JURNAL TEKNOLOGI DAN OPEN SOURCE
Vol. 8, No. 2, December 2025, pp. 993~1006
e-ISSN: 2622-1659, accredited Four Grade by Kemenristekdikti, Decree No: 152/E/KPT/2023
DOI: https://doi.org/10.36378/jtos.v8i2.5108

Designing History Based Educational Games to Improve Cognitive Empathy in Children with ASD: A Pilot Study

Cai Gi[1], Nanang Hoesen Hidroes Abbrori[2]
[12] Department of Digital Business, Faculty of Economics and Business, Universitas Negeri Surabaya, Indonesia

Article Info

Article history:

Received 11 11, 2025
Revised 11 30, 2025
Accepted 12 17, 2025

Keywords:

Autism Spectrum Disorder
Cognitive Empathy
Game-Based Learning
Historical Narrative
Theory of Mind

ABSTRACT

Children with Autism Spectrum Disorder (ASD) face significant challenges in Theory of Mind (ToM), particularly in understanding others' perspectives. While serious games have shown potential, most existing interventions focus on basic emotion recognition without engaging narrative contexts. Objective: This pilot study aims to design a history-based serious game, " Petualangan Tarumanegara," and evaluate its preliminary effectiveness in improving cognitive empathy among children with high-functioning ASD. Using a mixed-methods case study approach with a one-group pretest-post-test design, six students with ASD (aged 8-15) participated in a 2-week intervention. The game integrates historical narratives as a logical scaffold with explicit emotion scaffolding and moral decision-making mechanics. Data were analyzed using the Wilcoxon Signed-Rank Test and triangulated with behavioral observations. Quantitative results showed a significant increase in cognitive empathy scores ($p = 0.0312$) with a large effect size (Cohen's $d = 3.19$). However, triangulation revealed a "knowing-doing gap," in which higher cognitive scores did not correlate significantly with spontaneous prosocial behavior ($\rho = -0.36$). The history-based narrative approach effectively enhances cognitive knowledge of empathy by providing a structured context. However, bridging the gap between cognitive understanding and spontaneous behavioral application remains a critical challenge. To address this, future designs should explicitly integrate behavioral prompting systems that scaffold the transition from cognitive rules to spontaneous social action, aligning with the systemizing strengths of children with ASD.

Corresponding Author:

Cai Gi
Digital Business
Universitas Negeri Surabaya
Surabaya, Indonesia
Email: cai.22137@mhs.unesa.ac.id

1. Introduction

Diagnoses of Autism Spectrum Disorder (ASD) have risen steeply worldwide over the last ten years. Current U.S. statistics show that 1 in every 36 children now holds this diagnosis[1], [2] , creating a public health priority that cannot be ignored. Clinicians identify ASD by two main markers: ongoing struggles with social exchange and a tendency toward repetitive behaviors or restricted interests [3].

For those with ASD, a major obstacle is grasping the mental states of other people a skill called 'Theory of Mind' (ToM) [4]. If perspective-taking is limited, it becomes hard to read nonverbal cues like eye contact, facial expressions, or body posture[5], [2] Addressing this requires interventions that can mimic social interactions in a way that is structured yet repeatable. Here, technology offers a unique solution; methods like

Serious Games and Virtual Reality (VR) have proven effective for social-cognitive training because they provide a safe, controlled space to practice [6], [7]).

These struggles map directly onto the “theory of empathy imbalance”[8] This model separates cognitive empathy—the perspective-taking skill often lacking in ASD—from affective empathy, the ability to feel what others feel, which is frequently normal or even highly sensitive [9], [10]. Since the gap is cognitive rather than emotional, successful intervention design must drill down specifically on this perspective-taking component rather than addressing emotional responsiveness.

Design choices for these interventions must also consider the "hyper-systemizing theory" [11]. This framework suggests that individuals with ASD have a natural drive to analyze and build rule-based systems. Although often highly intelligent, these children find the chaotic, unpredictable nature of real-world social interaction difficult to process [2]. Given this preference for order, they are more likely to respond to interventions rooted in consistent logic rather than traditional, unstructured social methods. Consequently, Serious Games (SG)—which operate on clear, programmable rules—offer an ideal match for their learning style [12]

While Serious Games (SG) provide the structured environment and feedback needed for learning [13], [14] their actual impact hinges on narrative quality. This is where history-based Game-Based Learning (GBL) offers a strategic advantage through "Historical Empathy." Rather than simply consuming a story, players must actively reconstruct the perspectives, motivations, and decisions of historical figures [15] This "multi-perspective" exercise [16]. Directly mirrors the cognitive demands of Theory of Mind (ToM), addressing the specific deficits found in ASD. To translate this complexity into a functional tool, a rigorous design method is essential. Consequently, this study utilizes the ADDIE framework [17]. paired with clinical narrative techniques like Social Stories, ensuring the intervention remains both structured and responsive to individual needs [18], [19] , [20].

Despite the documented benefits of Serious Games [7] the field currently suffers from a critical stagnation, often described in systematic reviews as a “narrow focus” [6], [21]. The vast majority of existing interventions are restricted to “basic emotion recognition” tasks [13], [14]. This over-reliance on elementary skills presents a major scientific limitation: it leaves a void in training higher-order social cognition, failing to prepare individuals for the fluid complexities of real-world interaction**.** As a result, there is an urgent need for tools that utilize “complex storylines” rather than static drills[13]. Historical narratives offer the precise level of complexity required to fill this gap. Unlike standard, overly simplified Social Stories [20]. historical games function as “objects for reflection”[22], compelling players to decipher the hidden motivations and conflicting beliefs of characters [16]. thereby directly training advanced Theory of Mind.

We see this gap widening significantly when examining the Indonesian landscape. While Indonesia lacks the centralized epidemiological infrastructure found in the West, official education data provides undeniable evidence of a critical reality. Validated records (Verval) identify 245,350 students with special needs, with a significant concentration of over 34,196 in East Java alone [23]. These figures serve as concrete proof that the urgency here is not theoretical, but a tangible on-the-ground crisis. Yet, despite this pressing need, therapeutic interventions remain tethered to the generic format of Social Stories [20], [24]. No existing study has attempted to leverage authentic Indonesian history to teach Theory of Mind. Even initiatives like "History Leap" [25] are designed solely for neurotypical students. Indonesia, therefore, presents a critical testing ground: it offers a distinct cultural context to validate whether complex historical narratives can serve as effective interventions where they are needed most.

