books, audio, and audio-visual materials but can also include learning media and teaching methods that can improve student achievement. Several instruments were used for data collection, such as interviews, questionnaires, and questionnaires.

(1) Interviews

A structured, planned interview is a form of interview in which the interviewer, in this case the researcher, prepares a detailed and systematic plan or guideline of questions according to a specific pattern using a standard format (Yusuf Muri, 2019).

(2) Documentation

This technique is used during media trials conducted directly with students and is documented.

e) Data Analysis Techniques

(1) Descriptive Statistical Analysis

Descriptive statistical analysis is a statistical technique used to analyze data by describing or depicting the collected data as it is without

the intention of drawing general conclusions or generalizations (Muhson, 2006). Descriptive statistical analysis describes research results in detail from before and after treatment.

(a) Using the following formula: According to Casta (2021), the data obtained from the research results are entered into a table and percentages are calculated to be converted into a data conversion table. The formula used to determine this percentage can be shown as follows: P = f / N X100%

Note:

P = Percentage

F = Frequency of observed data

N = Number of data Data processing results

(b) Based on the formula and data table above, the data is converted using the following conversion table: (Quoted from the PPM Correlation Coefficient table) (Casta, H 2021):

Tabel Tabel Konversi

Presentase	Criteria
80%-100%	Interpretation: Very High
60%-79%	High
40%-59%	Fairly High
20%-39%	Less High
<20%	Very Low High

(2) Statistical Analysis Prerequisites

(a) Distribution Normality Test

Data normality tests are used to ensure that data meet the requirements or assumptions for parametric tests such as regression, ANOVA, MANOVA, independent t-tests, paired t-tests, and univariate and multivariate tests. This test aims to determine whether the data distribution follows normal distribution. Data normality testing will be performed using IBM SPSS version 27.

(b) Data Homogeneity Test

Data homogeneity tests are commonly used in comparative analysis, for both ANOVA and sample t-tests. However, it is not an absolute requirement. The purpose of this test is to determine whether there is equal variance in the data variations originating from various variables. Using Levene's statistical criteria, a sigma value below 0.05 is considered non-homogeneous because the data variances in several populations have different variances. A sigma value above 0.05 is

considered homogeneous because the data variances in several populations have equal variances.

(3) Inferential Statistical Analysis

To assess the significance of the differences in critical thinking skills before and after using the Interactive-Based Digital Puzzle media, inferential statistical analysis was used. Two methods were used for this analysis: the paired t-test and the hypothesis test using the t-test. Data analysis was performed using the SPSS version 27 program.

(a) Paired Sample T-test: The paired sample t-test determines whether there is a significant difference between two related mean values (for example, the mean values before and after treatment). If the data used are normally distributed, this test cannot be performed. If the data does not meet the assumption of normality, the Wilcoxon test can be used instead.

(b) Gain Test (g) Next, the analysis continued with a gain test to determine whether there was an increase in the initial critical thinking skills of children aged 5-6 years before and after using the Interactive-Based Digital Puzzle media.

The following is the formula used:

G = $skor\ tes\ akhir - skor\ tes\ awal$

s$kor\ maksimal - skor\ tes\ awal$

After obtaining the gain value (g), it is then confirmed with the table below.

Gain Classification Table (Casta, H 2021)

No.	Gain Index	Interpretation
1.	g>70	High
2.	0.30<g≤0.70	Medium
3.	g≤0.30	Low

RESULTS AND DISCUSSION

The research results represent an attempt to process the research from the initial problem-finding stage to the final stage of the research. (Fatimah et al., 2020) The educational game created using the Interacty application is a digital slide-and-click puzzle featuring animated images of children watering flowers, scribbling on walls, and tidying shoes. To implement the game in the classroom, the researcher used a laptop as the learning medium, and the experiment was conducted in turns, with three children participating with one teacher assessor.

Before playing the game, the children were given information about the day's lesson, including the media and game types. Before playing, the researcher demonstrated how to play the puzzle game in front of the children. After completing one game, the children took turns playing the game. The application was implemented with students at PAUDQ Al Kalam and RA Nurullah. The results of the initial media development are as follows. Digital Puzzle Link: https://interacty.me/projects/147fd4eb4ddabf32, https://interacty.me/projects/990a4c0611021275

Statistical tests in this study were conducted using SPSS 17, which included descriptive statistical analysis, distribution normality tests, and paired sample tests. These tests were conducted on a small scale in the form of a pretest and posttest, and on a large scale also in the form of a pretest and posttest.

A. Small-Scale Test Results

1. Descriptive Statistical Analysis

Tabel Statistics

		Pretest	Posttest
N	Valid	21	21
	Missing	0	0
Mean		4.7619	6.9524
Median		4.0000	7.0000
Mode		4.00	7.00
Std. Deviation		.99523	1.20317
Variance		.990	1.448
Range		2.00	4.00
Minimum		4.00	4.00
Maximum		6.00	8.00
Sum		100.00	146.00

From the Number of Samples (N) There were 21 participants in the pretest and posttest. There was no missing data (missing = 0), so the analysis can be done with full confidence in the validity of the data. Average (Mean) Average pretest score: 4.76 Average posttest score: 6.95, This indicates an increase in scores after the intervention or learning, which is 2.19 points. Median and Mode, The pretest median is 4, while the posttest is 7. This means the middle score has increased. The mode in the pretest is 4, and in the posttest is 7, indicating that the most frequently occurring score has also increased. Standard Deviation and Variance, The standard deviation increased from 0.995 to 1.203, indicating that the posttest scores were slightly more spread out than the pretest scores. The variance

supports this by increasing from 0.990 to 1.448. Range, Minimum, and Maximum The score range increased from 2 to 4, meaning the distribution of scores became wider after the intervention. The minimum score remained at 4, but the maximum score increased from 6 to 8, indicating significant improvement among participants. The total score for all participants increased from 100 to 146, reflecting an aggregate improvement in overall learning outcomes.

