RedFame

Journal of Education and Training Studies
Vol. 14, No. 1; January 2026
ISSN 2324-805X E-ISSN 2324-8068
Published by Redfame Publishing
URL: http://jets.redfame.com

AI &VR Technology in the Application of ASD Language Development--A Systematic Review

Fangye Luo[1], Yiding Lu[2]

[1]School of Humanities and Languages, University of New South Wales, Sydney, Australia

[2]School of Education and Social Work, The University of Sydney, Sydney, Australia

Correspondence: Fangye Luo, School of Humanities and Languages, University of New South Wales, Sydney, NSW 2052, Australia; Yiding Lu, School of Education and Social Work, The University of Sydney, Sydney, Camperdown NSW 2050, Australia.

Received: September 8, 2025 Accepted: October 9, 2025 Online Published: October 10, 2025

doi:10.11114/jets.v14i1.8035 URL: https://doi.org/10.11114/jets.v14i1.8035

Abstract

AI and VR Technology have been widely utilized in developing the language abilities of individuals with Autism Spectrum Disorder (ASD) in recent studies. Nevertheless, this domain faces fragmented outcomes and various limitations. By highlighting the opportunities and challenges in recent studies, this review offers recommendations for future synthesized experimental models and interdisciplinary studies, aiming to foster a more unified and nuanced understanding of AI and VR in autistic language learning. This study examined and synthesized nine relevant empirical studies selected through a deductive thematic analysis framework aligned with PRISMA guidelines. These studies are analyzed in terms of experimental procedure details, outcomes, and challenges of AI and VR. The results showed the divergence of experimental procedures, a significant yet lacking generalized outcome, and the coexistence of opportunities and challenges in the applications. The limitations persist in inadequately unified intervention designs, generalizability of language use, technological and resource restrictions, and ethical risks. The conclusions highlight the requirements for model development, long-term experiment, multimodal affordances, and detailed outcomes. This research contributes to a focused and contemporary synthesis in the critical examination of applications of VR and AI technologies in language development for ASD.

Keywords: language development, ASD, Artificial Intelligence (AI), Virtual Reality (VR)

1. Introduction

Autism Spectrum Disorder (ASD) is a neurodevelopmental disorder, highlighting challenges in social communication and interaction (Lima et al., 2023). They face a deficit in language acquisition, characterized by the inappropriate use of language in social communication, and have difficulty with generalization (Koegel & Koegel, 2006; Loukusa & Moilanen, 2009). With the development of AI and VR technology and their broad application, the language educational field has been experiencing an unprecedented revolution. Leveraging the technologies facilitates autistic language development in the diagnosis and interventions. They can not only help autistic individuals learn English vocabulary by stimulating pronunciation skills and language articulation in a more engaging and meaningful way, serving as a platform for instilling leisure learning (Hashim et al., 2022), but also identify and reveal the differences in how learners with ASD use language (Themistocleous et al., 2024). Autistic individuals enjoy working with technological tools, which help them imitate, maintain eye contact, and engage in social interactions, thereby developing communication skills and awareness (Aresti-Bartolome & Garcia-Zapirain, 2014).

The existing literature presents significant fragmentation in analyzing AI & VR interventions in promoting communication skills. In the meta-analysis review, Sohn et al. (2025) investigated the implementation of AI technologies in the screening, diagnosis, intervention, and caregiver support for ASD. Voultsiou and Moussiades (2025) identified the main themes and challenges of the combination of AI, VR, and LLMs in special education. Xu et al. (2024) conducted a meta-analysis to explore the effectiveness of technology interventions in improving developmental skills for children and adults with ASD. Nevertheless, fewer reviews identified the potential challenges and opportunities in the field of language skills. Gu et al. (2025) explored the affordances and implications of an AI chatbot for intervention in conversational skills. Nevertheless, existing studies haven't systematically explored the tools for children with ASD in language learning. The

present studies warrant integration, thereby inspiring future studies that address language learning for minority groups. By synthesizing recent studies of applications of AI and VR technologies in language teaching for ASD, the significance of the study lies in identifying key areas leveraging the technologies to facilitate future language research and instruction support. Future studies can conduct multimodal and interdisciplinary studies, develop intervention models and long-term experiments, and emphasize user experience and outcomes.

This study aims to review the application of AI and VR technology in language skills, highlighting its interventional details, potential advantages, and challenges. The systematic review dives into the following questions:

Q1: What are the experimental procedure details in interventions?

Q2: What are the outcomes of language skills in these experiments?

Q3: What are the limitations of these studies?

2. Methodology

2.1 Research Method

To provide a more comprehensive and systematic response to the three core questions of this study, this review adopts a systematic literature review methodology and mostly follows the Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA) statement (Page et al., 2021). PRISMA is an important systematic review tool in psychology, educational research, and medical research (Fuentes, 2022). It guides researchers in conducting systematic reviews and meta-analyses. Additionally, it improves the transparency and reliability of research (Brennan & Munn, 2021; Liberati, 2019).

2.2 Inclusion and Exclusion Criteria

To comprehensively, accurately, and systematically identify empirical studies relevant to this research, this study adopted the PICO framework to establish inclusion and exclusion criteria. PICO originally stands for patient, intervention, comparison, and outcome (Richardson et al., 1995). The PICO framework is widely used in evidence-based medical research (Schiavenato & Chu, 2021), but according to Nishikawa-Pacher (2022), the PICO process is not limited to evidence-based medical research and can also be applied in the development of retrieval strategies across any discipline. Therefore, based on the objectives of this study, the researchers established specific inclusion and exclusion criteria, as detailed below:

Table 1. Inclusion and exclusion criteria for selecting studies in this review

Screening Dimension	Criteria
Population	Include: Participants are children or adolescents with autism spectrum disorder (ASD). Exclude: Non-ASD population, or study subjects consisting solely of adults with ASD.
Intervention	Include: Use AI or VR or AR technology. The purpose of the intervention is language-related. Exclude: Research that does not utilise AI or VR or AR technology. The application of technology is unrelated to language learning.
Study Design	Include: Empirical studies (e.g., experimental, quasi-experimental, observational, qualitative, mixed methods) Exclude: Review articles, theoretical papers, commentaries, opinion pieces, news/blogs
Outcome	Include: Key outcomes related to language development: language skills (vocabulary acquisition, pronunciation, conversational skills), social and communication skills (eye contact, social interaction), learning motivation, engagement, and emotional support. Exclude: No results related to language learning or communication skills were reported. The results were limited to medical and neurological indicators (such as brain imaging and drug efficacy) and were unrelated to language education.
Publication Year	Include: Published between 2015 and 2025 Exclude: Published before 2015 or after 2025
Publication Type	Include: Peer-reviewed journal articles, conference papers, academic theses Exclude: Unreviewed articles, social media content, teaching materials
Language	Include: Published in English Exclude: Non-English publications