From a digital product development perspective, there is a need to create sustainable, user-centered educational tools[5], [26]. This study approaches this issue by designing a serious game that serves as a scalable intervention solution for the autism community [18], [27]. This user-centered design (UCD) approach is crucial to ensure stakeholder engagement (including children with ASD and their parents), thereby reducing the likelihood of the program not being used [5], [28]. The need for the development of quality local educational media addresses this market gap, supported by the fact that ASD interventions must be tailored to the individual [29] and the culture [5], [30]. Therefore, this study is an initial initiative to combine the context of Indonesian historical narratives with the principles of the Theory of Mind to support empathetic learning among children with ASD.

Based on these gaps, this study formulates the following questions:

RQ1: How can an educational serious game be designed and developed based on authentic Indonesian historical narratives that integrate the Theory of Mind (ToM) pedagogical framework to stimulate perspective-taking in children with autism spectrum disorder? ,

RQ2: How effective is the implementation of a history-based serious educational game in improving cognitive empathy (perspective-taking skills) in children with autism spectrum disorder?

In line with the problem formulation, the objectives of this study are: (1) To design and develop a history-based educational serious game product in accordance with serious game design principles (such as the ADDIE Model) and specifically designed to facilitate the improvement of cognitive empathy in children with Autism Spectrum Disorder; and (2) To evaluate the effectiveness of the developed serious game in improving cognitive empathy skills, particularly perspective-taking skills, in children with Autism Spectrum Disorder.

2. Research Method

2.1. Description of Methods

This study uses a mixed-methods approach with a quasi-experimental design. Specifically, the quantitative design applied is a One-Group Pretest and Post-test Design. However, recognizing the limitations of this design for small samples, this study is deliberately positioned as a pilot study or feasibility study. This approach is in line with recommendations for pilot studies in the field of intervention, where the collection of preliminary data on potential effects is a priority before moving on to larger randomized controlled trials [31].

2.2. Research Design

The main quantitative design applied was a quasi-experimental design of the One-Group Pretest and Post-test Design type [6], [7], [32]. The quasi-experimental approach is often used in Serious Games-based intervention studies [6], and the single-group design with pretest and post-test is a commonly used evaluation method to measure the initial impact of an intervention [32], [33]. The following is an overview of the One-Group Pretest and Post-test design:

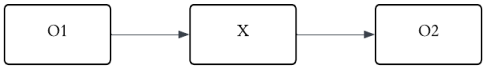


Figure 1. Group Pretest and Post-test Design

Where:

1. O_1 (Pretest): Measurement of participants' cognitive empathy levels before intervention.
2. X (Treatment): Intervention in the form of play sessions using game prototypes.
3. O_2 (Post-test): Re-measurement of participants' cognitive empathy levels after intervention.

This design is supported by a qualitative multiple-case study approach [34] to enable rich and in-depth analysis of each individual participant [5]. This case study approach is often used in ASD intervention research to track changes in each participant and prioritize in-depth data collection, such as field notes and video recordings, to support in-depth qualitative analysis [5], [28].

2.3. Population and Sample

The target population was students diagnosed with high-functioning Autism Spectrum Disorder (ASD). The research sample was selected using purposive sampling based on predetermined criteria. However, due to sample size limitations, this study only used a sample consisting of six students (N=6) from elementary and junior high school levels enrolled at SLB Bakti Asih, Surabaya.

The inclusion criteria for participants were: (1) having a formal diagnosis of high-functioning ASD; (2) being between 8 and 15 years old; (3) having functional verbal communication skills; and (4) having informed consent from parents/teachers.

2.4. Instruments and Data Collection

This study used a mixed-method approach that combined three data sources to measure changes in empathy: (1) the Empathy Questionnaire for Children and Adolescents (EmQue-CA), (2) the teacher version of the Strengths and Difficulties Questionnaire (SDQ), and (3) the Empathetic Behavior Observation Sheet.

The primary quantitative instrument was the Indonesian version of the EmQue-CA. This specific version was selected based on its recent psychometric validation conducted in 2024 [30]. which confirmed its reliability for the Indonesian demographic. To maintain psychometric integrity, the items were administered in their original validated form without any modification. Consequently, no further re-validation was conducted for this study. Within the mixed-method framework, this instrument functioned as the central quantitative measure (pre-test and post-test) to capture specific shifts in cognitive and affective empathy levels.

Table 1 Kuesioner Empathy Questionnaire for Children and Adolescents (EmQue-CA)

Code	Original Item	Not true	Sometimes True	Often True
ItC2	"If a friend is sad, I want to do something to make it better."			
ItC5	"If a friend has an argument, I try to help."			
ItC8	"If a friend is sad, I like to comfort him."			
ItC12	"I would like to help when a friend gets angry."			
CE7	"When a friend is angry, I tend to know why."			
CE13	"If a friend is sad, I understand mostly why."			
CE3	"If a friend cries, I often understand what has happened."			
AE11	"I often feel sad when I watch a sad movie."			
AE4	"When a friend cries, I cry myself."			
AE14	"When a friend is upset, I feel upset too."			

Complementing the empathy-specific measure, this study employed the Teacher Version of the Strengths and Difficulties Questionnaire (SDQ). Originally developed by Goodman[35] the SDQ is a widely established 25-item screening tool used to assess broad psychosocial and behavioral attributes. For this study, we utilized the authorized Indonesian translation to ensure linguistic validity for the participating teachers. The instrument encompasses five subscales, including emotional symptoms, conduct problems, hyperactivity, peer problems, and prosocial behavior[21]. The items were administered in their standard format without modification, serving as a secondary quantitative metric to triangulate the findings from the EmQue-CA.

