Artificial Intelligence Techniques and Best Practices to Improve Motivation and Learning

Maggie Mosher, Amber Rowland, Bruce Frey, Sean Smith
University of Kansas
Lawrence, Kansas
mosherku@ku.edu

Adam Carreon
Georgia Southern
Statesboro, Georgia
acarreon@georgiasouthern.edu

Tolulope Sulaimon
University of Central Florida
Orlando, Florida
tolulope.sulaimon@ucf.edu

ABSTRACT

Artificial Intelligence (AI), over the past decade, has improved substantially in its ability to mimic regular conversation and provide more accurate responses (Labadze et al., 2023). AI applications, such as ChatGPT, have recently developed the capacity to reject requests that are inappropriate or beyond the application's reach, challenge incorrect assumptions, and admit mistakes (Makridakis et al., 2023). Fryer and colleagues' (2019) even found that college students interacting with Chatbots reported greater curiosity and perseverance than those interacting similarly with peers. Despite these advancements, there are legitimate concerns over the use of AI, especially within education. Responses from AI apps continue to be unnatural, misleading, inaccurate, vague, and biased (Sallam et al., 2023). Research shows a decline in the quality of information if educators become over-reliant on AI (Chiu et al., 2023). Still, the capacity for AI to assist educators and motivate students has been displayed (Sanusi et al., 2022).

This paper discusses a mixed methods study on a randomized control trial evaluating teacher led versus AI facilitated student instruction. Possible solutions from this study include: (a) providing engineered detailed prompts within AI, (b) pulling data from a knowledgeable database by trusted experts, (c) providing specific questions within the prompt of response type warranted, (d) creating a clear pedagogical focus, and (e) using AI as one source within many sources to gather outside perspectives rather than replace human data collection. It will also discuss prompting questions to improve AI accuracy, such as: "Was the data the AI was trained on representative of the population affected by subsequent decisions?" and "Is the information given well-balanced and derived from an analysis of information from trustworthy sources?"

ABOUT THE AUTHORS

Maggie Mosher, PhD is an assistant research professor at the University of Kansas Achievement and Assessment Institute where she specializes in social skills, behavioral development, innovative technology, multisensory methods, and individualized instruction. Mosher's research focuses on the development of extended reality and AI for assessing, progress monitoring, and intervening in social-emotional-behavioral competencies for students with high-incidence disabilities.

Adam Carreon, PhD is an assistant professor of special education at Georgia Southern University, where he specializes in teaching, instructional design, technology, and innovation. Carreon's research interests include the use of effective instructional and behavioral strategies, specifically emerging technologies (e.g., augmented reality, virtual reality, wearables) for improving academic, adaptive, and behavioral outcomes for teachers of and students with disabilities.

Tolulope Sulaimon, PhD is a research associate at the University of Central Florida. He is also affiliated with the University of Kansas. Tolulope's research focuses on comprehension, word problems, and integration of technology to enhance evidence-based interventions for students with disabilities.

Sean J. Smith, PhD is a professor of special education in the Department of Special Education at the University of Kansas. Dr. Smith's research centers the role of technologies (e.g., virtual reality, extended reality, artificial intelligence) in advancing the overall growth and development (e.g., instruction, social emotional, adaptive) of knowledge and skills to promote further inclusion of struggling learners and their peers with identified with disabilities and subsequent positives outcomes.

Amber Rowland, PhD is an associate research professor with the Center for Research on Learning and the Achievement and Assessment Institute at the University of Kansas. Dr. Rowland's research centers on supporting educator professional learning at a distance and pursuing methods for integrating innovative technology to support all students, especially those with disabilities. Current work includes extended reality and AI with social skill learning, and progress monitoring and AI in writing.

Bruce Frey, PhD is a professor in the department of Educational Psychology. His areas of research include classroom assessment, instrument development and program evaluation. Dr. Frey is the author of several books, including Statistics Hacks and Modern Classroom Assessment.

Artificial Intelligence Techniques and Best Practices to Improve Motivation and Learning

Maggie Mosher, Amber Rowland, Bruce Frey, Sean Smith
University of Kansas
Lawrence, Kansas
mosherku@ku.edu

Adam Carreon
Georgia Southern
Statesboro, Georgia
acarreon@georgiasouthern.edu

Tolulope Sulaimon
University of Central Florida
Orlando, Florida
tolulope.sulaimon@ucf.edu

INTRODUCTION

As artificial intelligence (AI) continues to evolve and expand, its potential applications in education, particularly within K-12 classrooms, have become increasingly evident. The release of user-accessible generative artificial intelligence (AI) tools (e.g., ChatGPT, Gemini, and CoPilot) has caused a shift in how educational planning and education technology tools are utilized. While only 18% of K-12 teachers reported directly using AI for teaching in 2023, approximately 60% of school districts planned on training teachers about specific AI use by the end of the 2023-2024 school year (Diliberti et al., 2024). With this shift in technology, it becomes imperative that school stakeholders understand the importance of AI in education and the potential impact it can have on teacher and student success in the classroom.