2.3 Search Strategy

This review uses Web of Science (WoS) as the literature search engine. WoS is a long-established, widely used, and authoritative database of research publications and citations (Birkle et al., 2020). It supports a wide range of scientific tasks and disciplinary research across different knowledge domains and can also serve as a dataset for large-scale data-

intensive research (Li et al., 2018). Additionally, according to Birkle et al. (2020), WoS can support simultaneous queries of up to 50 Boolean logic keywords, and using this search strategy can yield a set of highly relevant and meaningful results. Therefore, the literature retrieval strategy in this paper also uses Boolean logic keywords as search keywords. Boolean logic operators "AND" and "OR" are employed in the search keywords, and the final retrieval formula is constructed using keywords and their derivative related terms (e.g., AI, VR, ASD, language). The "AND" operator is used to narrow the search scope, while the "OR" operator is used to avoid omitting relevant literature. The final specific retrieval formula is as follows:

("artificial intelligence" OR "AI" OR "virtual reality" OR "VR" OR "intelligent tutoring system") AND ("autism spectrum disorder" OR "ASD" OR "autistic children" OR "children with autism") AND ("language")

2.4 Screening Procedure

During the literature screening phase, this review followed the literature screening process recommended by the PRISMA statement (Page et al., 2021). We first entered the previously identified logical keywords into Web of Science (WoS) for searching, yielding an initial result of 45 relevant literature items. Subsequently, based on the inclusion criteria, we used WoS's built-in automatic screening tool to exclude 13 literature items that did not meet the requirements. There were two reasons for exclusion: one was that the publication date was prior to 2015 (N=3), and the other was that some search results were literature reviews (N=10) rather than empirical studies. Next, the two authors of this review independently screened the titles and abstracts. For any discrepancies in the screening results, the two authors reached a consensus through negotiation. After this screening stage, the two authors collectively excluded 22 articles that did not meet the criteria. The remaining 10 articles were all accessible for full-text reading directly through WoS. Finally, we carefully reviewed all the articles and found that one article reported results that did not align with the criteria (N=1). Therefore, through the aforementioned screening process, a total of 9 articles met the inclusion and exclusion criteria established for this review. The final screening date for the literature was 24 August 2025. Additionally, in accordance with the recommendations of the PRISMA statement, to clearly demonstrate the literature screening process and its reproducibility, we created a visual literature screening flowchart for this review based on the PRISMA-provided process template, as shown below:

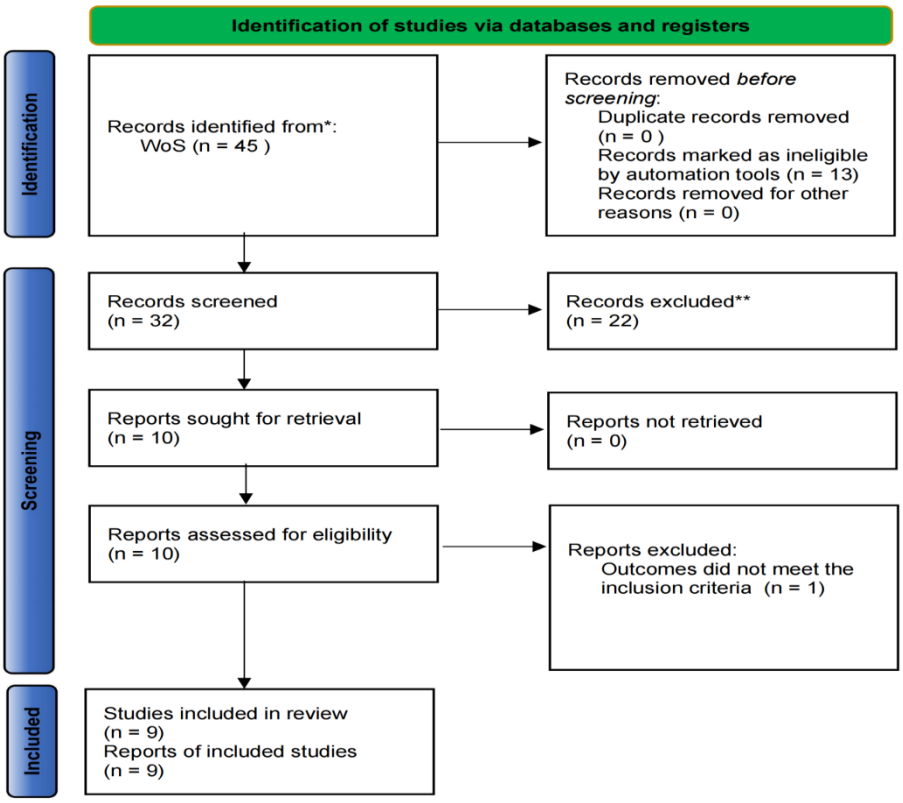

Figure 1. Literature screening flowchart, adapted from Page et al. (2021), licensed under CC BY 4.0

2.5 Coding Schema

Regarding the outcomes or benefits, this study refers to and adapts from the categorization of Xu et al. (2024). Outcomes include language skills and social-emotional skills. Language skills refer to communicative skills, language learning skills, and language use. Social-emotional skills refer to social interaction and stress management in language learning. To avoid including one domain of an article that overlaps with the two domains (Xu et al., 2024), some articles will be categorized under two criteria based on their contents.

To avoid potential bias, this study also followed the methodology of Voultsiou and Moussiades (2025), which adopted and developed a deductive thematic analysis framework. Following line-by-line coding, development of descriptive themes, and articulation of analytical themes, it serves as a methodological guide and a resource for synthesizing research (Voultsiou & Moussiades, 2025).

2.6 Validity

To ensure the validity and reliability of this systematic review, the following strategies were implemented throughout the research process. First, this study strictly adhered to the PRISMA guidelines (Page et al., 2021) to guarantee transparency and reproducibility in literature screening and reporting. Second, the PICO framework (Richardson et al., 1995) was used to clearly define inclusion and exclusion criteria, ensuring only empirical studies directly related to AI or VR interventions for language development in ASD were included. Third, two researchers independently conducted literature screening and data extraction, resolving discrepancies through discussion to minimize subjective selection bias. Finally, all retrieval processes and extracted data were meticulously documented and cross-checked to enhance the reproducibility and methodological transparency of the study. Collectively, these measures bolstered the validity of this systematic review's findings.

3. Results

3.1 Descriptive Analysis

Before detailing the findings of this review, we conducted a descriptive analysis of common basic information from the screened articles using SPSS and Datawrapper. The analysis was based on four dimensions: the region where the research was conducted, the year of publication, the AI, VR, or AR technology employed, and the research methodology. In systematic reviews, descriptive analysis aims to organize and summarize the overall characteristics of included studies, laying the groundwork for subsequent in-depth comparisons and syntheses (Brignardello-Petersen et al., 2024). As noted by Brignardello-Petersen et al. (2024), its core function lies in helping readers rapidly grasp the macro-level landscape of the literature—such as publication timing and regional distribution, sample sizes, research designs and methodological approaches, as well as the diversity of intervention tools and their application contexts. In other words, descriptive analysis serves as a "panoramic view" within systematic reviews, functioning both as a prerequisite for systematic synthesis and a crucial pathway for identifying research gaps (Brignardello-Petersen et al., 2024).