2. Distribution Normality Test

Mean Statistical Table

One-Sample Statistics				
	N	Mean	Std. Deviation	Std. Error Mean
Pretest	21	4.7619	.99523	.21718
Posttest	21	6.9524	1.20317	.26255

Explanation of Each Column

a. N (Number of Samples)

➢ The number of respondents who took both the pretest and posttest was 21.

b. Mean (Average)

➢ Pretest: The average score of participants was 4.76, indicating an initial level of understanding before the intervention.

➢ Posttest: The average increased to 6.95, indicating a significant improvement in learning outcomes after the intervention/learning process.

c. Standard Deviation

➢ Shows how spread out the values are from the average.

- Pretest: 0.995 → Relatively small variation in scores, with participants tending to cluster around 4–5.
- Posttest: 1.203 → Slightly higher variation, indicating some participants scored significantly better than others.

d. Standard Error of the Mean

Describes the accuracy of the mean as an estimate of the population value. Pretest: 0.21718 Posttest: 0.26255

→ The smaller this value, the more accurate the estimate of the mean. The values displayed indicate fairly good accuracy.

Conclusion

There was an increase in the mean score from 4.76 (pretest) to 6.95 (posttest), indicating that participants experienced significant improvements in knowledge

or skills after the treatment. The standard deviation indicates that the posttest data were slightly more dispersed, but still concentrated at high scores.

Mean Test Table

One-Sample Test

Test Value = 0

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Pretest	21.926	20	.000	4.76190	4.3089	5.2149
Posttest	26.480	20	.000	6.95238	6.4047	7.5001

This test examines whether the mean (mean) is significantly different from 0. This means we are testing whether the pretest and posttest scores are statistically significantly different from zero.

t (calculated t value) Pretest: 21.926 Posttest: 26.480 A high t value indicates that the mean results are far from zero, indicating a highly significant difference. df (degrees of freedom) The value is 20 because the sample size is 21 → df = N − 1 df = N − 1 df = N – 1. Sig. (2-tailed) The pretest and posttest both have a value of 0.001, which means p < 0.05. This indicates that the difference in the means is statistically highly significant from zero. Mean Difference (Difference from Zero) Pretest: 4.7619 Posttest: 6.9524 These are only averages because the comparison value is 0.

95% Confidence Interval of the Difference Pretest: between 4.3089 – 5.2149 Posttest: between 6.4047 – 7.5001 This means that, with a 95% confidence level, the pretest and posttest mean values are truly within this range—and are unlikely to be close to zero.

Key Conclusions

Both the pretest and posttest have mean values that are highly significantly greater than zero (p = .001). This test does not compare the pretest and posttest directly, but rather proves that each mean value is significantly different from 0. In general, these results indicate that: Participants had a significant initial score (Pretest ≠ 0) After treatment, there was a very significant increase in scores (Posttest ≠ 0) But to state that the posttest was significantly higher than the pretest.

Effect Table (Intervention)

One-Sample Effect Sizes

		Standardizer[a]	Point Estimate	95% Confidence Interval	
				Lower	Upper
Pretest	Cohen's d	.99523	4.785	3.247	6.311

	Hedges' correction	1.03460	4.603	3.124	6.071
Posttest	Cohen's d	1.20317	5.778	3.946	7.600
	Hedges' correction	1.25077	5.558	3.796	7.311

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation.
Hedges' correction uses the sample standard deviation, plus a correction factor.

Effect size measures the magnitude of the effect or influence of a treatment or intervention, not just whether the difference is statistically significant.
Cohen's d and Hedges' g are commonly used measures of effectiveness.
Cohen's d: Uses the sample standard deviation.
Hedges' correction (Hedges' g): Corrects Cohen's d for small samples, making it more accurate for small N values.
Interpretation of Results
✓ Pretest Cohen's d = 4.785 → This is a very large effect size. Hedges' g = 4.603 → A small correction due to the sample size (n = 21).
Confidence Interval (CI) for Cohen's d: 3.247 – 6.311
→ This means we are very confident that the mean pretest score is significantly greater than zero, and the effect is large.
✓ Posttest Cohen's d = 5.778 and Hedges' g = 5.558
→ The learning effect is even larger than the pretest.
The CI is also high: between 3.946 and 7.600 (Cohen's d), indicating a very large effect size and certainly not due to chance.

Table of Effect Size Categories (Cohen's d)

d Value	Interpretation
0.2	Small
0.5	Moderate
0.8	or greater
>2	Very large

Therefore, Cohen's d > 4 and > 5 for the pretest and posttest indicate an exceptionally large effect.