AI can provide teachers with tools to enhance student learning, automate tasks, and provide personalized instruction tailored to individual student needs (Marino et al., 2023). While AI is not new, the practical application and accessibility of this technology in education have grown exponentially in the last two years with the proliferation of generative AI tools (Mello et al., 2023). AI is an overarching term for a range of technologies that aim to mimic human intelligence and behavior, including machine learning, natural language processing, and computer vision. Generative AI has significantly broadened the scope of AI's utility in educational settings, from leveled texts, automated grading, and writing assistance to personalized tutoring and adaptive learning environments. Generative AI can produce original text, images, audio, and other forms of media in response to use prompts (Fulton et al., 2022). With generative AI, educators can enhance student learning outcomes, personalize the educational experience, and address educational inequalities. The following sections will outline how AI can and is being used to enhance the school environment.

Enhancing Student Learning Outcomes

Generative AI can be leveraged to create customized educational content and is revolutionizing education by significantly enhancing student learning outcomes across various domains (Mello et al., 2023). AI innately learns by adapting its behavior from existing data. This skill can be applied to education through applications that include adaptive learning. Adaptive learning technologies are pivotal in personalizing the educational experience and adaptive learning systems learn from student behavior, plan for the best learning path, and push students to complete personalized assignments and assessments that adjust based on performance. For instance, a study by Grams (2018) found that students using the adaptive learning tool DreamBox showed significant gains in mathematics proficiency from pretest to post-test. While some variables (time spent and external resources) were not accounted for, this research underscores the potential of adaptive learning technologies in fostering deeper learning and achievement among students.

Like adaptive learning systems, intelligent tutoring systems represent another common occurrence of AI in education. These systems utilize AI to analyze responses, understand personal gaps in learning, and deliver interventions that have been personalized for students. For example, tutoring systems like Tutor AI (e.g., https://www.tutorai.me) can provide personalized tutoring for students to learn at their own pace by answering questions and walking users through examples, explanations, formulas, and other academic information. Research highlights the effectiveness of intelligent tutoring systems in improving student understanding and retention of complex concepts (Lopes et al., 2019).

Addressing Educational Gaps with AI

Research indicates that AI-assisted learning platforms can effectively supplement traditional teaching methods (Topsakal & Topsakal, 2022), providing personalized educational experiences that adapt to individual learning capacities. AI interventions can demonstrate improved student engagement, retention rates, and academic performance (Hasan & Khan, 2023). Furthermore, research highlights AI's role in supporting differentiated instruction by providing adaptive learning pathways and personalized interventions that cater to varying learning styles and paces (Marino et al., 2023). These AI-driven approaches have been shown to increase student engagement, promote inclusive learning environments, and lead to enhanced learning outcomes for the diverse learner needs in the modern classroom (Shoaib et al., 2024).

While still a novice area of research, current special education research highlights the efficacy of AI tools in fostering inclusive learning environments and improving educational outcomes for students with disabilities (Sen & Akbay, 2023). Research on using communication boards that harness AI show further support for speech recognition, speech prediction, and contextual knowledge (Evangeline & Moorthy, 2024). Current research demonstrates the potential positive impact of AI in addressing the unique learning challenges faced by students with disabilities, promoting greater accessibility and inclusivity within educational settings (Marino et al., 2023).

Engineered Detailed Prompts within MagicSchool AI

MagicSchool's (2024) chatbot Merlin plays a crucial role in assisting middle school students in generating script ideas, dialogue, and narratives to improve their social skills within this AI tool. Leveraging natural language processing and generation capabilities, Merlin swiftly produces well-structured social narratives, aiding students in their storytelling endeavors. MagicSchool AI stemmed from a collaborative effort among educators, technologists, and AI experts aiming to transform the educational landscape for students and teachers globally. The platform was envisioned as a comprehensive toolset harnessing artificial intelligence to create personalized and engaging learning resources. Designed to offer tailored assistance and resources, the chatbots adapt to the specific needs of users at different educational levels, providing support for learning, teaching, and creativity across subjects and disciplines. MagicSchool AI offers a great opportunity that adapts content to the needs of different students. The specific prompts created to guide interaction with the chatbot have been critical to its success. These prompts are the primary means by which users connect with ChatGPT, allowing the model to create responses that are relevant to the user's goal. These are not merely scripted lines, but well-crafted dialogues that employ natural language processing (NLP) and machine learning to provide a personalized and engaging learning experience (Giray, 2023).

The engineered prompts are carefully constructed using educational philosophies that emphasize active learning and student interaction. Clark & Mayer (2023) asserted that learning is more effective when students are actively participating, a principle that detailed prompts can easily support. Prompts and scenarios based on educational research with imaginative content with MagicSchool AI are combined to provide a dynamic and useful learning experience. The platform employs a knowledge base derived from credible sources in both magical lore and educational research to provide adaptive prompts that respond meaningfully and educationally to user inputs. These prompts help students move beyond simple memorization by asking thought-provoking questions, such as how historical moments during the Greensboro sit-in contributed to the desegregation in America and what lessons can be applied today to promote equality, thereby fostering critical thinking and deeper cognitive processes. MagicSchool AI also promotes the development of students' knowledge with its well-constructed prompts as they connect material from diverse sources, such as when a question asks, "How did different events during the Civil Rights Movement contribute to changes in American laws?" This approach helps students become knowledge constructors, allowing them to synthesize information and obtain a better comprehension of historical events. MagicSchool AI also encourages problem-solving abilities and computational thinking through prompts that demand logical reasoning and pattern detection. For example, a prompt could instruct students to "create a timeline of major Civil Rights events and analyze the patterns of social change over time." This type of task helps students improve their ability to arrange and analyze data systematically.