3.1.1 National and regional distribution

From a geographical perspective, the relevant research exhibits a pronounced international trend, with the United States (N=4), Taiwan (N=2), and Pakistan (N=2) contributing a significant proportion of findings. Additionally, a small number of studies have emerged from Greece, Norway, Malaysia, Qatar, and Australia. This distribution pattern indicates that ASD intervention research has gradually transcended single-region boundaries, evolving into collaborative exploration across multiple regions. However, North America and Asia remain the primary focus overall. The specific regional distribution map is shown below:

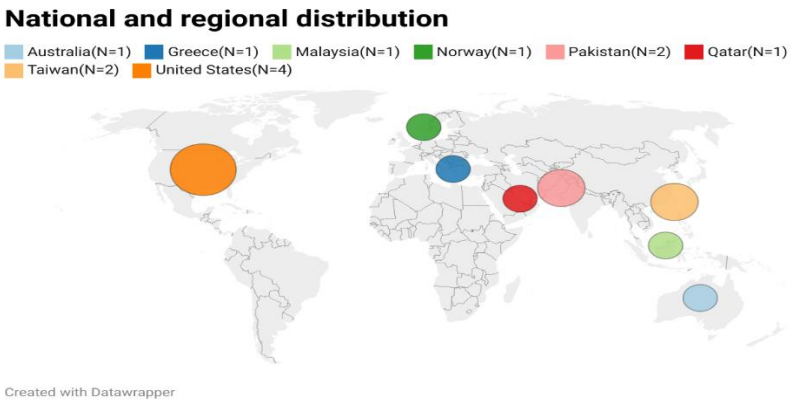

Figure 2. Distribution by Country and Region

3.1.2 Publication Year

From the perspective of publication timing, this field has maintained stable research output since 2017, reaching a relatively steady phase between 2020 and 2022, followed by a significant increase in 2025 when publication numbers peaked. This trend may indicate growing researcher interest in leveraging AI, AR, and VR technologies for ASD interventions amid their rapid advancement. However, given the limited sample size of only 9 included studies, this descriptive analysis warrants cautious interpretation. The specific publication year trend is illustrated below:

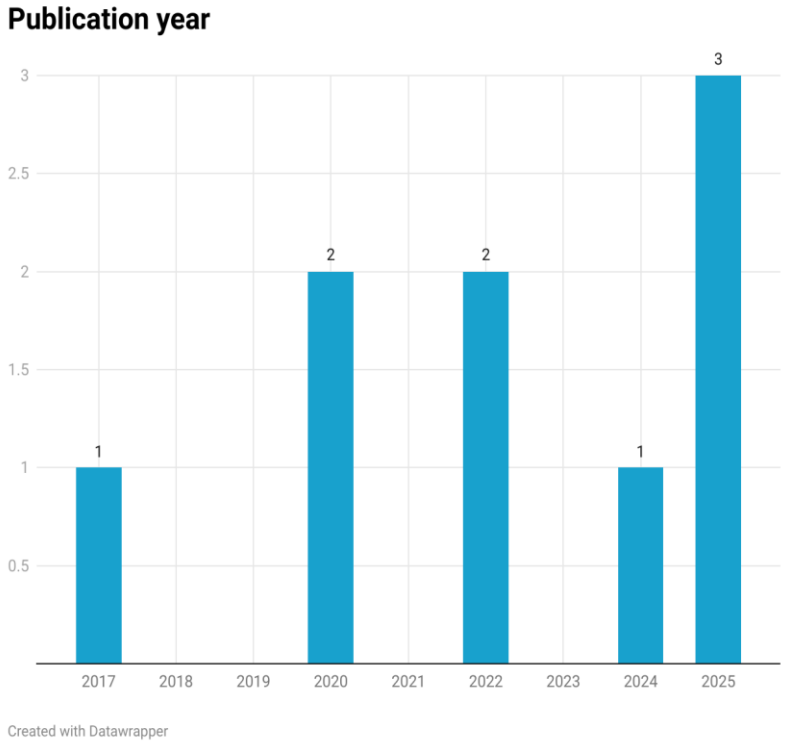

Figure 3. Year of Publication

3.1.3 Research Methods

In terms of research methodology, mixed-methods studies predominate, with five papers employing a combination of quantitative and qualitative approaches, while four studies utilized purely quantitative methods. None of the literature relied solely on qualitative research, indicating that researchers tend to enhance the credibility of their conclusions through quantitative data or multiple lines of evidence. However, this approach may also lead current research to overlook ASD patients' autonomous perceptions of technology or their caregivers' perspectives on emerging technologies. Therefore, future research should appropriately focus on the attitudes of ASD individuals themselves. The specific distribution of research methods is shown below:

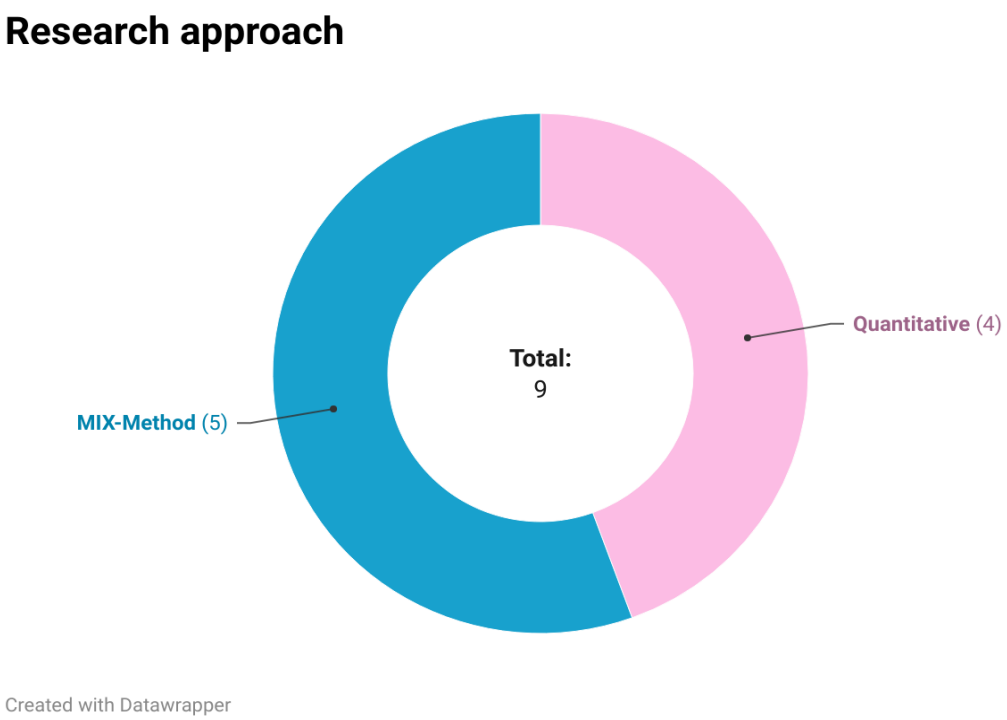

Figure 4. Research Methodology Statistics

3.1.4 Sample Size

The distribution of sample sizes exhibits significant variation, ranging from case studies involving only one participant (Lan, 2020) to large-scale studies with up to 99 participants (Sulek et al., 2022). Most studies, however, remain within the small-scale range of 1–20 participants, indicating persistent challenges in conducting large-scale randomized controlled trials within this field. The specific sample distribution is illustrated as follows:

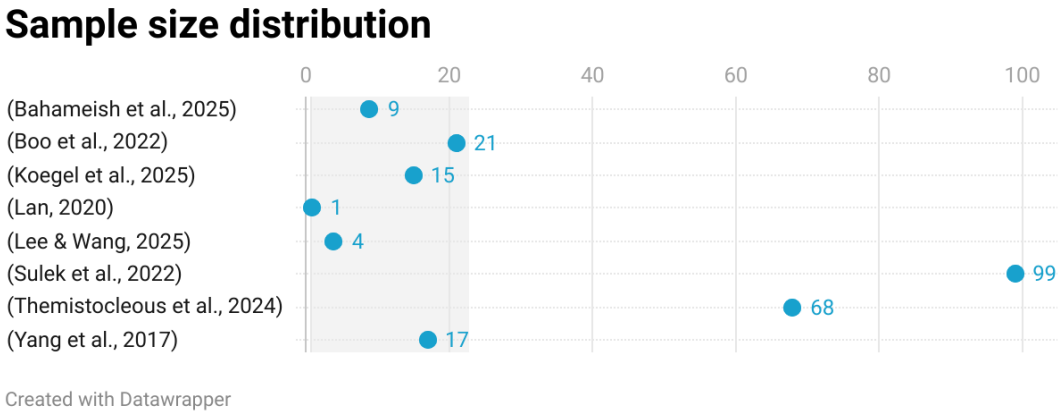

Figure 5. Sample Size Distribution

3.1.5 Age Distribution of the Sample

Regarding the age distribution of samples, the groups covered by different studies span a wide range, with the youngest subjects being only 1–4 years old and the oldest extending to adults aged 35. However, overall, most studies still focus on school-age children, particularly those aged 7–12. This phenomenon indicates that current research prioritizes interventions and training for school-age children with ASD, while studies involving younger children and adults remain limited. Therefore, we recommend that future research should focus more on early intervention for preschool children and studies involving adult populations. Additionally, among the nine included studies, two did not specify the sample age range. We also suggest that future research should establish appropriate sample age ranges, particularly for evidence-based intervention studies. The specific sample age range is illustrated in the figure below:

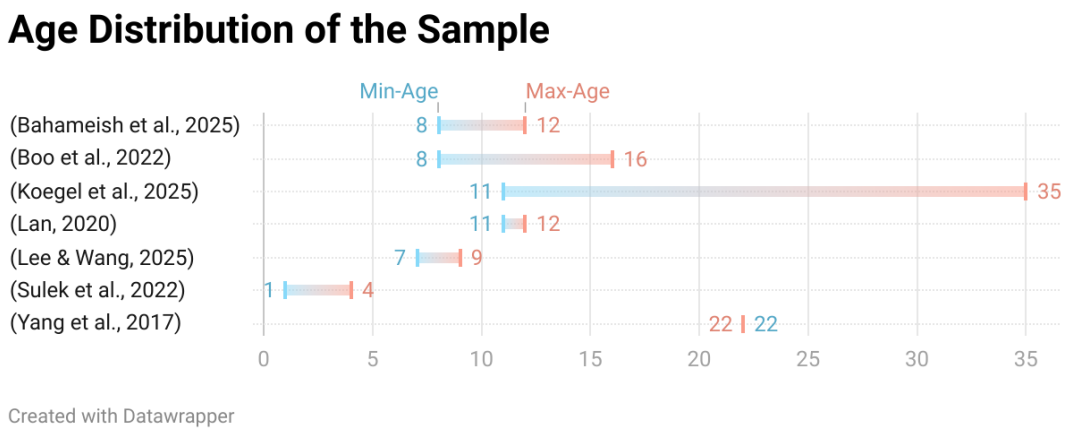

Figure 6. Sample Size Distribution

3.1.6 Research Design Employed in the Selected Literature

Regarding the research designs of each article, we found no overlapping components; each study employed a distinct methodology. Consequently, we do not specify the number of research designs per article, but present them in a table as follows:

Table 2. Research Designs Employed in Each Article

Research Design	
Author (year)	*Design*
(Bahameish et al., 2025)	Single-subject withdrawal design
(Boo et al., 2022)	Experimental comparison
(Khowaja & Salim, 2020)	Framework design & applicability evaluation
(Koegel et al., 2025)	Randomized controlled trial (RCT)
(Lan, 2020)	Case study + small-scale experiments
(Lee & Wang, 2025)	Multiple-baseline design
(Sulek et al., 2022)	Longitudinal evaluation
(Themistocleous et al., 2024)	Comparative study
(Yang et al., 2017)	Pre-post predictive study

3.2 Experiment Procedure Details

This section details the experiment procedure in the intervention design, divided into three approaches: VR, combined VR and AR, and AI. Bahameish et al. (2025) structure the research design into four phases: familiarization, baseline, intervention, and maintenance. Following the introduction of technology for teachers and students in the familiarization stage, the baseline serves as a pre-test of the vocabulary intervention. The intervention consists of two tasks: the word task and the letter task. In the word task, a 3D model generated by VR displays three words. Students can adjust the zoom for detailed inspection, alter viewing angles, virtually grasp the object, and ask questions (e.g., color and usual settings). In the letter task, after playing the audio of a letter, the students select one actual card on the table that corresponds to the audio. Then the cards are scanned, and the letters are displayed on the screens for the learning process, similar to the previous task. The process occurs under the direction of teachers. The maintenance stage is conducted similarly to the baseline stage to test vocabulary retention. Lan (2020) experimented with VR intervention in Chinese as a Foreign Language (CFL) classroom, where students observed virtual objects and collected knowledge for language writing by controlling their avatars under the guidance of teachers. Yang et al. (2017) utilized the Virtual Reality-Social Cognition Training (VR-SCT) in ten sessions during the intervention. In each session, the persona introduced by the clinician with pre-set conversational openings and emotional styles acts as a conversation partner with the participants.