3. Paired Sample T-test

Table Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Pretest	4.7619	21	.99523	.21718
	Posttest	6.9524	21	1.20317	.26255

Column Explanation:

a. Mean
   - Pretest: 4.76 → Average initial score of participants.
   - Posttest: 6.95 → Average score after treatment (intervention or learning).
   - This indicates that, on average, scores increased by approximately 2.19 points after learning.

b. N (Number of Samples)
   - The number of participants who took both tests (pre and post) was 21.

c. Std. Deviation
   - Describes the variability of scores in each test.
   - Pretest: 0.995 → The scores are fairly centralized, meaning participants have fairly uniform results.
   - Posttest: 1.203 → The variation is slightly larger, indicating that results after the intervention are more varied.

d. Std. Error of the Mean
   - Shows how accurate the average estimate is.
   - A smaller value indicates a more stable and representative estimate of the mean.

Initial Conclusions from These Statistics

- There was an increase in the average score from 4.76 to 6.95, indicating that participants experienced an increase in their skills or knowledge after the treatment.
- This difference appears quite large descriptively.
- However, to determine whether this difference is statistically significant, you must look at the output of the Paired Samples Test (paired t-test), which is the next section of this analysis.

Table Paired Samples Correlations

			N	Correlation	Sig.
Pair 1	Pretest & Posttest		21	.700	.000

Column Explanation:

a. N (Number of Samples):
   - A total of 21 participants took both tests: pretest and posttest.

b. Correlation (Correlation Coefficient):
   - The value is 0.700, which indicates a strong positive correlation between pretest and posttest scores.
   - This means that, in general, participants who scored high on the pretest tended to also score high on the posttest—and vice versa.

c. Sig. (Significance, p-value):
   - The p-value = 0.001, meaning this correlation is statistically significant (because p < 0.05).
   - Therefore, the relationship between pretest and posttest scores is not due to chance.

Interpretation of a Correlation of 0.700

0.700 falls into the following category: 0.60 – 0.79 = strong correlation. This correlation does not imply a causal relationship, but rather indicates that participants' performance on the pretest can predict their performance on the posttest, even though the intervention still has an effect.

Table Paired Samples Test

		Paired Differences							
					95% Confidence Interval of the Difference				
		Mean	Std. Deviation	Std. Error Mean	Lower	Upper	t	df	Sig. (2-tailed)
Pair 1	Pretest - Posttest	-2.19048	.87287	.19048	-2.58780	-1.79315	-11.500	20	.000

Explanation:

a. The Paired Samples Test is a statistical test used to compare the averages of two paired conditions, for example, pretest and posttest scores from the same subject.
b. Mean (Average) Paired Differences = -2.19048. This means that the average posttest score is approximately 2.19 lower than the pretest (because the mean difference is negative).
c. Std. Deviation indicates the variation in differences between pairs, which is approximately 0.87.
d. Std. Error of the Mean indicates the accuracy of the estimated average difference; the smaller the value, the better (here, approximately 0.19).
e. Confidence Interval (95%): Range -2.59 to -1.79. This means we are 95% confident that the true average difference between the pretest and posttest falls between these numbers. Since this interval does not include 0, there is a significant difference.
f. t = -11.500 and df = 20. A large absolute t-value (|t| = 11.5) indicates a significant difference.
g. Sig. (2-tailed) = 0.001 (usually p < 0.05 is considered significant). A very small p-value means there is very strong statistical evidence that there is a significant difference between the pretest and posttest scores.

Conclusion:

There is a statistically significant difference between the pretest and posttest scores. The posttest mean is approximately 2.19 points lower than the pretest, and this difference is not due to chance.

Table Paired Samples Effect Sizes

			Standardizer[a]	Point Estimate	95% Confidence Interval	
					Lower	Upper
Pair 1	Pretest - Posttest	Cohen's d	.87287	-2.510	-3.384	-1.619
		Hedges' correction	.88968	-2.462	-3.320	-1.589

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation of the mean difference.
Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.

Explanation:

a. Effect size is a measure of the magnitude of the difference between two groups, in this context, between the pretest and posttest. This helps us understand how large the effect is in practical terms, not just whether the difference is statistically significant.
b. Cohen's d
   - Cohen's d = 0.87287
   - This means the effect size is approximately 0.87, which, according to Cohen's general rule, can be considered a medium to large effect (0.2 = small, 0.5 = medium, 0.8 = large).
   - Cohen's d is calculated by dividing the mean difference by the standard deviation of that difference.
c. Hedges' correction
   - Value = 0.88968
   - This is a corrected version of Cohen's d that takes into account small sample sizes to provide a more accurate estimate of the effect. Its value is similar to Cohen's d, slightly larger due to the correction.
   - Typically used for small samples (df = 20 here is considered small).
d. Confidence Interval (CI)
   - There is likely a typo in the confidence interval you provided, as the CI value is negative and inconsistent with the positive point estimate (0.87). Typically, CIs for effect sizes like Cohen's d are positive and surround the point estimate.
   - This CI indicates the range of values within which the true effect size is most likely to lie, with a 95% confidence level. If the CI range does not include zero, this indicates that the effect size is significant.

e. Notes on the denominator:
   - Cohen's d uses the standard deviation of the difference in scores as the divisor to calculate the effect size.
   - Hedges' correction uses the same standard deviation, but with a correction factor to correct for biases that occur in small samples.