Evidence-Based Practice (EBP) Social Narratives

Currently, there are several evidence-based practices (EBPs) for teaching social skills. The EBP chosen in this study fits well with technology integration and requires minimal training prior to effective implementation: social narratives (Wong, et al., 2015). Social narratives appear to be particularly beneficial when introducing and teaching a variety of skills to individuals with autism, such as expressive communication (Charlop & Milstein, 1989), emotion processing

(Corbett, 2003), perspective-taking (LeBlanc et al., 2003), and play skills (D'Ateno et al., 2003). Gresham (2017) has also documented a long history of success in using social narratives to improve social communication skills, especially for students with autism. Traditionally, social narratives are presented by an adult who reads the story to the student. Today, technology has increased the delivery options for such narratives, making narratives available to students independent of adults. Software programs, applications, simulations, augmented reality, and virtual environments are used to present narratives to students. Social narratives remain an evidence-based practice whether they are read to children or presented in VR programs (Ghanouni et al., 2019). However, there is a need for additional research on the effectiveness of social skill instruction through innovative technology methods (Bernardini, 2014).

Sutherland and colleagues (2019) identified 22 intervention elements associated with increased social competencies for students with disabilities. Of these elements, the following effective practices can be implemented within the delivery of content and conversation with an AI chatbot focused on valuable and reliable sources: choice, instructional antecedent manipulation, opportunities to respond, praise, precorrection, reinforcement, rewards, routines, and self-management. These elements presented through AI are particularly important because they are feasible for implementation in general education, cost-effective, create an environment in which behavioral and academic instruction occur simultaneously, increase the likelihood of engagement, prevent the occurrence of social challenges, and require little specialized training.

Investigating the attitudes towards the use of AI as a learning aid to assist in writing social narratives and answering social questions based on student prompting is important as schools continue to add additional AI into the classroom. This study focused solely on student attitudes. Such information is useful in supporting current social skill practices within middle schools as well as in shaping future social skill interventions to improve the attitudes towards these interventions for this population. Therefore, the following research question was developed to gather valid and reliable data that may assist educators and curriculum developers in making decisions on adopting AI as a prompting tool for social narrative creation:

1. Is there a difference in students' attitudes about creating and reading social narratives on their technology device (i.e., iPad, Chromebook) when they generated the narratives in collaboration with an AI chatbot versus in collaboration with their social skill educator?

METHODS

Researchers, educators, and developers all point to a need for quality research on AI's potential within education (Chiu, 2023). AI can free up valuable time by automating daily tasks and personalizing content based on prior searches and saved resources, allowing students and teachers to focus on building relationships and fostering learning (Hashem et al., 2024).

Participants

Approval was received by the Office of Research Institutional Review Board (IRB) prior to all research activity. Participants in this study included middle school students and their primary educators. A call went out for participants at multiple conferences. Emails were also sent in March 2022 to statewide organizations that provide services to those with social communication delays. Inclusionary criteria were applied to 152 students, resulting in 109 student participants ages 10 to 13 years (M= 11) and 10 educators. Student participants' inclusionary criteria consisted of: 1. being of middle school age (i.e., age 10 to 15); 2. being identified by a qualified educator or practitioner familiar with the adolescent to be in need of a pragmatic or expressive language social narrative intervention determined by a valid and reliable assessment measure (e.g. Clinical Assessment of Pragmatics, Test of Pragmatic Language); 3. being able to complete rating scales on their perceptions; 4. being willing to participate for the duration of the study and follow up; 5. being able to participate in a technology-based intervention; 6. having an educator willing to be trained and complete rating scales; and 7. having the communication abilities to participate in the interventions (i.e., English-speaking with a minimum third-grade reading level). A participant's diagnosis, if any, was not a prerequisite for participation, though it was documented. Students were recruited to ensure the main qualifier for participation is a deficit in social skill as determined by a qualified school professional (e.g., speech and language pathologist, special education teacher, occupational therapist, school psychologist). Parent consent and student assent were obtained for 101 students. However, one student's family situation changed prior to training, and was no longer able to participate for the duration of the study. Table 1 provides a summary of the demographics for 100 student participants.

Table 1: Student Demographic Information

Race and Ethnicity	Group 1		Group 2			Group 1		Group 2	
	n	%	*n*	%	**Student Age**	*n*	%	*n*	%
African American	1	2	1	2	10 years old	22	44	18	12
American Indian/Alaska Native	5	10	3	10	11 years old	17	34	17	34
Asian	3	6	1	2	12 years old	6	12	8	16
Hispanic/Latino	3	6	4	8	13 years old	5	10	7	14
White	38	76	41	82					
Gender	*n*	%	*n*	%		*n*	%	*n*	%
Female	36	72	26	52	Male	14	28	24	48

Note. $N = 100$ students ($n = 50$ per group)

Upon comparing Group 1 and Group 2 within the entire group of 100 students, we can observe some similarities and differences. Both groups have a similar total number of students (50 each) and a comparable distribution of male and female students. Both groups have students across a range of ages from 10 to 13 years old, with diagnoses (see Table 2) such as Autism Level 1, ADHD, and Depression, and education plans like IEPs and Social Skill groups

Overall, while there are some differences in racial distribution, diagnoses, and specific education plans between the two groups, they were fairly equal in terms of total students, gender distribution, age range, and the presence of a common diagnoses and education plans. Each group contributes unique characteristics to the overall demographic composition of the 100 students.