Lee and Wang (2025) designed a combined intervention with AR-TBRP and VR-TBRP games. In both games, participants interact with various characters and identify their missing symbolic capacities, which are demonstrated visually. They need to send symbolic objects to characters as a gift. The assessment tools in the pre-test and post-test include social cognition assessment questionnaires by participants, subjective behavior observation forms by research teams, and parent observation records.

Khowaja and Salim (2020) utilized AI to create and control non-player characters (NPCs), whose actions and reactions are dynamic. The study uses the input, process, and output (IPO) model to conduct the intervention. Specifically, in the input stage, the characteristics of behavior and instruction information are identified, including autistic handicaps in

learning vocabulary, intended learning outcomes, and instructional contents. In the process stage, the game types include action-adventure, a combination of action, adventure, role-playing, simulations, strategy, and sports. The overall story consists of storytelling, narrative, and non-player characters (NPCs). The output stage focuses on recording the user profile. Koegel et al. (2025) claimed that an AI tool can afford personalized practice and mitigate the shortcomings of classroom in-person. In the intervention, participants are asked to rate the sentiment of statements generated by AI and receive real-time gradings according to their responses. Then, they are guided to reply with empathy and receive a sample verified answer. To ensure reliability, researchers randomly selected samples and examined them.

3.3 Outcomes

All studies regarding the intervention reveal different degrees of improvement in language outcomes in vocabulary learning, communicative, and verbal skills. Based on the serious game design framework (SGDF), Khowaja and Salim (2020) designed and verified a proposed framework of AI intervention that improves vocabulary learning for children with ASD. The evaluation results of the serious game prototype show its effectiveness in vocabulary learning. Specifically, the correct responses increase while the number of attempts decreases. Koegel et al. (2025) found that AI can significantly improve empathetic responses for children with ASD. Most participants reported high satisfaction with this intervention in a qualitative study. The reason lies in the self-adjusted pacing ability to provide context that enables autistic users to participate less stressfully. Lan (2020) found that writing outcomes in the VR intervention have better cohesion and organization than those in the traditional classroom. Lee and Wang (2025) found that training strategies incorporating Theatre-Based Role-Playing and augmented reality/virtual reality (AR/VR) can improve verbal expression skills as well as inner psychological state. The advantage of this approach lies in stimulating the theory of mind capabilities. They acquire the ability to infer other people's intentions and emotions, thereby interpreting the non-literal language. The story contexts and visual instruments in the interventions facilitate their understanding of symbolic vocabulary. In a Virtual Reality-Social Cognition Training (VR-SCT), the neural predictors that support language comprehension and process affective information are identified (Yang et al., 2017). Specifically, two brain regions that serve in language comprehension and socio-emotional experience processing are activated.

Notably, Bahameish et al. (2025) found that the results vary between individuals in the interventions. Some exhibits notable improvements in vocabulary learning outcomes, while others present minimal changes. The reasons behind the difference lie in sustaining attention, personalized content, and advantages in interaction with VR. It reveals high variability between individuals in retaining novel words and found a correlation between expressive language abilities and retention outcomes. Additionally, it emphasizes the importance of the ability to generalize acquired vocabulary to new contexts in vocabulary learning. Nevertheless, the results show noticeable improvement in engagement during the learning process.

The other three studies have verified the applications of AI and VR in language diagnosis and assessment. Boo et al. (2022) designed three phases, including a non-social phase, a social phase, and a higher-demand phase. A VR classroom can create more socially and cognitively demanding parts by guiding children to concentrate on virtual avatars (Boo et al., 2022). The results show that the structural language of autistic children is less than that of typically developing (TD) peers. Nevertheless, autistic children who respond with decreased language show similarity to the TD group during the higher-demand phase, indicating that the VR paradigm demands not only social skills but also cognitive load. Sulek et al. (2022) found that the VR metric of Language Environment Analysis (LENA) can predict verbal and non-verbal skill improvement. But it cannot significantly contribute to the prediction when considering baseline skills. The VR correlates with improved expressive and receptive language scores on both parent-report and direct assessment, showing that LENA can complement traditional language assessments. The study found that a Machine Learning (ML) Model using artificial intelligence (AI) can distinguish children with ASD from typically-developing peers by constructing models to analyze their narrative and vocabulary abilities (Themistocleous et al., 2024). The detailed and objective information of autistic children's language, including morphosyntactic features, the usage of pronouns, and lexical characteristics, can be captured and analyzed in future studies.

3.4 Limitations of Studies

There are limitations of research design and technologies in studies of interventions. Themistocleous et al. (2024) focused the study on participants aged 4-10 years. Future studies can select broader samples with various ages and different settings, thereby determining the most accurate ML models. Boo et al. (2022) found the lack of counterbalance in different VR phases and the control group, resulting in ambiguity in factor detection for cognitive load. Furthermore, the experimental questions raised with participants can be the attributing factors of the results. Future studies can be conducted in different contexts. Bahameish et al. (2025) found that VR is incapable of providing practice with dynamic difficulty for children with ASD, thus affecting the learning outcomes. During the evaluation stage of the intervention model, some experts questioned the necessity of an AI-modified character in the game and proposed replacing it with

fixed dialogue (Khowaja & Salim, 2020). Future studies can also investigate the impact of combining new modalities (e.g., virtual reality, augmented reality, and mixed reality) with VR in the acquisition and generalization of the instructional content. Koegel et al.'s (2025) study lacks an in-person intervention control group. Sulek et al. (2022) found a lack of testing reliability of the Language Environment Analysis (LENA) and insights into LENA's validity in the recording context. Future studies can investigate how VR metrics can affect adult responsiveness to speech versus non-speech-like sounds. The role of VR as a moderator of intervention outcomes can contribute to further investigating the utility of VR. The generalization difficulty lies in the study that only conducts experimentation to a certain degree of ASD, thus affecting the VR's effectiveness for broader groups (Lee & Wang, 2025). Yang et al. (2017) conducted a single experimental group that can be optimized for adding a waitlist control group. It can also expand the number of participants to detect a smaller degree of effect in the experiment. Furthermore, it also doubts the adaptability to all degrees of participants (e.g., high-functioning autistic participants). Future studies can delve into the underlying reasons for VR-SCT's effectiveness at the brain level. It can be answered by conducting a comparison of brain activations or tracking of behavioral changes.