Conclusion:

- The effect size of the difference between the pretest and posttest is moderate to large (around 0.87 to 0.89), indicating a difference that is practically meaningful, not just statistically.
- Hedges' correction provides a slightly more conservative estimate but is nearly identical to Cohen's d.

Table

Brief Comparison of Pretest and Posttest Results

High Score (%)	Pretest: 38.1% (score 6.00)	Posttest: 80.95% (scores 7.00 & 8.00)
Low Score (%)	Pretest: 61.9% (score 4.00)	Posttest: 19.05% (scores 4.00 & 5.00)

- This demonstrates the effectiveness of the treatment in improving participant learning outcomes.

Figure 4.1 Pretest Histogram

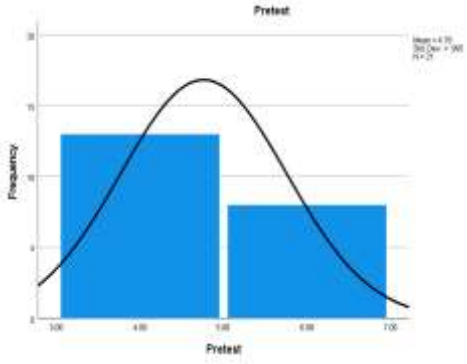
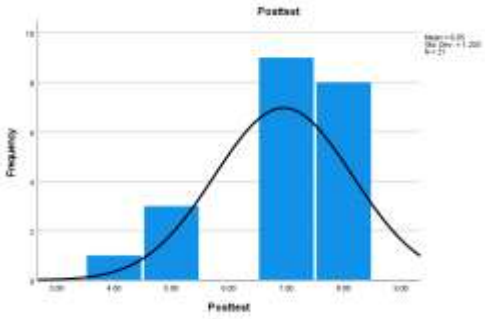


Figure 4.2 Posttest Histogram



The average pretest score was relatively low (4.76), indicating a limited level of initial mastery of the material. Most participants (more than half) scored 4.00. The data distribution was relatively narrow, indicating that participants' scores were quite concentrated (not very varied).

The average posttest score increased significantly compared to the pretest (from 4.76 to 6.95). The distribution of scores shifted to the right, indicating improved learning outcomes. Most participants scored high, indicating a greater understanding of the material after the intervention. Scores remained variable, but the majority clustered at the upper end.

B. Large-Scale Test Results

1. Descriptive Statistical Analysis
   a. Descriptive Statistics Results for Pretest and Posttest

This large-scale study involved 35 respondents from two institutions, PAUDQ Al Kalam and RA Nurullah, who took the pretest and posttest.

The assessment results were analyzed using IBM SPSS 27, which aimed to measure the increase in participants' understanding of the material after receiving the treatment or intervention designed in this study. The statistical table of the pretest and posttest is below:

Tabel 4.15.Statistics

		Pretest	Posttest
N	Valid	35	35
	Missing	0	0
Mean		5.0571	6.8286
Median		4.0000	7.0000
Mode		4.00	8.00
Std. Deviation		1.41302	1.20014
Variance		1.997	1.440
Range		4.00	4.00
Minimum		4.00	4.00
Maximum		8.00	8.00
Sum		177.00	239.00

Based on the results above, the average pretest score was 5.06, while the average posttest score increased to 6.83. This indicates an increase in scores after the intervention.

The median score also increased from 4.00 in the pretest to 7.00 in the posttest, indicating that most participants scored higher after the intervention. The mode, the most frequently occurring score, also increased from 4.00 in the pretest to 8.00 in the posttest.

The standard deviation decreased from 1.41 to 1.20. This indicates that the data distribution in the posttest was more even than in the pretest, meaning posttest results tended to be more consistent.

Although the range of scores, which is the difference between the maximum and minimum scores, remained the same (4.00), the increase in the average score and the shift in the median and mode indicate that overall posttest results were higher. Thus, descriptively, it can be concluded that there was an increase in participants' learning outcomes after the intervention in this study. Further

analysis will be discussed in the inferential statistical tests in the next sub-chapter.

2. Distribution Normality Test

a. One-Sample Statistics: Pretest and Posttest

One-Sample Statistics analysis was used to obtain an overview of the mean, standard deviation, and standard error of the mean from the pretest and posttest data collected from 35 participants.

Mean Test Table

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Pretest	35	5.0571	1.41302	.23884
Posttest	35	6.8286	1.20014	.20286

The average pretest score was 5.06, indicating that participants' initial scores before the intervention were relatively low. The average posttest score increased to 6.83, indicating improved learning outcomes after the intervention.

The standard deviation decreased from 1.41 in the pretest to 1.20 in the posttest, indicating that participants' scores became more uniform after the intervention.

The standard error of the mean indicates the level of confidence in the sample mean. This value also decreased from 0.239 in the pretest to 0.203 in the posttest, indicating that the average posttest results were more stable and representative of the population.

Overall, these data confirm previous findings that there was an improvement in participants' abilities after the intervention. To test the statistical significance of this improvement, an inferential analysis using a paired sample t-test will be conducted, as explained in the next subchapter.

b. One-Sample t-Test

A one-sample t-test analysis is used to test whether the average pretest and posttest scores are significantly different from the reference value, in this case, compared to a value of zero (test value = 0) as the basis for statistical analysis.