Table 2: Participant Diagnosis and School Plan

Diagnosed Disability	Group 1		Group 2	
	n	%	*n*	%
Attention Deficit Hyperactivity Disorder (ADHD)	18	36	16	32
Autism Spectrum Disorder (ASD)	29	58	18	36
Obsessive Compulsive Disorder (OCD)	2	4	1	2
Depression	2	4	2	4
Learning Disabilities (LD)	9	18	3	6
Anxiety	1	2	0	0
Intellectual Disability (ID)	1	2	0	0
Student Plan Type				
Individual Education Program (IEP)	22	44	20	40
BIP	2	4	0	0
Student Improvement Plan (SIP)	5	10	1	2
Social Skill Group	8	16	6	12
504 Plan	1	2	1	2

Note. $N = 100$ students ($n = 50$ per group)

The study included classrooms in private, public, and charter schools (see Table 3). All ten educators reported adequate access to technology and broadband internet in their classrooms. All student and educator participants reported extensive (i.e., three or more years) experience utilizing the technology generating the social narratives (i.e., Chromebook, iPad) in their classrooms. Educators' time spent teaching ranges from 3 years to over 25 years. The same was true for the number of years providing instruction in developing social skill competencies.

Table 3: Educators' School Demographic Information

Location	*n*	%	State	*n*	%	Type of School	*n*	%
Rural	3	30	KS	2	20	Public	5	50
Suburban	4	40	NM	1	10	Charter	1	10
Urban	3	30	VA	4	40	Private	4	40
			NC	3	30			

Participants assigned to each paired group received intervention in the same room, at the same time of day, using the same technology device, with the same educator, excluding unforeseen absences from school. Interventions were administered through the same readily available technology (i.e., Chromebook or iPad) used daily by the student in

the classroom. This was done to control the technology device displaying the intervention and the student's familiarity with this device.

Setting and Materials

To ensure maximization of learners' time and to comply with school COVID protocol regulations, all sessions occurred over Zoom in a room with tables or desks, a computer connected to Zoom, and a Chromebook or an iPad. A physical presence by a researcher was made available upon request, but no physical presence was requested. All participants had their educator in the room during all stages of the study. Throughout this study, the word "educator" refers to the adult (e.g., teacher, counselor) who typically provides the student's social skill instruction. This study began in September 2022 and ended in April 2023 with the intervention phase all occurring between February 2023 and March 2023.

Social Narrative Training

Educators and students spent two sessions (60 minutes total) within a two-week time span being trained in creating social narratives. The first author, with twenty years of experience in special education and social skill instruction, implemented all training over Zoom.

Program and Training

MagicSchool AI chatbots assist middle schoolers in crafting social narratives by providing prompts on topics like friendship, empathy, and conflict resolution. They aid in character development, plot structuring, dialogue practice, feedback provision, and fostering improved communication skills, empathy, conflict resolution abilities, collaborative storytelling, and self-confidence. MagicSchool AI's chatbots are particularly beneficial for middle school students, educators, and education professionals. While suitable for elementary school users as well, the platform is optimized for middle school and higher age groups due to the complexity of topics and resources available. The chatbots offer tailored guidance and resources to enrich the learning experience for middle school students, aiding in their academic and social skill development.

Intervention

The intervention occurred during the student's normally scheduled social skill instructional time. The intervention session length per day was dependent on the typical class intervention schedule for that school and ranged from 20 to 60 total minutes per day and one to four sessions per week. Each matched peer received the same scheduled intervention day, time, and amount unless absent from school. If absent from school, the participants made up the time the day they arrived back at school in the same classroom environment and with the same educator as their matched peers. The varying intervention lengths occurred because the intervention was intended to be delivered at the educator's normal social skill instructional time to ensure students were not pulled from any academic instruction or extracurricular events. Every participant received an estimated 150 minutes of intervention within a two-month time span. Two coders observed randomly selected sessions to ensure coding for a minimum of 34% of sessions for reliability purposes.

Technology Delivering the Intervention

The chatbot and social skill narrative creation was delivered through the same iPad or Chromebook through which the student participant typically received their instruction. There were 84 students who utilized a Chromebook and 16 who utilized an iPad. Their assigned matches were also using the same device in the same classroom. On these devices, the students hear and see all visuals and sounds from the iPad or Chromebook's speakers and can move and respond while tapping on buttons on the iPad screen or Chromebook mouse. It was decided to use the device the students currently use most frequently in their classroom in attempts to decrease the amount of time needed to familiarize the student with the technology device, to decrease the possibility of gains or losses being caused by the novelty of the device, and to ensure students who had accessibility needs were able to utilize their everyday instructional device.

