

Evaluation of MagicSchool AI-Generated Elementary Teaching Materials: A Self-Study

by

Leah Carr

A Signature Honors Project Presented to the

Honors College

East Carolina University

In Partial Fulfillment of the

Requirements for

Graduation with Honors

by

Leah Carr

Greenville, NC

January, 2025

Approved by:

Dr. Kristen H. Gregory

College of Education, Department of Elementary and Middle Grades Education

Evaluation of MagicSchool AI-Generated Elementary Teaching Materials: A Self-Study

Leah M. Carr

Department of Elementary Education and Middle Grades Education, East Carolina University

Signature Honors Project

Dr. Kristen H. Gregory

January 15, 2025

Abstract

Recent developments in Artificial Intelligence (AI) have resulted in many possible tools, such as MagicSchool AI, that educators could use in their practice. Possible outcomes of using AI in their practice could be reduced teacher burnout, reduced time for developing materials, and increased time for other aspects of their role as an educator. Thus, the purpose of this mixed methods self-study is to investigate my experiences using MagicSchool AI to generate elementary classroom materials for a two-week integrated instructional unit, the quality of the AI-generated materials, and my pedagogical decisions surrounding using those materials in my personal teaching practices. I collected the following data: AI-generated lesson plans, lesson materials, and assessments; evaluation of AI-generated materials; personal journal entries; and recorded and transcribed critical friend conversations. Quantitative data analysis included descriptive statistics, and qualitative data analysis included open coding and axial coding to identify themes. The findings address experiences utilizing MagicSchool AI to generate classroom materials, professional development opportunities, best AI practices for using in the elementary classroom, and my pedagogical decisions based on the quality of the materials. Implications for future practice are discussed for pre-service and in-service teachers as they use AI to generate classroom materials and future research.

Keywords: MagicSchool AI, Artificial Intelligence, elementary education, lesson planning, instructional materials

Evaluation of MagicSchool AI-Generated Elementary Teaching Materials: A Self-Study

Artificial Intelligence (AI) is a field of study that combines the applications of machine learning, algorithm productions, and natural language processing whose applications transform the tools of education (Akgun, 2022). One application, Chat Generative Pre-Trained Transformer, or ChatGPT, was first introduced in 2022 and sparked widespread usage and ethical concerns surrounding the use of AI in the classroom (OpenAI, 2024). More recently, large language models developed specifically for educators, such as MagicSchool AI, are becoming available. MagicSchool AI is a LLM specifically designed to cater towards the needs of educators, providing an abundance of classroom-related tools to assist teachers in classroom planning, communication, and instruction. Due to unknown information about the new technology, some educators are concerned with whether using AI to generate classroom materials is a violation of ethical and professional standards (Kehoe, 2023). Regardless of such ethical considerations, educators are tasked with determining if AI is a viable tool, what kinds of classroom materials can be generated, and if those materials are aligned with learning standards and students' learning needs.

Research Problem

Recent developments in AI have resulted in an ethical dilemma surrounding its use in education. AI can be used by students for a variety of purposes (e.g., information-seeking, summaries, thesis statements), but it can also be utilized by educators in the creation of teacher materials and lessons. As educators continue to balance many aspects of their role, they may benefit from AI as a tool to generate classroom materials. However, teachers currently do not know whether or not to use AI to generate classroom materials because they are unsure of the model's credibility, efficiency and accuracy, alongside a lack of research on how teachers should

use AI to generate materials (Trust et al., 2023; van den Berg & du Plessis, 2023). Possible outcomes of utilizing AI in the classroom could be reduced teacher burnout, reduced time for developing materials, and increased time for other aspects of their role as an educator (Pearce, 2023). As a pre-service elementary teacher, I am concerned about the lack of research and guidance about using AI, and more specifically MagicSchool AI, to generate classroom materials.