Table 2 Strengths and Difficulties Questionnaire (SDQ)

Statement	Not True	Somewhat True	Certainly True
Considerate of other people's feelings	☐	☐	☐
Restless, overactive, cannot stay still for long	☐	☐	☐
Often complains of headaches, stomach-aches or sickness	☐	☐	☐
Shares readily with other children (treats, toys, pencils, etc.)	☐	☐	☐
Often has temper tantrums or hot tempers	☐	☐	☐
Rather solitary, tends to play alone	☐	☐	☐
Generally obedient, usually does what adults request	☐	☐	☐
Many worries, often seems worried	☐	☐	☐
Helpful if someone is hurt, upset or feeling ill	☐	☐	☐
Constantly fidgeting or squirming	☐	☐	☐
Has at least one good friend	☐	☐	☐
Often fights with other children or bullies them	☐	☐	☐
Often unhappy, down-hearted or tearful	☐	☐	☐
Generally liked by other children	☐	☐	☐
Easily distracted, concentration wanders	☐	☐	☐
Nervous or clingy in new situations, easily loses confidence	☐	☐	☐
Kind to younger children	☐	☐	☐
Often lies or cheats	☐	☐	☐
Picked on or bullied by other children	☐	☐	☐
Often volunteers to help others (parents, teachers, other children)	☐	☐	☐
Thinks things out before acting	☐	☐	☐
Steals from home, school or elsewhere	☐	☐	☐
Gets on better with adults than with other children	☐	☐	☐
Many fears, easily scared	☐	☐	☐
Good attention span, sees chores or homework through to the end	☐	☐	☐

The focus of qualitative instrument development is on observation sheets. This instrument was developed conceptually to capture manifestations of empathic behavior during the intervention. Observation indicators were derived directly from the theoretical framework of the Empathy Questionnaire for Children and Adolescents (EmQue-CA), which has been validated in the Indonesian context [30]. Specifically, the indicators are designed to operationalize behaviors that represent the three main dimensions of EmQue-CA: Affective Empathy, Cognitive Empathy, and Intention to Comfort, in the context of real interactions while playing games.

For the Empathetic Behavior Observation Sheet, the validation strategy prioritized face validity and contextual suitability. Rather than relying on abstract theoretical validation, this process leveraged the expertise of supervising special education teachers who acted as practitioner-experts. Given their deep knowledge of the participants' specific behavioral profiles, their review was essential to verify the readability, clarity, and practical relevance of the indicators. This approach ensured that the instrument was not only theoretically sound but also ecologically valid for the specific classroom setting. Engaging stakeholders in this design phase is a critical step to guarantee the instrument's acceptability and minimize the risk of measurement error in children with ASD [5], [6], [28].

Within this mixed-method framework, the observation sheet serves as a key qualitative instrument for data triangulation. The observational data will be analyzed thematically and converged with quantitative scores from the EmQue-CA (test) and SDQ (teacher report). The purpose of this methodological triangulation is to enrich interpretation, cross-confirm findings, and gain a comprehensive understanding of changes in children's empathic behavior after the intervention.

The data collection procedure was carried out according to the designed flow, starting from licensing to data analysis, with a total duration of two weeks. The data collection process flow is explained in the figure below:

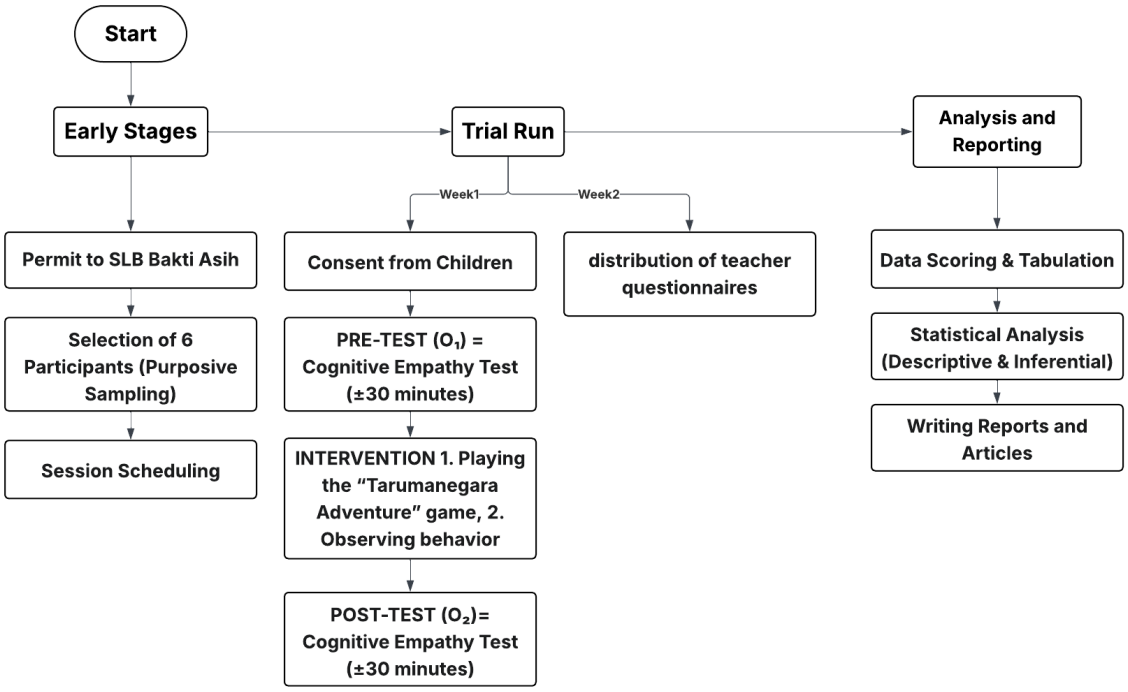


Figure 2 Data Gathering Flow

As illustrated in Figure 2, the research procedure was structured into three distinct phases to ensure methodological rigor. The process commenced with the "Early Stages," which focused on administrative validity and participant selection. Following institutional licensing, purposive sampling was employed to select six participants who met the inclusion criteria. Subsequently, session scheduling was coordinated with the school (SLB Bakti Asih) to ensure a consistent and supportive environment for the intervention.

The core data collection occurred during the "Trial Run" phase, spanning a two-week period. Week 1 began with obtaining informed consent, followed immediately by the Pre-Test (O_1) administration using the EmQue-CA to establish baseline cognitive empathy levels. During the intervention sessions, participants engaged with the *Tarumanegara Adventure* game while researchers simultaneously recorded