Measures

The Computer Simulation Evaluation Scale for Students (CSES-S)

The Computer Simulation Evaluation Scale for Students (CSES-S) is designed to assess attitudes toward computer-based simulations, focusing on learning, quality, and engagement. It is a reliable and valid tool for assessing attitudes toward the use of computer-based simulations as a learning aid. It provides valuable insights into learning, quality, and engagement constructs and can be used in a variety of educational settings. Importantly, the same items can be used to report perceptions on interactions with a human, like a teacher, so that allowed for comparison between groups. While it's not specifically designed for AI chatbots, some aspects of the scale are applicable. For instance, the learning construct gauges whether students feel they are learning effectively from the technology, which is the primary goal of the project. The quality construct assesses the perceived quality of the chatbot's responses. The engagement construct measures how engaging students find interactions with the AI chatbot. Internal reliability estimates for the three constructs (learning, quality, and engagement) on the CSES survey were calculated using Cronbach's reliability estimates. These values ranged from 0.78 to 0.89, which are considered moderate to high and acceptable for measures in the social sciences. The survey consists of twelve statements about beliefs in learning. Respondents are asked to read and assess each statement by selecting a number between 1 and 5 that best indicates their feelings about each statement. The numbers indicate: 1–Strongly Disagree, 2–Disagree, 3–Neutral, 4–Agree, and 5–Strongly Agree. This five-point Likert scale examines perceptions of learning (learning construct), the quality of the computer-based simulation (quality construct), and engagement (engagement construct) with computer-based simulations.

Survey Administration

Written surveys can be subject to coverage error and item nonresponse, where some questions can be inadvertently or intentionally skipped (Salant & Dillman, 1994). To resolve the possibility of coverage error, the questions of the survey were electronically randomized by classroom to help limit biased context results and ensure that if people quit part way through the survey, the data collected would not be substantially affected. Randomization also limited the possibility of the order influencing the participants' responses. The surveys were distributed to all matched participants within the same time frame to ensure the surveys did not reflect seasonal or temporal differences. Data was analyzed immediately following collection. Qualtrics (Provo, UT) was chosen for the survey platform because of its accessibility, data security, and randomization features. Experts were consulted to ensure appropriate language and response options as well as to assess whether the surveys measured the target construct.

Procedures

The experimental design consisted of 60 minutes of training on creating social narratives broken into 2 or 3 sessions for all participants, the intervention phase, and post-test survey with questions. Each student and educator were given a checklist with each day of the study procedures listed as well as every social skill to be learned in a social narrative. There was a box for the student and educator to mark off as they completed each task. The educator was also trained on and given a fidelity implementation checklist, to assist the educator with implementing the procedures in a way the educator was trained.

Groups

Students received intervention in 2 groups. Group 1 generated and read social stories created in collaboration with their educator. Group 2 generated and read social stories created in collaboration with the MagicSchool Social Stories Chatbot. Survey questions on students' attitudes were responded to using Qualtrics immediately following the use of the technology. Students were matched who have similar scores in the CELF-5 Pragmatic Profile. From those matched students, each was randomly assigned by SPSS to either Group 1 or 2. The participants paired are of similar demographics and taught similar curricula outside of the intervention sessions. Student demographic information was obtained from the educator prior to randomization and included gender, age, experience with the technology being used to deliver the intervention, any diagnosis, and any current school plan or program (e.g., individualized education program, behavioral intervention plan, student improvement plan). Paired participants were then entered into the Statistical Package for Social Sciences (SPSS) for one of the pairs to be randomly assigned a number one (the control condition with no AI) or a two (the training condition with AI generation). After all students were paired, an additional assessment of each group's traits (i.e., diagnosed disability, chronological age, race, gender, and type of educational plan) was considered to ensure groups were not disproportional. Priority was given to age and gender because of prior literature reporting differences in acceptability and outcomes for these two groups. Having no variance in all variables for a population of 100 students was not possible, but this allowed for a good measure of accuracy in pairing similar students.

Training

All participants were trained in creating social narratives based on the skill to be learned in collaboration with feedback. The AI grouping received 30 minutes of instruction on how to get on and locate the MagicSchool social skill chatbot on the iPad and Chromebook. All training sessions occurred utilizing a "teach, model, and do" with a guided feedback approach, with time for questions and answers and three examples. The training took place via Zoom based on the desire of the participants and schools during continued COVID protocols. Participants and educators also received visuals with written reminders able to be read aloud in Google Classroom on a mouse click, in case they forgot a step or needed a reminder on how to navigate to the AI chatbot. Practice sessions were utilized with participants in both groups until the participants were able to independently navigate the chatbot. The training was complete when the researcher observed the participant independently accessing and successfully opening the MagicSchool website on their device.

Fidelity

To ensure that the students were trained to create quality social narratives, the lead author and a trained graduate student independently evaluated the teaching. Inter-observer reliability data was collected through direct observational recording of agreements and disagreements: (agreements/ [agreements + disagreements] x 100 = percent of agreement). Agreement on the session was defined as at least 90% agreement on the identified procedural steps. Cohen's kappa (κ) was calculated to determine the level of agreement between the coders while controlling for agreement due to chance. There was substantial agreement ($\kappa = 0.89$, $p < .005$) that all core pieces necessary to accurately instruct on social narrative generation were provided by the educator.