Justification for the Study

The integration of AI into education is a critical area in need of investigation, emphasized by many recent studies. According to Celik and colleagues, educators currently lack a fluent understanding of how to utilize AI effectively and ethically in the classroom (2022). Aside from the call for educators to develop knowledge of how to use AI fluently, Kehoe (2023) advocates for teachers to be trained on how these technologies can positively impact their classrooms. van den Berg and du Plessis (2023) concur with this notion, going on to say that if large language models are

used by student teachers and teachers to evaluate and critique existing lessons, they might be equipped with the skills they need to succeed in the technology-driven world by focusing on applying knowledge; creating new meanings from existing knowledge; and developing skills relevant to the 21st century, such as critical thinking and problem solving. (p. 10)

Therefore, a study centered around integrating AI, and more specifically MagicSchool AI, into the classroom efficiently and ethically is highly relevant, as it will not only address current gaps in the literature but also contribute to the development of skills needed for teachers to succeed in a technology-enhanced classroom.

Gaps in the Research

Research on AI use in education has become more prevalent in the last four years and is projected to increase in the future; however, there is a lack of research in the field of education centered around the usage of AI, due to resistance from decision makers (Celik et al., 2022). According to Su and Weipeng (2023), there is a need for future research to focus on “developing and evaluating AI-based educational applications,” bringing to light the benefits and challenges of its use across a variety of educational settings. These benefits were shed light on by Gregory et al. (2024), as practicing teachers in a graduate education program used ChatGPT to generate classroom assessments. In fact, a participant in the study stated, “We are educating students for a world that we are creating as it goes. The more familiar we are with AI, the better we can support them” (Gregory et al., 2024, p. 151). While some educators found this to be a beneficial tool to create classroom materials, others were concerned about adequate training opportunities and the ethical use of AI potentially impacting their professionalism. Crompton et al. (2022) conducted a systematic review of challenges and affordances of AI in K-12 education, determining that there is reason for ethical and privacy concerns. To combat this issue, the researchers call for additional research and teacher training to minimize ethical risks (Crompton et al., 2022; Gregory et al., 2024).

Due to the current state of research, a specific gap has been noted in understanding the perspectives and usage of AI in preservice teachers. However, resistance from decision-makers could be disregarded after demonstrating the improvements on positive self-efficacy and efficiency in the classroom (Celik et al., 2022). These deficiencies in the current research will be addressed in the research conducted.

Intended Audience

The research completed in the study will benefit many individuals and organizations interested in improving the field. First, teachers, curriculum specialists, and administrators alike will benefit greatly from this study due to the insights on the implementation of AI-generated materials in a classroom setting. Understanding the benefits and drawbacks across a variety of educational settings will allow for informed decision making and effective implementation. In addition, the study will benefit preservice teachers, allowing for targeted preparation to enter the classroom in a technology-driven society, such as understanding how MagicSchool AI can automate tasks and assist in daily classroom tasks. Furthermore, state boards of education and local school districts will benefit from the knowledge of how AI can be used, for consideration in formative of new standards and trainings. Finally, students will benefit from this study, as the implementation of these new technologies has the potential to offer a better educational experience for all.

Research Purpose and Question

Thus, the purpose of this mixed methods self-study is to investigate my use of MagicSchool AI to generate elementary classroom materials for a two-week integrated instructional unit, the quality of the AI-generated materials, and my pedagogical decisions surrounding using those materials in my personal teaching practices. It is the aim that the findings will provide insight for future teaching practices utilizing AI. As a pre-service elementary teacher, I am concerned about the lack of research and guidance on using AI, and more specifically MagicSchool AI, to generate classroom materials. The following research questions will be addressed:

- What are my perceptions of using Artificial Intelligence to generate elementary classroom materials for a two-week integrated instructional unit?
- How do the AI-generated materials align with standards, learning objectives, and student learning needs?
- What are my pedagogical decisions surrounding using the AI-generated materials?

Through this study, I will investigate how AI could play a role in my own teaching practices.

This will not only benefit my own pedagogical decisions, but it will also improve the quality of education for my students.

Literature Review

The use of Artificial Intelligence in the field of education is a rising trend, which must be understood before taking on the study. The following sections will cover four main topics related to this study: Overview of lesson planning, frameworks to consider for AI in education, large language models, and considerations for the use of AI in education. Looking to how AI is being used in education must be addressed, highlighting its use in content areas, lesson planning, creating assessments, generating data, and designing differentiation. I will present a review of the frameworks, highlighting the IDEE Framework, TPACK Framework, and AI Literacy. Moving forward, the large language models used in the study and highly relevant in the field, ChatGPT and MagicSchool AI, will be addressed. Finally, the literature review will conclude with considerations taken from the review of current literature, highlighting the benefits, concerns, and ethics surrounding AI.