Mean Statistics Table

One-Sample Test

	Test Value = 0					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference Lower	Upper
Pretest	21.173	34	.000	5.05714	4.5718	5.5425
Posttest	33.661	34	.000	6.82857	6.4163	7.2408

The pretest results showed a t-value of 21.173 with a significance level of 0.000 ($p < 0.05$), meaning the pretest score was significantly different from zero. This indicates that participants already had a baseline understanding before receiving the treatment, albeit at a relatively low level.

The posttest results also showed a t-value of 33.661 with a significance level of 0.000, meaning the posttest score was also significantly different from zero. This confirms that participants experienced a significant improvement in learning outcomes after the intervention.

The mean difference (mean difference) in the pretest was 5.06, while in the posttest it increased to 6.83. The 95% confidence interval (CI) indicates that we are 95% confident that the average pretest score is between 4.57 and 5.54, and the posttest score is between 6.42 and 7.24.

Conclusion:

Although this test uses a test value of 0 (which generally only indicates that the mean is not zero), the results still confirm that: Both scores (pretest and posttest) are statistically highly significant.

c. Effect Size

In addition to statistical significance, it is also important to assess the extent of the treatment's effect on learning outcomes. This can be seen from the effect size using Cohen's d and Hedges' correction.

Effect Size Test Table Cohen's d

One-Sample Effect Sizes

			Point	95% Confidence Interval	
		Standardizer[a]	Estimate	Lower	Upper
Pretest	Cohen's d	1.41302	3.579	2.665	4.484
	Hedges' correction	1.44518	3.499	2.606	4.384
Posttest	Cohen's d	1.20014	5.690	4.300	7.072
	Hedges' correction	1.22745	5.563	4.204	6.915

a. The denominator used in estimating the effect sizes.

Cohen's d uses the sample standard deviation.

Hedges' correction uses the sample standard deviation, plus a correction factor.

Cohen's d for the pretest was 3.579, increasing to 5.690 for the posttest. Based on Cohen's interpretation:

d = 0.2 → small effect

d = 0.5 → medium effect

d = 0.8 → large effect

Therefore, effect size values above 3 to 5 indicate a very large effect.

Hedges' correction was used to reduce bias in small sample sizes, and the results still showed very large effects for both the pretest (3.499) and posttest (5.563). The 95% confidence interval (CI) was also well above the threshold for a "large effect," indicating that the results are statistically and practically convincing.

Conclusion:

The very large effect size indicates that the treatment/intervention administered in this study had a strong and significant impact on improving participants' learning outcomes. Not only statistically, but also practically, the treatment was proven to be highly effective.

3. Paired Sample T-test
   a. Paired Samples Statistics

Paired Samples Statistics analysis is used to examine descriptive comparisons between two measurements taken from the same subject, namely the pretest and posttest scores of 35 participants.

Table Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Pretest	5.0571	35	1.41302	.23884
	Posttest	6.8286	35	1.20014	.20286

The average pretest score was 5.06, while the average posttest score increased to 6.83. This indicates an improvement in learning outcomes after the intervention.

The standard deviation decreased from 1.413 to 1.200, indicating that the posttest scores were more concentrated and less dispersed.

The standard error of the mean also decreased from 0.239 to 0.203, indicating that the estimated average posttest score was more precise than the pretest score.

Initial conclusion:

Based on these descriptive statistics, it can be seen that participants experienced consistent improvement after the intervention. However, to determine whether this improvement is statistically significant, the results of the Paired Samples Test, which compares the two averages inferentially, are needed.

b. Paired Samples Correlations

Paired Samples Correlations analysis aims to determine the relationship or correlation between pretest and posttest scores from the same subject.

Table Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Pretest & Posttest	35	.682	.000

The correlation value (r) of 0.682 indicates a strong and positive relationship between pretest and posttest scores. This means that participants who scored high on the pretest also tended to score high on the posttest.

A significance value (p) of 0.000 (<0.05) indicates that the correlation is statistically significant, meaning the relationship between the two scores did not occur by chance.

Conclusion:

The significant and fairly strong correlation between the pretest and posttest indicates consistency in scores across individuals, despite an overall increase in scores. This suggests that the intervention not only improved overall scores but also maintained the performance order across participants.

c. Paired Samples t-Test

The Paired Samples t-Test is used to determine whether there is a statistically significant difference between pretest and posttest scores after the treatment/intervention is administered to participants.

Table 4.23 Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Pretest - Posttest	-1.77143	1.05957	.17910	-2.13540	-1.40745	-9.891	34	.000

The average difference between the pretest and posttest was -1.77, meaning there was an average increase in posttest scores of 1.77 points.

The t-value of -9.891 indicates a very large and measurable difference between the two scores.

A significance value (p-value) of 0.000 (<0.05) indicates that the difference between the pretest and posttest is statistically significant.

The 95% confidence interval for the difference in scores ranges from -2.135 to -1.407, which is completely below zero, confirming that the posttest scores were consistently higher than the pretest.

Conclusion:

Based on the results of the Paired Samples t-Test, it can be concluded that there was a significant difference between the pretest and posttest scores. This indicates that the intervention or treatment provided in this study was effective in improving participants' learning outcomes.

d. Paired Samples Effect Sizes

Effect sizes are used to measure the extent of the treatment/intervention's influence on the difference in scores between the pretest and posttest. In this case, Cohen's d and Hedges' correction were used for the Pretest - Posttest pair.