The total twenty-seven sessions randomly selected for procedural fidelity included sessions that occurred during the training, intervention, and survey phases. Inter-observer agreement (IOA) was taken across participants and conditions using point-by-point agreement ([agreements/total number possible] × 100). Each of the two independent graduate students viewed either a live or recorded randomly assigned session and completed the checklist to measure adherence to the checklists. The observer assessed whether the implementer (a) presented the correct materials (Chromebook/iPad), (b) ensured the participant accessed the right collaborator (educator/Magic School AI Social Story Chatbot), (c) the accuracy of reading survey questions, response options, and procedural script to participants across all conditions, (d) whether the participant remained on task until the social story was completed, (e) whether participants completed the session checklist as listed, and (f) whether the implementer followed implementation guide procedures accurately.

Experimental Design and Data Analysis

A randomized control trial (RCT) design with two randomly assigned and matched groups was utilized to evaluate any differences in responses between the educator group and the AI chatbot group. A power analysis suggested 60 total participants would be adequate for a medium effect size of .50 (d=.5) to be observed at a statistical power at the level of .70.

The data was analyzed using the Statistical Package for the Social Sciences (SPSS). Descriptive statistics were calculated for the end-of-intervention scores for each group. The distribution of scores within each group was visualized using bar charts and box plots. To determine whether there were significant differences in the end-of-intervention scores between the groups, a one-way Analysis of Variance (ANOVA) was conducted. The group was treated as the independent variable, and the end-of-intervention score was the dependent variable.

RESULTS

The purpose of this study was to explore differences in students' attitudes towards creating and reading social narratives on their technology devices, such as iPads and Chromebooks, when the narratives were generated in collaboration with an AI chatbot compared to those created in collaboration with their educator. Social narratives, which are used to teach social skills and improve social understanding, can be an effective tool for students with disabilities. To address this question, we gathered data from students who participated in the creation of social narratives through two different methods. In one condition, students worked with an AI chatbot to develop their narratives. In the other condition, they collaborated with their teacher. By comparing the attitudes of students towards

both the creation and reading of these narratives, we aimed to uncover any significant differences that might inform future educational practices and the integration of AI tools in social skill development.

Analyses of variance were conducted comparing the educator group and AI group on the total score of the 12-item *Computer Simulation Evaluation Scale for Students.* Additionally, comparisons were made on each of the 12 items individually. The analysis of variance for the total score was significant, F (1,98) = 24.89, $p < .001$, with a large effect size, eta-squared = .20. Students preferred the AI support (M= 4.11 on a 1 to 5 scale) much more than teacher support (M=3.19). Results for analyses of variance for each of the individual items are shown in Table 4. All items, with the exception of Item #8, are significantly different with students consistently preferring the AI training. Most effect sizes are large.

Table 4: Comparisons of Individual *Computer Simulation Evaluation Scale for Students* Items

	Educator Group		AI Chatbot Group				
Item	**Mean**	**SD**	**Mean**	**SD**	***F* (1, 98)**	***p***	**η^2**
1. Working with the educator/AI helped me learn effectively.	2.86	1.46	4.16	1.02	26.76	<.001	.21
2. The feedback from the educator/AI helped me learn well.	2.76	1.41	4.08	.97	29.89	<.001	.23
3. The graphics and animation created by the educator/AI helped me learn effectively.	3.14	1.39	3.78	1.23	5.95	.02	.06
4. The educator/AI taught me a new concept.	2.66	1.38	3.88	1.14	23.30	<.001	.19
5. Overall, the educator/AI helped me understand the concept.	3.16	1.60	4.30	.91	19.27	<.001	.16
6. The help features from the educator/AI were useful.	2.50	1.33	3.86	1.31	26.58	<.001	.21
7. The instructions from the educator/AI were easy to follow.	3.88	1.21	4.32	1.00	3.95	.05	.04
8. The educator/AI was easy to use.	4.24	1.02	4.50	.76	2.08	.15	.02
9. The educator/AI was well organized.	3.70	1.54	4.34	.90	6.44	.01	.06
10. I liked the overall theme, visuals and graphics given by the educator/AI.	3.28	1.23	3.94	1.06	8.28	.01	.08
11. I found the educator/AI's social narratives motivating.	3.00	1.18	4.10	1.07	23.81	<.001	.20
12. I would like to use the educator/AI's help to create these social narratives again.	3.10	1.40	3.94	1.33	9.43	.003	.09

Note. Answer options range from 1= strongly disagree to 5= strongly agree.

DISCUSSION

This study aimed to determine if there is a difference in students' attitudes towards creating social narratives when these narratives were generated in collaboration with an AI chatbot versus in collaboration with their teacher. Understanding these attitudes is pivotal in assessing the effectiveness and acceptability of incorporating AI-driven tools in classroom instruction. Our findings revealed that students overwhelmingly preferred training with AI chatbots. Instruction to facilitate social competence development is essential for many students with disabilities. This deficit impacts learning and overall development within the school and broader community setting that expects students to have knowledge and skills associated with essential social skill capacity (Gresham, 2017). The results of this study showed effective social skill instruction can be furthered through the assistance of AI, specifically using an AI chatbot.