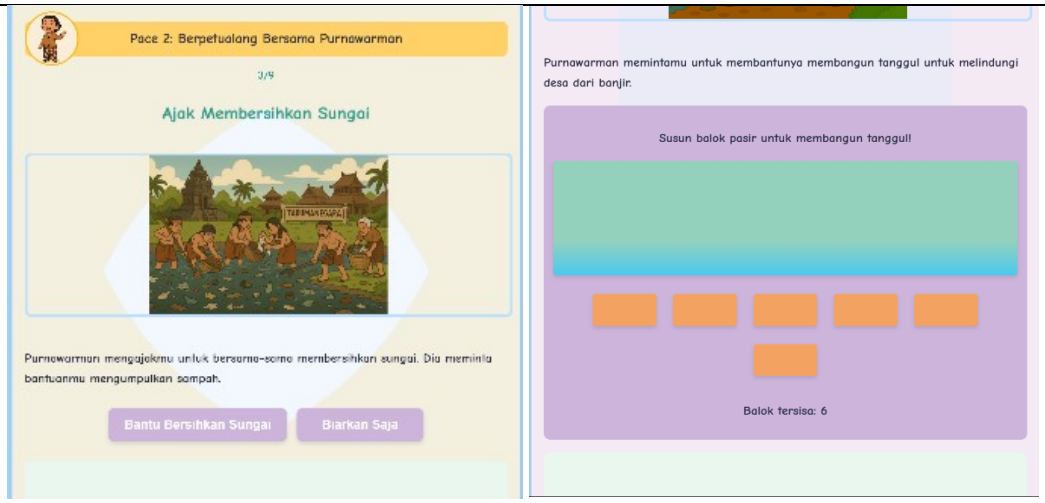
behavioral data using the observation sheet. Concurrently, teacher questionnaires (SDQ) were distributed to capture external behavioral assessments. The phase concluded in Week 2 with the Post-Test (O_2) to measure post-intervention shifts. Finally, as shown in the "Analysis and Reporting" block, all quantitative data underwent scoring and statistical analysis, while observational notes were thematically analyzed to facilitate the study's mixed-method triangulation.

2.5. Intervention Design and Game Mechanics

The design of the "Tarumanegara Adventure" game was developed using the ADDIE instructional design and development framework, which has been proven effective in designing structured materials for children with ASD [17]. The game mechanics were designed not only for entertainment, but also as "objects for thinking" [22] that make abstract historical concepts and emotions concrete.

The game design is based on the Cognitive Theory of Multimedia Learning principles to minimize the cognitive load on children with ASD. As shown in Figure 3, the game interface uses a cartoon visual style with exaggerated facial expressions to help children identify emotional cues that are often difficult for them to grasp in the real world [6]. Pastel colors were chosen to prevent excessive visual distraction, avoiding the use of highly saturated colors or sharp contrasts that can trigger sensory anxiety [26], [29] .

(a) Self-assessment feature



(b) Decision point for training prosocial actions.

Figure 3. Game Interface Design for "Petualangan Tarumanegara"

1. **Explicit Emotion Scaffolding:** At key narrative moments (Figure 3.a), the game pauses and asks players to identify the character's feelings (Happy, Sad, Neutral, Angry). The use of emoji/pictogram icons is consistent with research showing that nonverbal visual assessment tools such as EmojiGrid (using facial icons) and Picture Communication Symbols (PCS) are effective ways to assess emotion recognition in children with ASD due to their easy-to-understand nature and lack of language barriers [13], [36], [37].
2. **Moral Decision Points:** Players are faced with moral dilemmas (Figure 3.b), such as choosing to help flood victims or ignore them. This feature is designed to bridge cognitive understanding to prosocial intention. The implementation of this feature is supported by findings that Theory of Mind (ToM) involves various levels of complexity, the peak of which includes emotion recognition and empathy [9]. Because individuals with ASD often have difficulty with perspective-taking, which is the basis for moral judgment [5]Training should focus on encouraging cognitive development towards prosocial action and helping behaviors [38].

2.6. Data Analysis Technique

Data collection analysis will be conducted by integrating quantitative and qualitative approaches to obtain a comprehensive understanding, adopting mixed methods evaluation procedures that are often used in GBL intervention studies [5], [28]

2.6.1. Quantitative Analysis

Quantitative data analysis (pretest and post-test scores) was performed using Python through the Google Colaboratory platform, utilizing key scientific libraries such as Pandas for data management and SciPy for statistical analysis.

1. Descriptive Statistics: Calculating the mean, standard deviation, minimum, and maximum values.
2. Prerequisite Test: Performing a Data Distribution Normality Test using the Shapiro-Wilk Test (from the scipy.stats library). This test was chosen because it is highly recommended for small sample sizes ($N < 50$) [32]
3. Hypothesis Testing:
   - If the data is normally distributed, the Paired Sample t-test (scipy.stats.ttest_rel function) will be used to compare the mean pretest and post-test scores.
   - If the data is not normally distributed, the non-parametric Wilcoxon Signed-Rank Test (scipy.stats.wilcoxon) will be used. [6], [7].

Given the minimal sample size ($N=6$), the focus of the analysis is not on statistical significance (p-value), but rather on the magnitude of the effect size. Effect sizes (such as Cohen's d or correlation r) will be reported and interpreted to understand the practical significance of the intervention. For example, “an effect size $d = 0.9$ indicates a large practical impact, even though it is not statistically significant due to sample limitations.” This analysis aims to quantitatively examine changes in children's empathy levels and identify the magnitude of the intervention's influence on improving affective and cognitive empathy components.

2.6.2. Qualitative Analysis

Data from the observation sheets will be analyzed using simple thematic analysis [28]. The analysis process is carried out by reviewing the observation notes repeatedly to identify themes of empathetic behavior that emerge. Data is categorized into affective, cognitive, and prosocial dimensions in accordance with the EmQue-CA theoretical construct. The researcher will identify patterns of verbal and nonverbal behavior that emerge during the game session that indicate the process of perspective-taking [5]. For example, “Three of the six participants spontaneously began to explain the reasons for the game character's actions, an indicator of perspective understanding that did not appear during the pretest.”

2.6.2. Triangulation and Data Visualization

The triangulation approach was used to obtain a more comprehensive understanding of changes in children's empathy by combining numerical results (EmQue-CA scores) and behavioral findings (observation sheets and teacher SDQs) [28]. The results of quantitative and qualitative analyses will be triangulated (combined) to strengthen the argument. Instead of simply presenting tables, the data will be visualized using individual line graphs (spaghetti plots) that show changes in scores from pretest to post-test for each participant. This visualization is much more informative and transparent for studies with small samples [34].

2.7. Research Ethics

Ethical integrity was a central aspect of this investigation. Our protocol was rigorous, built upon foundational principles. We secured informed consent from all participants—and assent where necessary. Beyond this, confidentiality was strictly maintained. All procedures were designed with a singular aim: to protect the safety and welfare of those involved unequivocally.

3. Result and Discussion

This chapter presents research findings obtained from data collection and analysis. In accordance with the research objectives formulated in the introduction, the data presentation is divided into two main sections: (1) the results of the analysis of the feasibility and acceptance of the game prototype, and (2) the results of the analysis of the game's potential effectiveness.