Table 4.24 Paired Samples Effect Sizes

					95% Confidence Interval	
			Standardizer[a]	Point Estimate	Lower	Upper
Pair 1	Pretest - Posttest	Cohen's d	1.05957	-1.672	-2.183	-1.150
		Hedges' correction	1.07144	-1.653	-2.158	-1.138

a. The denominator used in estimating the effect sizes.
Cohen's d uses the sample standard deviation of the mean difference.
Hedges' correction uses the sample standard deviation of the mean difference, plus a correction factor.

Cohen's d = -1.672 indicates a very large effect (since it significantly exceeds 0.8), and the negative sign indicates that the posttest score is higher than the pretest (due to the order of pretest and posttest).

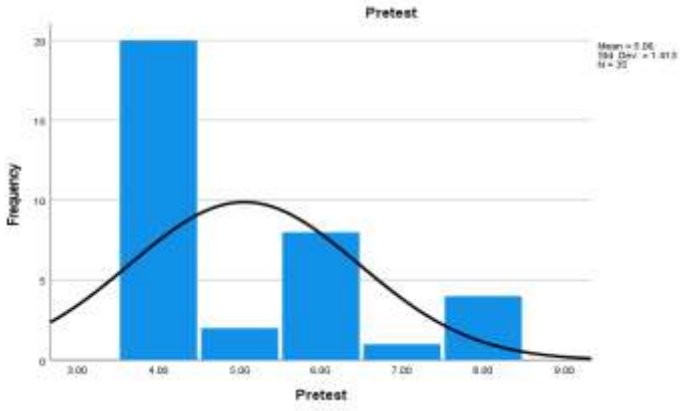
The Hedges' correction = -1.653 also confirms a very large effect, with minor adjustments to address sample size bias.

The 95% confidence intervals for both do not cross zero and are entirely negative, strengthening the evidence for a significant and large effect of the treatment.

e. Visualization of Pretest Score Distribution

The following figure shows a histogram of the distribution of pretest and posttest scores from 35 participants before and after the treatment.

Large-Scale Pretest Histogram Image



The graph above shows that:

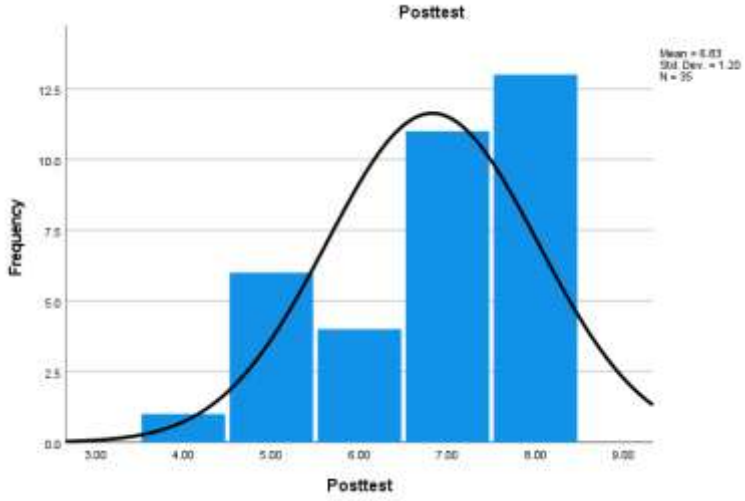
- Score 4 was the most dominant, obtained by 20 participants, indicating the highest frequency.

- Score 6 was obtained by 8 participants, while the other scores (5, 7, and 8) were obtained by fewer participants.
- The distribution appears positively skewed, meaning that most participants scored below the average, in the range of 4 to 6.
- The curved line in the graph represents the normal curve used to compare the data distribution. It appears that the distribution of pretest scores tends to be uneven and does not fully follow a normal distribution pattern.

The average pretest score was 5.06 with a standard deviation of 1.413, indicating that most participants had relatively low initial abilities and were not spread too far from the average.

This visualization reinforces the previous descriptive findings that before treatment, participants still demonstrated a low level of mastery of the material. Therefore, intervention or treatment is needed to improve their understanding, which will then be evaluated through a posttest.

Large-Scale Post-Test Histogram Image



The graph above shows that:

- The score of 8 had the highest frequency, obtained by 13 participants.
- This was followed by the score of 7, obtained by 11 participants.
- Low scores such as 4 and 5 were only obtained by a small proportion of participants.
- The distribution of the graph approximates a normal curve, as seen from the bell curve in the graph, indicating that participants' scores are more evenly distributed and concentrated in the middle to high range.
- The average posttest score was 6.83, with a standard deviation of 1.20, indicating an improvement from the pretest score, both in terms of the average and a more concentrated distribution of scores.
- This distribution indicates that after the treatment was administered, there was a shift in scores toward higher scores, as reflected in the majority of

participants scoring between 7 and 8. This indicates that the intervention provided in the study had a positive impact on improving participants' understanding.

f. Visual Comparison of Pretest and Posttest

Based on the comparison between Figure 4.1 and Figure 4.2, it can be concluded that there was a significant improvement in the distribution of scores after the treatment was administered. In the pretest, most participants were in the low range (score 4), while in the posttest, the distribution shifted to the high range (scores 7 and 8). This indicates the treatment's success in improving participants' learning outcomes.