Potential moderating factors for children with ASD and unexplored language skills warrant further investigation. Bahameish et al. (2025) found that potential moderating factors for children with ASD may include language proficiency level, cognitive abilities, and unique symptoms, which warrant investigation in future studies. They call for future refinement of intervention strategies and a longitudinal study on the effect of AR and VR technology. Themistocleous et al. (2024) found that other factors (e.g., socioeconomic status and IQ profiles) cannot be neglected because they can refine the applicability in various diagnostic contexts. To ensure high accuracy in diagnosis, future studies can integrate complementary diagnostic tools (e.g., narrative evaluation). Furthermore, other language domains, including pragmatics, receptive language, and language expressions, can also be experimented with in future studies. Koegel et al. (2025) found that the results cannot exclude the roles of trials involved in the experiments, which can act as mediators for outcomes. Khowaja and Salim (2020) found that the details of components, including the identification of participant behaviors and evidence-based instruction methods, in the intervention model can be expanded to develop other language skills. Yang et al. (2017) focus on the basic abilities of socio-emotional and socio-cognitive. Thus, future studies can conduct more measures to test other kinds of abilities.

Notably, AI reveals limitations in its adaptability to different scenarios and the accuracy of generated content. While the study recorded the sample responses for AI model training to ensure their responses better align with interventional aims, it has a slight possibility of mistakenly identifying correct answers rated by humans as incorrect ones (Koegel et al., 2025). Themistocleous et al. (2024) found the requirements for testing the efficacy of AI tools in ASD screening, thereby providing further interventional suggestions for educators.

4. Discussion

4.1 Experimental Procedure

The experimental procedure in these analyzed studies aligns with the principles of Naturalistic Developmental Behavioral Intervention (NDBI) models for ASD that highlight the nature of learning targets, learning contexts, and development-enhancing strategies (Dechsling et al., 2021). By simulating a virtual object and environment related to the real world, the participants can interact naturally with the characters and obtain both language knowledge and common skills (Lan, 2020; Khowaja & Salim, 2020). Incorporating an AR and VR combined intervention in a gamified strategy, participants are guided to initiate interactions with various characters to identify the information for tasks (Lee & Wang, 2025).

Nevertheless, recent studies in computer-assisted language learning demonstrate a lack of synthesis in the theoretical basis of experimental procedure. In the intervention-related papers, fewer studies verified the effectiveness of embedding the technologies into traditional intervention models. Specifically, except for that, Khowaja and Salim (2020) designed an updated serious game design framework (SGDF) that integrated VR technology and was verified by experts, aiming to improve vocabulary learning for children with ASD. Most studies leverage AI and VR to identify ASD issues in language handicaps, conduct interventions in ASD without a basis of theory, and rarely test the effectiveness of AI and VR in language learning. The lack of intervention models or theory supports may lead to doubt about the generalizability and usefulness of the outcomes. Notably, virtual characters play a vital role in interventions by possessing empathy, warmth, and morality, thereby successfully building dependable relationships with participants (Voultsiou & Moussiades, 2025). Participants can interact and practice language with characters embedded in the interventional process (e.g., Lee & Wang, 2025; Yang et al., 2017).

4.2 Outcomes

Despite effective language outcomes in aspects of vocabulary learning and verbal communicative skills in recent studies (e.g., Bahameish et al., 2025; Koegel et al., 2025), maintenance of these language skills is underexplored (Gu et al., 2025). The generalization ability is also lacking in several studies (Bahameish et al., 2025; Khowaja & Salim, 2020). However,

Bahameish et al. (2025) emphasize the importance of the generalization ability of acquired vocabulary to new contexts in vocabulary learning. AI-assisted interventions can mitigate the shortcomings of in-person interventions by providing more practice for the generalization of language ability (Koegel et al., 2025). Specifically, AI such as conversational agents, VR avatars, and social assistive robots (SARs) can provide support for autistic individuals through consistent interactions (Sohn et al., 2025). Future studies can investigate the maintenance and generalization ability of lexical knowledge, verbal skills, and other unexplored skills (e.g., writing) by conducting longitudinal studies. Notably, autistic children who respond with decreased language show similarity to the typically developing (TD) group during the higher-demand phase, indicating that the VR paradigm demands not only social skills but also a high cognitive load (Boo et al., 2022).

The effectiveness of learning outcomes lies in the baseline of participants. It reveals high variability between individuals in retaining novel words and found a correlation between expressive language abilities and retention outcomes (Bahameish et al., 2025). Similarly, some participants need to improve similar outcomes by participating in additional weeks of usage (Koegel et al., 2025). The heterogeneity of outcomes necessitates an AI that provides individualized and dynamic practice for students to adapt to different baselines of participants in the interventions (Koegel et al., 2025; Sohn et al., 2025).

4.3 Opportunities and Challenges

The contradiction lies in the coexistence of both opportunities and challenges. AI can play the roles of companions, co-learners, and virtual assistants (VAs) (Voultsiou & Moussiades, 2025). Lan (2020) found that VR can enhance the motivation and autonomy of students with ASD while studying languages. The benefits of AI lie in interventions and predictions for autistic populations by leveraging contextual understanding and interdisciplinary integration, thereby overcoming limited clinical resources (Garg & Chauhan, 2024; Koegel et al., 2025).

However, the debate on the role and challenges of AI and VR persists in studies of different autistic populations. The limitations of recent studies lie in the adaptation for various scenarios due to the immature technology, which weakens the reliability and generalizability of the findings. This aligns with Gu et al.'s (2025) study, which found that limited datasets and a shortage of personalization restrict the reliability of AI suitable for the unique needs of individuals with ASD. Specifically, AI cannot accurately capture and predict the complex severities and individual differences for ASD (Garg & Chauhan, 2024). While hallucinations of AI cannot be neglected in the study, especially in the role of ASD diagnosis and intervention (Koegel et al., 2025; Sohn et al., 2025), the errors can potentially lead to risks of misleading clinical judgments and delayed treatments.

Furthermore, the resource constraint can limit their usage and broad applications (Lee & Wang, 2025). Challenges such as resource limitations are highly mentioned in the previous studies (e.g., Voultsiou & Moussiades, 2025). Lan (2020) claimed that a VR classroom requires teacher training that enables teachers to familiarize themselves with technology usage. The challenges also lie in the complexity of intervention training and assessment tools that cannot directly validate the interventions (Lee & Wang, 2025). In summary, AI applications for autistic individuals are in the early proof-of-concept stages, thereby warranting further refinement in research for effective implementation (Sohn et al., 2025).

5. Implications

By synthesizing recent studies of applications of AI and VR technologies in language teaching for ASD, the review identifies key areas leveraging the technologies to facilitate future language research and instruction support.

5.1 Development of Models and Long-term Experiment

The interventions should be designed and validated in a reliable, effective, and safe way (Sohn et al., 2025). Therefore, future studies should synthesize intervention models and language acquisition models, then consider how to integrate them in AI and VR. The important elements essential for designing serious games are missing in recent studies (Khowaja & Salim, 2020). To better structure the methodology, the call for rigorous procedure, including top-down metatheory, bottom-up guided explorations, top-down theory/hypothesis formation, and bottom-up empirical testing, also provides future studies with references (van Dijk et al., 2024). This acknowledges that while individual cases are crucial, they are not studied separately in isolation. In summary, future studies should prioritize the development of interventions that apply theoretical research into achievable actions, embedding the various technological affordances (e.g., tailored prompting tools and applications) (Voultsiou & Moussiades, 2025). Furthermore, future studies are warranted to investigate the long-term effects and the effectiveness for participants with all clinical levels and ages (Bahameish et al., 2025; Yang et al., 2017).