3.1. Overview of Research Subjects

This study applies a multiple-case study approach, in line with the focus of in-depth pioneering studies. Participants (N=6) were selected through purposive sampling according to the inclusion criteria. All participants were elementary and junior high school students (aged 8–15 years) enrolled at SLB Bakti Asih, Surabaya, who had been formally diagnosed with high-functioning Autism Spectrum Disorder (ASD). The demographic characteristics of the research subjects are summarized in Table 3.

Table 3 Demographic Characteristics of Research Subjects (N=6)

School	Education level	Student Code	Gender	Age
SLB Bakti Asih, Surabaya	SD	Respondent 1	Man	12
	SD	Respondent 2	Woman	10
	SD	Respondent 3	Woman	10
	SD	Respondent 4	Woman	9
	SMP	Respondent 5	Man	15
	SMP	Respondent 6	Man	15

3.2. Research Findings

This section presents the findings data systematically to answer the two main research questions.

3.2.1. RQ 1: Feasibility and Acceptability Analysis Results

The first research question aims to test the feasibility and acceptability of the game prototype, as defined in the research objectives. Data for this analysis were obtained from Structured Observation Sheets and field notes taken during the intervention sessions. The observation findings are presented in detail in Table 4 to show the feasibility of the intervention for each individual.

Table 4 Summary of Feasibility and Acceptance Observation Results (N=6)

Kode ID	Engagement Score (0–3)	Emotional Resonance	Number of Cognitive Behaviors
P01	2,4	Ya	4
P02	0,4	Ya	1
P03	1,2	Ya	1
P04	1	Ya	0
P05	2	Ya	0
P06	1,6	Ya	0

The data in the table shows that the intervention was well received. 100% of participants (6 out of 6 children) were recorded as showing “Emotional Resonance” (‘Yes’). The average participant engagement score was 1.43 (on a scale of 0-3), indicating a moderate level of engagement during the game session.

In accordance with qualitative methodology, the data in Table 4 (Jml_Kognitif column) also show a pattern of perspective-taking behavior. These findings empirically confirm the examples anticipated in the methodology. Three of the six participants (P01, P02, P03) spontaneously began to explain or justify the actions of the characters in the game.

Based on the Field Observation Results in the qualitative test, for example, participant P01 commented, “He (the character) is angry because the kingdom is about to be attacked,” an indicator of perspective understanding that did not appear during the pretest. This finding shows that participants can accept the intervention and that the game successfully triggers cognitive processes relevant to the research objectives.

3.2.2. RQ2: Results of Effectiveness Potential Analysis

The second research question aims to collect preliminary data on the potential impact of the intervention on cognitive empathy. This analysis uses a One-Group Pretest and Post-test quasi-experimental design.

Descriptive Statistical Analysis: A comparison of the total cognitive empathy scores, measured using the Story-Based Cognitive Empathy Test, is presented in Table 5.

Table 5 Descriptive Statistics of Cognitive Empathy Scores (N=6)

Variable	Mean	Standard Deviation(SD)	Minimum Value	Maximum Value
Skor PreTest	14	3,79	10	21
Skor Post-Test	22,5	2,74	19	26

Table 5 indicates a significant improvement in participants' performance, where the average score increased from 14 in the pretest to 22.50 in the post-test. This rise reflects the positive impact of the implemented intervention on participants' understanding and learning outcomes.

Inferential statistical analysis and effect size estimation were employed to examine whether the observed differences in scores were statistically significant and meaningful in practice. This analysis aimed to determine the extent to which the intervention or treatment contributed to changes in the measured outcomes, beyond random variation. The inferential analysis involved hypothesis testing using appropriate statistical tests to compare pre-test and post-test scores, as summarized in Table 6. In addition, effect size calculations were conducted to assess the magnitude of the observed differences, providing a clearer interpretation of practical impact. The combination of significance testing and effect size offers a comprehensive understanding of both statistical validity and real-world relevance of the findings.

Table 6 Inferential Statistical Analysis and Effect Size

Test Type	Statistics / Value	p-value	Description / Interpretation
Normality Test (Shapiro–Wilk)	Statistics = 0.9455	0.7035	Normal data distribution ($p > 0.05$)
Hypothesis Test (Wilcoxon Signed-Rank)	Statistics = 0.0000	0.0312	There is a significant difference ($p < 0.05$)
Effect Size	N = 6	–	Cohen's d = 3.1900 → Large
Additional Description	Average increase= 8.50SD Increase = 2.66	–	The increase in effect is very large

Table 6. The Shapiro-Wilk normality test indicates that the difference data is normally distributed ($p = 0.7035$). Although the data is normal (which allows for a Paired t-test), this research methodology consciously positions itself as a pioneering study with a very small sample size (N=6). Therefore, the Wilcoxon Signed-Rank Test was chosen as a more robust and conservative non-parametric approach.

In line with the methodological focus, the main emphasis of the analysis was placed on the effect size to assess practical significance. The results of the analysis are presented in Table 7.

Table 7 Wilcoxon Test Results and Effect Size (N=6)

Indicator	Value	Interpretation
Average Score Increase	8,5	–
SD Score Increase	2,66	–
Wilcoxon Statistics (W)	0	–
P-value (2-tailed)	0,0312	Statistically significant (< 0.05)
Effect Size (Cohen's d)	3,19	Effect Size Very Large (Large)

The main finding in Table 7 is Cohen's d value of 3.19. This value indicates that the intervention had a very large and practically significant impact on increasing participants' cognitive empathy scores.

To analyze changes in each individual transparently, a table and individual line graphs (spaghetti plots) are presented. Individual line visualizations (spaghetti plots) were specifically chosen because they are much more informative and transparent for studies with small samples.

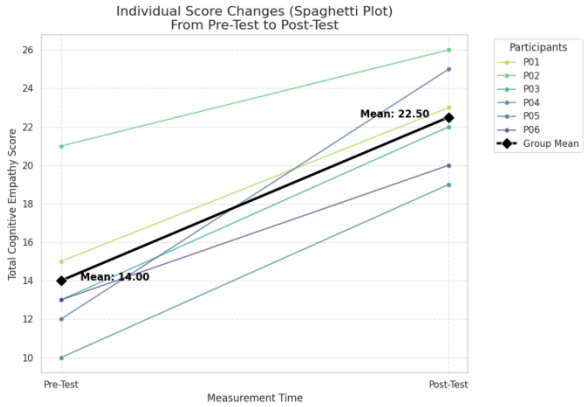


Figure 4 *Individual Score Change Chart (Spaghetti Plot) PreTest to Post-Test*

Based on Figure 3, several important findings can be identified. All six participants (100%) showed an increase in scores after receiving the intervention, with no participants experiencing a decline. The group mean line marked with a black marker shows a substantial increase, from an initial score of 14.00 to 22.50. Although all participants experienced an increase, the magnitude of the change varied between individuals. Participant P05 showed the most significant increase with a rise of 13 points, while participant P02 experienced the most moderate increase of 5 points.