AI directly helped the students in this study further understand the demands of the social skill interaction and in turn, learned the essential skills being introduced and further demonstrated when the AI chatbot was available. Likewise, group members shared that the just-in-time feedback, available within the context of the social competence being learned, helped the participant learn the knowledge and skills associated with the social narrative. There was a significant difference in group members who learned from the teacher-supported social narrative and the AI chatbot instruction. Students reported that AI helped the student understand the nuances of the social skill instruction. The ongoing support, particularly through the various tech-based features, further assisted the learner during the social

narrative use. Overall, participants responded to the immediacy of AI feedback and the ability to receive the necessary guidance when engaged in the social narrative instruction while gaining further knowledge and skill of the expected social skill.

Our study involved the essential elements of an effective social competence intervention, social narratives, combined with AI chatbot supports to facilitate the efficiency and effectiveness of supporting students' social skill development. The ability for the students to successfully interact with engaging social narratives that motivated their learning experience was noted as an essential part of the learning experience. For students with disabilities who have trouble understanding and using verbal and nonverbal communication, difficulty interpreting social cues and body language, struggles with taking turns in conversation, and difficulty managing emotions and reactions in social situations, AI chatbot assistance was found to be beneficial in supporting students' engagement in social narrative applications for social skill development. In turn, the supports and engaging features of the AI chatbot enhanced student motivation and interest in continuing with the social narrative process.

Recommendations for Future Research and Intervention

Overall, our findings supported the use of AI to further engage, motivate, and support student social narrative application to gain knowledge and skills to address social skill deficits. This is promising for future study as we seek to further support learning for students with disabilities, enhance the just-in-time feedback students often require for successful social skill instruction, and further personalize the learning experience to the specific and unique needs and complexity of social skill instruction. Results of this initial study can be used to inform recommendations for future research and intervention, particularly towards social competence development.

Educators struggle to implement social narratives because it is difficult to maintain the fidelity of the intervention while adapting to meet the individual student needs. More systematic methods are needed. Our results highlight the importance and value of integrating AI into the complexity of social skill instruction through social narratives. Prior research for students with disabilities reinforces the need to plan and design for student choice, opportunities to respond, gain praise and precorrection, and receive consistent reinforcement and rewards. Our initial findings highlight the value of the integration of the social narrative process with AI to allow students to further engage, self-manage, and benefit from the social skill instruction. Likewise, secondary outcomes indicated an enhanced ease of use, further organization, essential visual supports, and an interest, on the part of the student, to continue with further social skill instruction. Schools and educators should be encouraged to combine interventions with the increased support of AI tools to facilitate student learning. Intentional study of these combinations should be conducted to support other educators in similar settings. AI as a simple tool can be a powerful intervention when combined with proven strategies.

Limitations

One statistical limitation concerns the multiple comparisons using each individual item as dependent variables. One concern is that many consider Likert-type items to be at only the ordinal level of measurement, not at the interval level which is an assumption when conducting analyses of variance. So, individual *p* values may not be precise. The overall pattern, however, along with the large difference found on the total scale scores, supports a clear preference for AI support for this activity. Conducting 12 different statistical comparisons at the item level increases the risk of inflation of type I error, or the possibility that some comparisons might be significant only by chance. A more conservative approach would be to test each of those comparisons at the .05/12 level, or .004. Even at that level, though, most of the items still show a significant difference.

REFERENCES

Bernardini, S., Porayska-Pomsta, K., & Smith, T. J. (2014). ECHOES: An intelligent serious game for fostering social communication in children with autism. *Information Sciences*, 264, 41-60. https://doi.org/10.1016/j.ins.2013.10.027

Clark, R. C., & Mayer, R. E. (2023). *E-learning and the science of instruction: Proven guidelines for consumers and designers of multimedia learning*. John Wiley & sons.

Charlop, M. H., & Milstein, J. P. (1989). Teaching autistic children conversational speech using video modeling. Journal of Applied Behavior Analysis, 22(3), 275-285.

Chiu, T. K., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of artificial intelligence in education. *Computers and Education: Artificial Intelligence*, *4*.

Corbett, B. A. (2003). Video modeling: A window into the world of autism. *The Behavior Analyst Today*, 4(3), 367. https://doi.org/10.1037/h0100025

D'Ateno, P., Mangiapanello, K., & Taylor, B. A. (2003). Using video modeling to teach complex play sequences to a preschooler with autism. *Journal of Positive Behavior Interventions*, 5(1), 5-11.

Diliberti, M. K., Schwartz, H.L., Doan, S., Shapiro, A., Rainey, L.R., and Lake, R.J., *Using Artificial Intelligence tools in K–12 classrooms.* Santa Monica, CA: RAND Corporation, 2024. https://www.rand.org/pubs/research_reports/RRA956-21.html.

Evangeline, S. B., & Moorthy, A. D. (2024, March). Investigating AI applications in communication tools for individuals with speech impairments: An in-depth analysis. In 2024 *IEEE International Conference on Interdisciplinary Approaches in Technology and Management for Social Innovation (IATMSI)* 2, 1-6).

Fryer, L. K., Nakao, K., Thompson, A. (2019). Chatbot learning partners: Connecting learning experiences, interest and competence. *Computers in Human Behavior*. doi10.1016/j.chb.2018.12.023

Fulton, R., Fulton, D., & Kaplan, S. (2022). Artificial Intelligence: Framework of driving triggers to past, present and future applications and influencers of industry sector adoption. *arXiv* preprint arXiv:2204.01518.