5.2 Multimodal and Interdisciplinary Study

A technology-assisted language intervention can integrate cues from vocal tone, facial expressions, eye movement, and biological data to capture participants' behaviors, thus providing individualized guidance and feedback (Sohn et al., 2025). Integrated AI and VR technologies, the systems are capable of seamlessly collecting, analyzing, and providing

information to participants. Specifically, using VR to stimulate a flexible real-world learning environment is compensated with AI providing personalized guidance and assistance (Voultsiou & Moussiades, 2025). To mitigate the bias in the limited applicability of systems, multimodal databases potentially facilitate equitable studies applicable to various cultural backgrounds around the global research community.

While multimodal AI can potentially simulate interdisciplinary human clinical teams by synthesizing a variety of data to support nuanced diagnoses and real-time support in personalized interventions, the necessity of interdisciplinary cooperation between technology and human therapists remains prevalent. The VR plays a role as a moderator of intervention outcomes but does not exert a determining impact (Sulek et al., 2022). For instance, Koegel et al. (2025) conduct further refinement of the feedback generated by AI by providing suggestions for sampled responses before the experiment. This highlights the importance of human participation before the intervention phases. Using multimodal is conducive to enhancing the quality of interactions (Voultsiou & Moussiades, 2025). To remain ethical, future studies can also integrate stakeholder or expert feedback with empirical evidence to comprehensively evaluate the tools, moving beyond the mere validation of effectiveness (Voultsiou & Moussiades, 2025).

6. Limitations and Conclusion

Through analyzing the outcomes of technology and study limitations in language learning for children with ASD, this study conducts a systematic review of the potential of AI and VR applications in language learning in empirical studies. Our analysis revealed the prevalence of effectiveness in the intervention stage as well as debates regarding the ethical risk. This research contributes to a focused and contemporary synthesis in the critical examination of applications of VR and AI technologies in language development for ASD, advocating for a more rigorous and integrated framework to guide the applications. By highlighting the opportunities and challenges in recent studies, this review offers recommendations for future synthesized experimental models and interdisciplinary studies, aiming to foster a more unified and nuanced understanding of AI and VR in autistic language learning.

Two limitations exist in the present review. The study includes only nine articles from SSCI journals in English, ensuring their high quality and wide citations. Nevertheless, the limited number of articles from the SSCI overlooks valuable contributions from other sources, potentially leading to one-sidedness of results. By incorporating more resources, such as journal articles indexed by Scopus, Eric, etc (Zhang & Zou, 2022), future studies can synthesize a broader range of application approaches and empirical outcomes. This thereby strengthens the overall comprehensiveness and objectivity of the analysis. Moreover, the review collected studies on language acquisition rather than communication or social skills. According to the exclusion criteria, most excluded studies on communication skills may involve some language learning elements, which results in the potential omission of relevant research.

Acknowledgments

We greatly appreciate the valuable contributions of all those who provided insights and support during the preparation of this study. The authors would like to express sincere gratitude to Dr. Wenjun Xie from Sun Yat-sen University and Zixiang Wei from the University of New South Wales for their encouragement and advice that assisted in developing the initial ideas and framework of this study. The authors also appreciate the insightful comments by Dr. Kathryn Sidaway from the University of Bedfordshire and Prof. Sithembinkosi Dube from the University of New South Wales throughout the research process.

Authors contributions

Conceptualization: Fangye Luo, Methodology: Yiding Lu, Fangye Luo, Data collection and analysis: Fangye Luo, Yiding Lu, Writing-original draft: Fangye Luo, Yiding Lu, Writing-review & editing: Fangye Luo, Yiding Lu.

Supervision: Fangye Luo

Funding

This article received no funding support.

Competing interests

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Informed consent

Obtained.

Ethics approval

The Publication Ethics Committee of the Redfame Publishing.

The journal's policies adhere to the Core Practices established by the Committee on Publication Ethics (COPE).

Provenance and peer review

Not commissioned; externally double-blind peer reviewed.

Data availability statement

The data that support the findings of this study are available on request from the corresponding author. The data are not publicly available due to privacy or ethical restrictions.

Data sharing statement

No additional data are available.

References

Aresti-Bartolome, N., & Garcia-Zapirain, B. (2014). Technologies as Support Tools for Persons with Autistic Spectrum Disorder: A Systematic Review. *International Journal of Environmental Research and Public Health*, *11*(8), 7767-7802. https://doi.org/10.3390/ijerph110807767

Bahameish, M., Khowaja, K., Abdelaal, Y., & Al-Thani, D. (2025). Pathways to learning: exploring the impact of augmented reality on vocabulary development in children with autism spectrum disorder. *Interactive Learning Environments*, 1-24. https://doi.org/10.1080/10494820.2025.2485407

Birkle, C., Pendlebury, D. A., Schnell, J., & Adams, J. (2020). Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies*, *1*(1), 363-376. https://doi.org/10.1162/qss_a_00018

Boo, C., Alpers-Leon, N., McIntyre, N., Mundy, P., & Naigles, L. (2022). Conversation During a Virtual Reality Task Reveals New Structural Language Profiles of Children with ASD, ADHD, and Comorbid Symptoms of Both. *Journal of Autism and Developmental Disorders*, *52*(7). https://doi.org/10.1007/s10803-021-05175-6

Brennan, S. E., & Munn, Z. (2021). PRISMA 2020: a reporting guideline for the next generation of systematic reviews. *JBI Evidence Synthesis*, *19*(5), 906-908. https://doi.org/10.11124/jbies-21-00112

Brignardello-Petersen, R., Santesso, N., & Guyatt, G. H. (2024). Systematic reviews of the literature: An introduction to current methods. *American Journal of Epidemiology*, *194*(2). https://doi.org/10.1093/aje/kwae232

Dechsling, A., Shic, F., Zhang, D., Marschik, P. B., Esposito, G., Orm, S., ... & Nordahl-Hansen, A. (2021). Virtual reality and naturalistic developmental behavioral interventions for children with autism spectrum disorder. *Research in Developmental Disabilities, 111*, 103885. https://doi.org/10.1016/j.ridd.2021.103885

Fuentes, A. (2022). Reseña de sitio web: Preferred Reporting Items for Systematic Reviews and Meta-Analyses (PRISMA). Declaración PRISMA 2020. *Revista de Estudios E Investigación En Psicología Y Educación*, *9*(2), 323-327. https://doi.org/10.17979/reipe.2022.9.2.9368