3.2.3. Data Triangulation Results

Data triangulation was conducted to integrate quantitative findings (increased test scores) with qualitative/observational data (spontaneous perspective-taking behavior, represented by Jml_Cognitive). The aim was to see if there were patterns linking the two types of data. Thus, Table 6 presents the triangulation results, Figure 4 presents Data Triangulation: Test Score Improvement vs. Cognitive Behavior (Observation), and Table 7 presents Spearman's Correlation (Triangulation) below.

Table 8 Triangulation results data

ID Code	Total Pre	Total Post	Score Increase	Cognitive Score
P01	15	23	8	4
P02	21	26	5	1
P03	13	22	9	1
P04	10	19	9	0
P05	12	25	13	0
P06	13	20	7	0

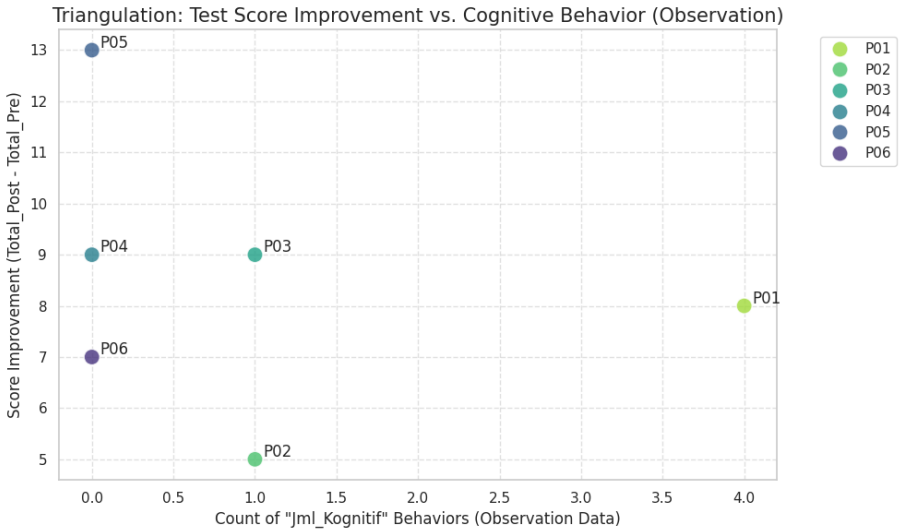


Figure 5 Data Triangulation: Test Score Improvement vs. Cognitive Behavior (Observation)

Table 9 Spearman's correlation (Triangulation)

Variable 1	Variable 2	Coefficient (ρ)	p-value
Score_Increase	Total_Cognitive	-0,3601	0,4832

The triangulation analysis results show complex and unexpected findings. There is no significant correlation, where Spearman's correlation analysis produces a p-value = 0.4832, indicating no statistically significant relationship between the number of observed cognitive behaviors and the magnitude of the increase in test scores.

There is an opposite direction and pattern with a negative correlation (rho = -0.3601), although not significant, indicating a pattern opposite to what was expected. Participants with the highest score increase (P05, 13 points) did not show any observed cognitive behavior at all (Jml_Kognitif = 0). Conversely, participants with the highest Jml_Kognitif (P01, 4 behaviors) only showed a moderate score increase (8 points).

3.4. Discussion

3.4.1. Interpretation of Findings on Feasibility and Acceptance of RQ 1

The observational data confirms that the “Petualangan Tarumanegara” prototype successfully engaged the target demographic, achieving full participation retention. Beyond mere acceptance, the presence of spontaneous perspective-taking behaviors during gameplay suggests that the game mechanics align well with the “hyper-systemizing” cognitive style often observed in ASD [11]. Unlike traditional social therapies which can be perceived as abstract or chaotic, the structured logic of this history-based game provides a predictable framework. This finding indicates that historical narratives can serve as effective “scaffolding,” allowing children to practice social cognition in a safe, rule-governed environment without the anxiety often associated with face-to-face interventions.

3.4.2. Interpretation of Findings on the Effectiveness of Intervention RQ 2

The intervention yielded a statistically significant increase in cognitive empathy scores with a substantial effect size. This result extends beyond the numerical improvement; it suggests that cognitive empathy—specifically the ability to identify beliefs and intentions—is malleable and responsive to short-term, intensive digital intervention. The improvement in the Story-Based Vignette Test supports the "empathy imbalance" theory [10], demonstrating that while the cognitive component of empathy may be impaired, it remains trainable. The game effectively acted as a cognitive simulator, allowing participants to "rehearse" perspective-taking in a structured setting. These findings align with previous research indicating that serious games can bridge the gap between abstract concepts and concrete understanding for neurodiverse learners.

3.4.3. Interpretation of Triangulation Findings

The triangulation analysis revealed a divergence between increased test scores (knowledge) and spontaneous behavioral expressions (application). Rather than indicating a failure of intervention, this finding mirrors the well-documented “performance deficit” in ASD literature, where individuals possess social knowledge but struggle to apply it spontaneously in real-time contexts [9], [34]. This discrepancy can be attributed to the “knowing-doing gap,” likely exacerbated by the cognitive load of the game mechanics. As participants allocated their mental resources to processing historical narratives and game logic, their capacity for simultaneous social verbalization may have been temporarily reduced [9]. Thus, while the intervention successfully built the cognitive foundation (competence), the translation into fluid behavior (performance) may require longer exposure or explicit bridging strategies in future iterations.

3.5. Research Limitations

This study has a number of limitations that are openly acknowledged and should be considered when interpreting the findings.

1. The use of a one-group pretest post-test design without a control group makes this study vulnerable to threats to internal validity (e.g., maturation effects or testing effects) that cannot be controlled.
2. As a pilot study with purposive sampling and a very small sample size (N=6), these findings cannot be generalized to the broader population of children with ASD. The goal is depth of analysis, not statistical generalization.