CONCLUSION

Based on the research and development of an interactive digital puzzle educational game to stimulate critical thinking in children in kindergartens in Cirebon Regency, the following conclusions can be drawn:

1. The Need for Engaging Media
   Preliminary studies indicate that early childhood requires engaging learning media that aligns with their developmental characteristics. Digital-based educational games are an innovative solution to stimulate critical thinking in children from an early age.
2. Development Process Using the ADDIE Model
   The game was developed through the stages of Analysis, Design, Development, Implementation, and Evaluation (ADDIE). During the development stage, revisions were made based on expert input, such as replacing animations with real-life images to increase relevance to everyday life.
3. The Game's Effectiveness in Stimulating Critical Thinking
   Based on pretest and posttest results on 35 students at PAUDQ Al Kalam and RA Nurullah, there was a significant increase in children's critical thinking skills. This is supported by statistical test results, which showed an increase in the average score from 5.06 to 6.83, as well as a Cohen's d effectiveness value of 0.87, which falls into the large effect category.
4. Positive Responses from Children
   Most children demonstrated high enthusiasm for using this media. They felt happy, interested, and more focused when learning using the digital puzzle game. Many students even expressed their hope that similar media would be used in future lessons.

REFERENCES

Amalia, S., & Patiung, D. (2021). Pengembangan Media Puzzle Untuk Menumbuhkan Kemampuan Mengenal Huruf Latin Anak Usia Dini. Nanaeke: Indonesian Journal of Early Childhood Education, 4(1), 53. https://doi.org/10.24252/nananeke.v4i1.21598

Ayu bunga Dibtasari et al., 2023. (2023). Buku panduan membuat instrumen pembelajaran dengan puzzle maker (I). ipusnas2.perpusnas.go.id. https://ipusnas2.perpusnas.go.id

Defi, R. A., Abidin, Z., & Susilaningsih, S. (2021). Pengembangan Media Puzzle Game Materi Gaya Untuk Kelas IV Sekolah Dasar. JKTP: Jurnal Kajian Teknologi Pendidikan, 4(4), 329–338. https://doi.org/10.17977/um038v4i42021p329

Fatra, M., & Maryati, T. K. (2018). Implementasi k13 pada pembelajaran matematika dalam meningkatkan kemampuan berpikir kritis dan kreatif. 1, 1–28. http://repository.uinjkt.ac.id/dspace/handle/123456789/42272

Hasanah, U., & Fajri, N. (2022). Konsep Pendidikan Karakter Anak Usia Dini. Edukids : Jurnal Inovasi Pendidikan Anak Usia Dini, 2(2), 116–126. https://doi.org/10.51878/edukids.v2i2.1775

Hendrawan, G. B., & Marlina, R. (2022). Persepsi Siswa Terhadap Penggunaan Game Edukasi Digital Pada Pembelajaran Matematika. JPMI (Jurnal Pembelajaran Matematika Inovatif), 5(2), 395. https://doi.org/10.22460/jpmi.v5i2.10288

Karseno, Sariyasa, & I.G. Astawan. (2021). Pengembangan Media Game Edukasi Berbasis Android Pada Topik Bilangan Bulat Kelas VI Sekolah Dasar. Jurnal Teknologi Pembelajaran Indonesia, 11(1), 16–25. https://doi.org/10.23887/jurnal_tp.v11i1.621

Kozachenko, E. E., & Davydova, E. M. (2019). A “labyrinth” toy for 6-7 years old children as a way of intellectual development of a child. CEUR Workshop Proceedings, 2485, 202–206. https://doi.org/10.30987/graphicon-2019-2-202-206

Mardianingsih, & Nuris, A. D. (2022). Karakteristik Perkembangan Anak Usia 5-6 Tahun Dalam Aspek Kognitif Melalui Permainan Balok. Jurnal Caksana : Pendidikan Anak Usia Dini, 5(1), 29–36. http://trilogi.ac.id/journal/ks/index.php/PAUD/article/view/1332

Nufus, H., & Kusaeri, A. (2020). Analisis Tingkat Kemampuan Berpikir Kritis Siswa Dalam Memecahkan Masalah Geometri. Jurnal Pendidikan Matematika Indonesia, 5(2), 49–55.

Nuryanti, N., Amirul, B., Preset, C., Detected, R., Check, I., Check, I., Important, F., & Engine, U. A. (2022). Analyzed document : Nila Nuryanti fiks . docx Licensed to : Bagus Amirul Comparison Preset : Rewrite Detected language : Id Check type : Detailed document body analysis : Relation chart : Distribution graph : Processed resources details : Important notes. 4–8.

Pradana, A. G., & Nita, S. (2019). Rancang bangun game edukasi "amudra" alat musik daerah berbasis android. Prosiding Seminar Nasional Teknologi Informasi Dan Komunikasi (SENATIK), 2(1), 49–53.

Purnomo, I. I. (2020). Aplikasi Game Edukasi Lingkungan Agen P Vs Sampah Berbasis Android Menggunakan Construct 2. Technologia: Jurnal Ilmiah, 11(2), 86. https://doi.org/10.31602/tji.v11i2.2784

Safitri, M. Sarman, F; Rosadi, A; Hidayah, N; Rizki Nur Khalifah, D; Dailami; Dewi, M; Setiawan, E; Robert Tuerah, P; Atikah, C. (2023). Pengembangan Media Pembelajaran (P. Sarwandi(Mifandi Mandiri Digital (ed.); June 2023). Mifandi Mandiri Digital.