Garg, S., & Chauhan, A. (2024). Insights into the potential benefits and challenges of AI-driven large language models/ChatGPT-4 for predicting Autism Spectrum Disorder. *Archives of Psychiatric Nursing*, *53*, 85-87. https://doi.org/10.1016/j.apnu.2024.10.009

Gu, P., Xu, F., Chen, L., Ma, Z., Zhang, M., & Zhang, Y. (2025). Technological affordances and applications of chatbots for conversational skill interventions in autism: A scoping review. *Education and Information Technologies*, *30*(7). https://doi.org/10.1007/s10639-024-13191-z

Hashim, H. U., Yunus, M. M., & Norman, H. (2022). "AReal-Vocab": An Augmented Reality English Vocabulary Mobile Application to Cater to Mild Autism Children in Response towards Sustainable Education for Children with Disabilities. *Sustainability*, *14*(8), 4831. https://doi.org/10.3390/su14084831

Khowaja, K., & Salim, S. S. (2020). A framework to design vocabulary-based serious games for children with autism spectrum disorder (ASD). *Universal Access in the Information Society*, *19*(4). https://doi.org/10.1007/s10209-019-00689-4

Koegel, L. K., Ponder, E., Bruzzese, T., Wang, M., Semnani, S. J., Chi, N., ... & Lam, M. S. (2025). Using Artificial Intelligence to Improve Empathetic Statements in Autistic Adolescents and Adults: A Randomized Clinical Trial. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-025-06734-x

Koegel, R. L., & Koegel, L. K. (2006). *Pivotal response treatments for autism : communication, social & academic development*. Paul H. Brookes.

Lan, Y. J. (2020). Immersion into virtual reality for language learning. *Psychology of Learning and Motivation*, *72*, 1-26. https://doi.org/10.1016/bs.plm.2020.03.001

Lee, I. J., & Wang, W. Z. (2025). Applying theatre-based role-playing combined with AR and VR game strategies to enhance imagination and symbolic play skills in children with autism spectrum disorder. *Universal Access in the Information Society*, 1-22. https://doi.org/10.1007/s10209-025-01206-6

Li, K., Rollins, J., & Yan, E. (2018). Web of Science use in published research and review papers 1997–2017: a selective, dynamic, cross-domain, content-based analysis. *Scientometrics*, *115*(1), 1-20. https://doi.org/10.1007/s11192-017-2622-5

Liberati, A. (2019). The PRISMA Statement for Reporting Systematic Reviews and Meta-Analyses of Studies That Evaluate Health Care Interventions: Explanation and Elaboration. *Annals of Internal Medicine*, *151*(4). https://doi.org/10.7326/0003-4819-151-4-200908180-00136

Lima, M. P. P., Carneiro, M. O. V., Mendonça, É. M., Oliveira, C. B. G. C. de, Benevides, S. B., Forte, B. L. ... & Neto, J. A. de A. (2023). Language acquisition in Autism Spectrum Disorder: Strategies for early diagnosis and intervention – A review of literature. *Research, Society and Development*, *12*(12), e106121243837–e106121243837. https://doi.org/10.33448/rsd-v12i12.43837

Loukusa, S., & Moilanen, I. (2009). Pragmatic inference abilities in individuals with Asperger syndrome or high-functioning autism. A review. *Research in Autism Spectrum Disorders*, *3*(4), 890-904. https://doi.org/10.1016/j.rasd.2009.05.002

Marijn van Dijk, Lowie, W., Smit, N., Marjolijn, V., & Paul, van G. (2024). Complex dynamic systems theory as a foundation for process-oriented research on second language development. *Second Language Research*. https://doi.org/10.1177/02676583241246739

Nishikawa-Pacher, A. (2022). Research Questions with PICO: A Universal Mnemonic. *Publications*, *10*(3), 21. MDPI. https://doi.org/10.3390/publications10030021

Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., ... & McGuinness, L. A. (2021). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *British Medical Journal*, *372*(71). https://doi.org/10.1136/bmj.n71

Richardson, W. S., Wilson, M. C., Nishikawa, J., & Hayward, R. S. A. (1995). The well-built Clinical Question: a Key to Evidence-based Decisions. *ACP Journal Club*, *123*(3). https://doi.org/10.7326/acpjc-1995-123-3-a12

Schiavenato, M., & Chu, F. (2021). PICO: What It Is and What It Is Not. *Nurse Education in Practice*, *56*(1). https://www.sciencedirect.com/science/article/pii/S1471595321002304

Sohn, J. S., Lee, E., Kim, J. J., Oh, H. K., & Kim, E. (2025). Implementation of generative AI for the assessment and treatment of autism spectrum disorders: a scoping review. *Frontiers in Psychiatry*, *16*, 1628216. https://doi.org/10.3389/fpsyt.2025.1628216

Sulek, R., Smith, J., Bent, C. A., Hudry, K., Trembath, D., Vivanti, G., & Dissanayake, C. (2022). The utility of LENA as an indicator of developmental outcomes for young children with autism. *International Journal of Language & Communication Disorders*, *57*(1), 103-111. https://doi.org/10.1111/1460-6984.12678

Themistocleous, C. K., Andreou, M., & Peristeri, E. (2024). Autism Detection in Children: Integrating Machine Learning and Natural Language Processing in Narrative Analysis. *Behavioral Sciences*, *14*(6), 459. https://doi.org/10.3390/bs14060459

Voultsiou, E., & Moussiades, L. (2025). A systematic review of AI, VR, and LLM applications in special education: Opportunities, challenges, and future directions. *Education and Information Technologies*. https://doi.org/10.1007/s10639-025-13550-4

Xu, F., Gage, N., Zeng, S., Zhang, M., Iun, A., O'Riordan, M., & Kim, E. (2024). The Use of Digital Interventions for Children and Adolescents with Autism Spectrum Disorder—A Meta-Analysis. *Journal of Autism and Developmental Disorders*, 1-17. https://doi.org/10.1007/s10803-024-06563-4

Yang, Y. J. D., Allen, T., Abdullahi, S. M., Pelphrey, K. A., Volkmar, F. R., & Chapman, S. B. (2017). Brain responses to biological motion predict treatment outcome in young adults with autism receiving Virtual Reality Social Cognition Training: Preliminary findings. *Behaviour Research and Therapy*, *93*, 55-66. https://doi.org/10.1016/j.brat.2017.03.014

Yu, Z. (2025). Promoting Language Perception in Children with ASD Through Targeted Teaching Methods. *Communications in Humanities Research*, *59*(1), 41-45. https://doi.org/10.54254/2753-7064/2024.22508

Zhang, R., & Zou, D. (2022). Types, features, and effectiveness of technologies in collaborative writing for second language learning. *Computer Assisted Language Learning*, *35*(9), 1-31. https://doi.org/10.1080/09588221.2021.1880441