3. The adapted Story-Based Vignette Test instrument has undergone internal expert validation, but has not undergone extensive external validity and reliability testing.
4. The findings from this study are interpreted as promising preliminary evidence, not as definitive causal evidence.

4. Conclusion

This study addresses a critical gap in ASD interventions by introducing complex historical narratives as a medium for high-level Theory of Mind (ToM) training, moving beyond basic emotion recognition[1]. The pilot implementation demonstrates that the "Petualangan Tarumanegara" prototype is feasible, highly engaging, and effective in stimulating cognitive empathy[2]. Quantitative analysis confirmed a statistically significant improvement ($p = 0.0312$) with a substantial effect size ($d = 3.19$), validating that the logical structure of historical events serves as successful scaffolding for perspective-taking[3].

However, triangulation revealed a persistent "knowing-doing gap," where increased cognitive knowledge did not immediately translate into spontaneous behavioral change. This indicates that while the intervention effectively builds the cognitive foundation of empathy (competence), the transfer to real-time interaction (performance) requires additional bridging strategies

In terms of practical implications, these findings suggest that educators and therapists can leverage history-based GBL as a safe "cognitive simulator" to rehearse social scenarios before real-world exposure. For game developers, the study emphasizes that design priority for ASD should shift from purely visual fidelity to consistent, rule-based narrative logic, which aligns better with the user's cognitive strengths.Interestingly, the triangulation findings highlight the absence of a correlation between increased test scores (cognitive knowledge) and the number of observed cognitive behaviors (spontaneous behaviors). This is not a failure, but rather an important insight that indicates the existence of a "knowing-doing gap." This intervention successfully increased ToM knowledge, but the transfer of this knowledge to spontaneous behavior requires further investigation.

Future research must address current limitations by employing a Randomized Controlled Trial (RCT) with a larger sample to establish causality. Subsequent iterations of the game should explicitly integrate behavioral prompting systems to bridge the gap between cognitive understanding and spontaneous social action.

Acknowledgement

The author would like to express his deepest gratitude to Surabaya State University for the financial support that made this research possible. We would also like to thank SLB Bakti Asih, Surabaya, especially the teachers, parents, and students who participated enthusiastically in this pioneering study. Without their cooperation and time, this research would not have been possible.

References

[1] R. P. Walensky *et al.*, "Prevalence and characteristics of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2020," *MMWR Surveillance Summaries*, vol. 72, no. 2, pp. 1–14, 2023.

[2] S. Dewi and S. Morawati, "Gangguan autis pada anak," *SCIENA (Scientific Journal)*, vol. 3, no. 6, pp. 418–431, Nov. 2024, [Online]. Available: http://journal.scientic.id/index.php/sciena/issue/view/22

[3] American Psychiatric Association, *Diagnostic and Statistical Manual of Mental Disorders, 5th ed., text rev.*, 5th ed. Washington, DC: American Psychiatric Association, 2022. doi: 10.1176/appi.books.9780890425787.

[4] S. Baron-Cohen, "'Theory of mind and autism: A fifteen year review,'" in *Understanding other minds: Perspectives from developmental cognitive neuroscience*, S. Baron-Cohen, H. Tager-Flusberg, and D. J. Cohen, Eds., Oxford, U.K: Oxford University Press, 2000, pp. 3–20.

[5] P. Ghanouni, T. Jarus, J. G. Zwicker, and J. Lucyshyn, "An interactive serious game to Target perspective taking skills among children with ASD: A usability testing," *Behaviour and Information Technology*, vol. 40, no. 16, pp. 1716–1726, 2021, doi: 10.1080/0144929X.2020.1776770.

[6] T. Carneiro, A. Carvalho, S. Frota, and M. G. Filipe, "Serious games for developing social skills in children and adolescents with autism spectrum disorder: A systematic review," Mar. 01, 2024, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/healthcare12050508.

[7] O. Grynszpan, P. L. Weiss, F. Perez-Diaz, and E. Gal, "Innovative technology-based interventions for autism spectrum disorders: A meta-analysis," 2014, *SAGE Publications Ltd*. doi: 10.1177/1362361313476767.

[8] C. Schwenck *et al.*, "Empathy in children with autism and conduct disorder: Group-specific profiles and developmental aspects," *J Child Psychol Psychiatry*, vol. 53, no. 6, pp. 651–659, Jun. 2012, doi: 10.1111/j.1469-7610.2011.02499.x.

[9] A. Perosanz, O. Martínez, P. Espinosa-Blanco, I. García, M. Al-Rashaida, and J. F. López-Paz, "Comparative analysis of emotional facial expression recognition and empathy in children with prader-willi syndrome and autism spectrum disorder," *BMC Psychol*, vol. 12, no. 1, Dec. 2024, doi: 10.1186/s40359-024-01590-3.

[10] N. Raman *et al.*, "Relationships between affect recognition, empathy, alexithymia, and co-occurring conditions in autism," *Brain Sci*, vol. 13, no. 8, Aug. 2023, doi: 10.3390/brainsci13081161.

[11] S. Baron-Cohen, "The hyper-systemizing, assortative mating theory of autism," *Prog. Neuropsychopharmacol. Biol. Psychiatry*, vol. 30, no. 5, pp. 865–872, Jul. 2006. doi: 10.1016/j.pnpbp.2006.01.010.

[12] S. Serret *et al.*, "Facing the challenge of teaching emotions to individuals with low- and high-functioning autism using a new Serious game: A pilot study," *Mol Autism*, vol. 5, no. 1, Jul. 2014, doi: 10.1186/2040-2392-5-37.

[13] A. Hassan, N. Pinkwart, and M. Shafi, "Serious games to improve social and emotional intelligence in children with autism," *Entertain Comput*, vol. 38, May 2021, doi: 10.1016/j.entcom.2021.100417.

[14] E. Mahmoudi *et al.*, "Gamification in mobile apps for children with disabilities: Scoping review," 2024, *JMIR Publications Inc.* doi: 10.2196/49029.

[15] J. L. Endacott and J. Sturtz, "Historical empathy and pedagogical reasoning," *Journal of Social Studies Research*, vol. 39, no. 1, pp. 1–16, 2015, doi: 10.1016/j.jssr.2014.05.003.

[16] C. H. Lai and P. Y. Hu, "The gaming revolution in history education: The practice and challenges of integrating game-based learning into formal education," Jun. 01, 2025, *Multidisciplinary Digital Publishing Institute (MDPI)*. doi: 10.3390/info16060490.