Ghanouni, P., Jarus, T., Zwicker, J. G., Lucyshyn, J., Mow, K., & Ledingham, A. (2019). Social stories for children with autism spectrum disorder: Validating the content of a virtual reality program. *Journal of Autism and Developmental Disorders*, 49(2), 660-668.

Giray, L. (2023). Prompt engineering with ChatGPT: a guide for academic writers. *Annals of biomedical engineering*, 51(12), 2629-2633.

Grams, D. (2018). *A quantitative study of the use of "DreamBox Learning" and its effectiveness in improving math achievement of elementary students with math difficulties* (Publication No. 10744570) [Doctoral dissertation, Northcentral University]. ProQuest Dissertations & Theses Global.

Gresham, F. (2017). *Effective interventions for social emotional learning*. Guilford Press.

Hasan, M. R., & Khan, B. (2023). An AI-based intervention for improving undergraduate STEM learning. *PloS one*, 18(7), e0288844.

Hashem, R., Ali, N., El Zein, F., Fidalgo, P., & Khurma, O. (2024). AI to the rescue: Exploring the potential of ChatGPT as a teacher ally for workload relief and burnout prevention. *Research & Practice in Technology Enhanced Learning*, 19, 23.

Kirschner, P. A., & De Bruyckere, P. (2017). The myths of the digital native and the multitasker. *Teaching and Teacher education*, 67, 135-142.

Labadze, L., Grigolia, M. & Machaidze, L. (2023). Role of AI chatbots in education: systematic literature review. *International Journal of Educational Technology Higher Education* 20, 56.

LeBlanc, L. A., Coates, A. M., Daneshvar, S., Charlop-Christy, M. H., Morris, C., & Lancaster, B. M.(2003). Using video modeling and reinforcement to teach perspective-taking skills to children with autism. *Journal of Applied Behavior Analysis*, 36(2), 253-257.

Lopes, A. M. M., de Magalhães Netto, J. F., de Souza, R. A., Mourão, A. B., Almeida, T. O., & de Lima, D. P. (2019, October). Improving students' skills to solve elementary equations in K-12 programs using an intelligent tutoring system. In *2019 IEEE Frontiers in Education Conference (FIE)*. 1-9.

MagicSchool. (2024). *http://magicschool.ai*.

Makridakis, S., Petropoulos, F., Kang, Y. (2023). Large language models: Their success and impact. *Forecasting*, 5, 536-549. https://doi.org/10.3390/forecast5030030

Marino, M. T., Vasquez, E., Dieker, L., Basham, J., & Blackorby, J. (2023). The future of Artificial Intelligence in special education technology. *Journal of Special Education Technology*, https://doi.org/01626434231165977.

Mello, R. F., Freitas, E., Pereira, F. D., Cabral, L., Tedesco, P., & Ramalho, G. (2023). Education in the age of Generative AI: Context and Recent Developments. *arXiv* preprint arXiv:2309.12332.

Olejnik, S. F., & Algina, J. (1984). Parametric ANCOVA and the rank transform ANCOVA when the data are conditionally non-normal and heteroscedastic. *Journal of Educational Statistics, 9*(2), 129–149. https://doi.org/10.2307/1164717

Ribble, M. (2015). *Digital citizenship in schools: Nine elements all students should know*. International Society for Technology in Education.

Rožman, M., Tominc, P., & Vrečko, I. (2023). Building skills for the future of work: Students' perspectives on emerging jobs in the Data and AI Cluster through artificial intelligence in education. *Environment and Social Psychology*, 8(2).

Sallam, M. (2023). The Utility of ChatGPT as an Example of Large Language Models in Healthcare Education, Research and Practice: Systematic Review on the Future Perspectives and Potential Limitations. *medRxiv*. doi:https://doi.org/10.1101/2023.02.19.23286155

Sanusi, I., Olaleye, S.,Agbo, F., & Chiu, T. (2022). The role of learners' competencies in artificial intelligence education. *Computers and Education: Artificial Intelligence*, 3.

Şen, N., & Akbay, T. (2023). Artificial intelligence and innovative applications in special education. *Instructional Technology and Lifelong Learning*, 4(2), 176-199.

Shoaib, M., Sayed, N., Singh, J., Shafi, J., Khan, S., & Ali, F. (2024). AI student success predictor: Enhancing personalized learning in campus management systems. *Computers in Human Behavior*, 158, 108301.

Sutherland K. S., Conroy M. A., McLeod B. D., Kunemund R., McKnight K. (2019). Common practice elements for improving social, emotional, and behavioral outcomes of young elementary school students. *Journal of Emotional and Behavioral Disorders*, 27(2), 76–85.

Topsakal, O., & Topsakal, E. (2022). Framework for a foreign language teaching software for children utilizing AR, voicebots and ChatGPT (large language models). *The Journal of Cognitive Systems*, 7(2), 33-38.

Wong, C., Odom, S. L., Hume, K. A., Cox, A. W., Fettig, A., Kucharczyk, S., Brock, M. E., Plavnick, J.B., Fleury, V. P., & Schultz, T. R. (2015). Evidence-based practices for children, youth, and young adults with autism spectrum disorder: A comprehensive review. *Journal of Autism and Developmental Disorders*, 45(7), 1951-1966. https://doi.org/10.1007/s10803-014-2351-z