Saidah, N. (2020). Konsep Dasar, Prinsip dan Peranan Supervisi Pendidikan. Jurnal Pendidikan, 1(2), 21.

Salsabila Samsyah, N. K., & Dheasari, A. E. (2022). Meningkatkan Kemampuan Kognitif Anak Usia 5-6 Tahun Melalui Media Tebak Warna Gambar Di Tk It Abatatsa Pasuruan. Incrementapedia: Jurnal Pendidikan Anak Usia Dini, 4(2), 47–55. https://doi.org/10.36456/incrementapedia.vol4.no2.a6544

San, F, Yudha, Mardiana, P. Safitri, S. Ardiansyah, T. (2024). Efektivitas pembelajaran sejarah berbasis kuis interaktif. 09 (September).

Saputri, N. E., Novianti, R., & Febrialismanto, F. (2022). Pengembangan Media Puzzle Shalat Edukatif dalam Meningkatkan Kemampuan Ibadah Shalat Anak Usia 5-6 Tahun. Journal of Education Research, 2(1), 26–36. https://doi.org/10.37985/jer.v2i1.43

Shofia, M., & Dadan, S. (2021). Media Pembelajaran untuk Anak Usia Dini di Pendidikan Anak Usia Dini. Jurnal Pendidikan Tambusai, 05(01), 1560–1561.

Sihotang, K. (2018). Berpikir Kritis: Kecakapan Hidup di Era Digital.

Sumani, Manan, A., & Supriadi, S. (2022). Pengembangan Media Puzzle Berbasis Make A Match Tema 8 Subtema 2 Keunikan DaerahTempat Tinggalku. Jurnal Al-Hikmah Way Kanan, 1(2), 79–88.

Surentu, O. S., Sumilat, J. M., & Tarusu, D. T. (2023). Pengembangan Media Puzzle Untuk Meningkatkan Minat Belajar Pkn Siswa Kelas Iii Sekolah Dasar. Jurnal Review Pendidikan Dan Pengajaran (JRPP), 6(4), 959–963.

https://journal.universitaspahlawan.ac.id/index.php/jrpp/article/view/20614/14844

Sutarman, A. B., Rahmanto, Y., & Puspaningrum, A. S. (2022). Pembuatan Game Edukasi Pembelajaran Kata Imbuhan Untuk Tingkat Sekolah Dasar (Studi Kasus Sd Negeri Karang Sari Lampung Utara). Jurnal Informatika Dan Rekayasa Perangkat Lunak, 3(2), 202–212. https://doi.org/10.33365/jatika.v3i2.2027

Sutisna, I., & Laiya, S. W. (2020). Metode Pengembangan Kognitif Anak Usia Dini. In UNG Press Gorontalo.

Tiaswani, D. C. (2020). Penelitian Relevan. Angewandte Chemie International Edition, 6(11), 951–952., 5–24.

Uswatun, U., Suryani, L., Liza, M. El, & Saputra, N. I. (2022). Analisis Deskriptif Penerapan Model Pembelajaran Bcct Dalam Mengembangkan Kemampuan Berpikir Kritis Pada Anak Usia Dini. Atthufulah : Jurnal Pendidikan Anak Usia Dini, 3(1), 26–36. https://doi.org/10.35316/atthufulah.v3i1.2425

Widyanarti, S. (2013). Penggunaan Media Puzzle Dalam Model Pembelajaran Langsung Untuk Meningkatkan Hasil Belajar Siswa Pada Mata Pelajaran Ips Kelas Va Sdn Rangkah I Tambaksari Surabaya. Jurnal Penelitian Pendidikan Guru Sekolah Dasar, 1(1), 1–5. http://ejournal.unesa.ac.id/index.php/jurnal-penelitian-pgsd/article/view/1007

Yuliati, C. L., & Susianna, N. (2023). Penerapan Model Pembelajaran Discovery Learning Dalam Meningkatkan Keterampilan Proses Sains, Berpikir Kritis, dan Percaya Diri Siswa. Scholaria: Jurnal Pendidikan Dan Kebudayaan, 13(1), 48–58. https://doi.org/10.24246/j.js.2023.v13.i1.p48-58

Yulinda, M. F. F., & Saifuddin, M. F. (2022). Digital puzzle: Alternative media for cell learning in middle school. BIO-INOVED : Jurnal Biologi-Inovasi Pendidikan, 4(2), 230. https://doi.org/10.20527/bino.v4i2.12600

Yus, A., Diputera, A. M., Agustiara, B., Sianipar, E., Boangmanalu, R. A., Naibaho, B. A., ... & Purba, C. A. (2023). Implementasi Instrumen Penilaian Perkembangan Kognitif Anak Usia 5-6 Tahun di TK Santo Thomas 2 Medan. Jurnal Pendidikan dan Keguruan, 1(4), 1238-1247.

Zakaria, I., Suyono, S., & Priyatni, E. T. (2021). Dimensi Berpikir Kritis. Jurnal Pendidikan: Teori, Penelitian, Dan Pengembangan, 6(10), 1630. https://doi.org/10.17977/jptpp.v6i10.15072