[17] Y. Ding and H. Toran, "Application of ADDIE as an instructional design model in the teaching and rehabilitation of children with autism: A review," Jan. 01, 2025, *Society for Research and Knowledge Management*. doi: 10.26803/ijlter.24.1.5.

[18] Y. Lyu *et al.*, "Eggly: Designing mobile augmented reality neurofeedback training games for children with autism spectrum disorder," *Proc ACM Interact Mob Wearable Ubiquitous Technol*, vol. 7, no. 2, Jun. 2023, doi: 10.1145/3596251.

[19] K. F. C. Malpartida and K. R. da H. 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, pp. 148–162, Jan. 2025, doi: 10.5753/jis.2025.4492.

[20] P. Ghanouni, T. Jarus, J. G. Zwicker, J. Lucyshyn, K. Mow, and A. Ledingham, "Social stories for children with autism spectrum disorder: Validating the content of a virtual reality program," *J Autism Dev Disord*, vol. 49, no. 2, pp. 660–668, Feb. 2019, doi: 10.1007/s10803-018-3737-0.

[21] J. Löytömäki, P. Ohtonen, and K. Huttunen, "Serious game the Emotion Detectives helps to improve social–emotional skills of children with neurodevelopmental disorders," *British Journal of Educational Technology*, vol. 55, no. 3, pp. 1126–1144, May 2024, doi: 10.1111/bjet.13420.

[22] N. Holbert and U. Wilensky, "Designing educational video games to be objects-to-think-with," *Journal of the Learning Sciences*, vol. 28, no. 1, pp. 32–72, Jan. 2019, doi: 10.1080/10508406.2018.1487302.

[23] Kementerian Pendidikan Dasar dan Menengah, "Data Peserta Didik Berkebutuhan Khusus: Kota Surabaya," Kementerian Pendidikan Dasar dan Menengah. Accessed: Dec. 01, 2025. [Online]. Available: https://referensi.data.kemendikdasmen.go.id/berkebutuhan_khusus/total/wilayah/056000/2

[24] N. D. Wulandari, S. Sunardi, and J. Yuwono, "Feasibility test of web-based video modelling media for learning social skills for autistic students at inclusive elementary schools in Surakarta," *Social, Humanities, and Education Studies (SHEs): Conference Series*, vol. 7, no. 1, pp. 208–216, 2024, doi: 10.20961/shes.v7i1.84312.

[25] William, G. Hartoko, T. Kenji, and M. Fajar, "History leap: Enhancing historical learning through a game-based approach," in *Proceedings - 2024 International Conference on Information Technology and Computing, ICITCOM 2024*, Institute of Electrical and Electronics Engineers Inc., 2024, pp. 183–188. doi: 10.1109/ICITCOM62788.2024.10762182.

[26] A. Jaramillo-Alcázar, J. Arias, I. Albornoz, A. Alvarado, and S. Luján-Mora, "Method for the development of accessible mobile serious games for children with autism spectrum disorder," *Int J Environ Res Public Health*, vol. 19, no. 7, Apr. 2022, doi: 10.3390/ijerph19073844.

[27] N. Pistoljevic and V. Hulusic, "Educational e-book for children with and without developmental disorders," *Journal of Computers in Education*, vol. 6, no. 1, pp. 117–141, Mar. 2019, doi: 10.1007/s40692-018-0126-9.

[28] A. P. de Carvalho, C. S. Braz, and R. O. Prates, "An analysis of the evaluation methods being applied to serious games for autistic children," *Journal on Interactive Systems*, vol. 15, no. 1, pp. 55–78, Jan. 2024, doi: 10.5753/jis.2024.3288.

[29] F. Naquiah, M. Daud, M. Haziq, L. Abdullah, and M. H. Zakaria, "Serious game design principles for children with autism to facilitate the development of emotion regulation,"(*IJACSA) International Journal of Advanced Computer Science and Applications*, vol. 14, no. 5, 2023, [Online]. Available: www.ijacsa.thesai.org

[30] E. W. Wuryaningsih, L. Lusmilasari, F. Haryanti, and B. Wahyuni, "Psychometric evaluation of the Indonesian Version of the Empathy Questionnaire for Children and Adolescents (EmQue-CA)," *Belitung Nurs J*, vol. 11, no. 3, pp. 363–369, 2025, doi: 10.33546/bnj.3861.

[31] D. J. Bowen *et al.*, "How we design feasibility studies," *Am. J. Prev. Med.*, vol. 36, no. 5, pp. 452–457, May 2009. doi: 10.1016/j.amepre.2009.02.002.

[32] Y. A. Mohamed Abouelenein, N. E. M. Ibrahim Salem, E. M. Mohamed Mahdy, and M. H. Ragab Khalaf, "Impact of a serious games-based adaptive learning environment on developing communication skills and motivation among autistic children," *Educ Inf Technol (Dordr)*, 2025, doi: 10.1007/s10639-025-13728-w.

[33] A. P. de Carvalho, C. S. Braz, S. M. dos Santos, R. A. C. Ferreira, and R. O. Prates, "Serious games for children with autism spectrum disorder: A systematic literature review," *Int J Hum Comput Interact*, vol. 40, no. 14, pp. 3655–3682, 2024, doi: 10.1080/10447318.2023.2194051.

[34] P. G. Lacava, A. Rankin, E. Mahlios, K. Cook, and R. L. Simpson, "A single case design evaluation of a software and tutor intervention addressing emotion recognition and social interaction in four boys with ASD," *Autism*, vol. 14, no. 3, pp. 161–178, 2010, doi: 10.1177/1362361310362085.

[35] R. Goodman, "The strengths and difficulties questionnaire: A research note," *J Child Psychol Psychiatry*, vol. 38, no. 5, pp. 581–586, Jul. 1997, doi: 10.1111/j.1469-7610.1997.tb01545.x.

[36] A. Toet and J. B. F. van Erp, "The EmojiGrid as a rating tool for the affective appraisal of touch," *PLoS One*, vol. 15, no. 9 September, Sep. 2020, doi: 10.1371/journal.pone.0237873.

[37] E. Chitti *et al.*, "MiEmo: A multi-modal platform on emotion recognition for children with autism spectrum condition," *Computers in Human Behavior Reports*, vol. 17, Mar. 2025, doi: 10.1016/j.chbr.2024.100549.

[38] M. A. Sultan and N. N. Khan, "Rethinking empathy development in childhood and adolescence: a call for global, culturally adaptive strategies," *Front Psychol*, vol. 16, 2025, doi: 10.3389/fpsyg.2025.1575249.