Overview of Lesson Planning

Educators have begun to utilize generative AI in developing educational materials for specific content areas, such as social studies and English language arts. Recent studies have noted

the specific, promising benefits of generative AI models in the formation of 5E lesson plans (Goodman et al., 2024). However, researchers have noted the need for this technology to be used to assist teachers, rather than replace, and stress the need for teachers to “critically evaluate, refine, and modify” the outputs from large language models to ensure they match the instructional goals of their lesson (Goodman et al., 2024, p. 2). Additionally, Trust et al. (2023) have found the time-saving ability of large language models to generate a number of classroom instructional materials, such as mathematical procedure or word problems, analytical passages and comprehensive questions on a subject or passage, bringing instructional goals to light. As additional research comes to light, educators will develop a more structured understanding of generative AI’s strengths and weaknesses in terms of content areas.

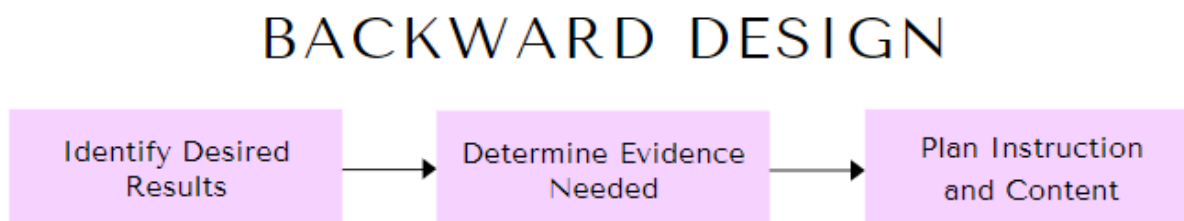
Backward Design

When determining how to approach a lesson plan, research has shown the use of Backward Design in using the learning objective to catapult planning. According to the creators, Wiggins & McTighe, backward design is “an approach to designing a curriculum or unit that begins with the end in mind and designs toward that end” (2006, p. 10). Even though this approach seems to make sense, and is what some teachers do in the classroom, many plan lessons around a desired activity or how the textbook is set up. The researchers call for educators to begin their planning by identifying the desired results of their students, then determining what evidence they need an assessment to contain to ensure students have met the desired results (see Figure 1; Wiggins & McTighe, 2006). By doing this, the teacher can be sure of what is needed of the students to move into the final stage of backwards design: planning. Educators must take the known desired results and planned assessment and determine all content students must need to achieve this result (Wiggins & McTighe, 2006). Taught as the standard for teachers at many

universities, the backward design approach is a well-established framework to ensure students are supported to meet their goals.

Figure 1

Phases of Backward Design



Educators have turned to generative AI for the opportunities associated with lesson planning, as researchers have found generative AI can produce an acceptable lesson plan (van den Berg & du Plessis, 2023). However, after studying the current literature published in the field, it has been determined that one must evaluate the generated material's correctness, creativity, demand and context; therefore, helping teachers will broaden their knowledge and perspectives and potentially develop a sense of creativity inspired by generated AI outputs (van den Berg & du Plessis, 2023). In addition, the researchers found some large language models excelled in creating higher order thinking questions and yielded better results when building a lesson plan in small chunks, rather than all at once (Goodman et al, 2024). This is because generative AI can modify its responses based on new requests quite well; therefore, teachers can work back and forth with AI until an acceptable output is provided (Goodman et al., 2024). Preservice educators who are completing their student teaching can utilize generative AI to evaluate concepts against academic standards and references. Materials must be evaluated against standards, adapted to meet the needs of learners, and applied to determine effectiveness.

The lesson must be evaluated to determine its relevance, feasibility in the classroom environment, and modifications needed to meet student abilities (van den Berg & du Plessis, 2023). Supporting this point, graduate students in their first years of teaching expressed benefits from learning how to use ChatGPT in their educational practices (Gregory et al., 2024). There is a long way to go in determining the specific best practices in using generative AI for lesson planning, but a good start has been made.

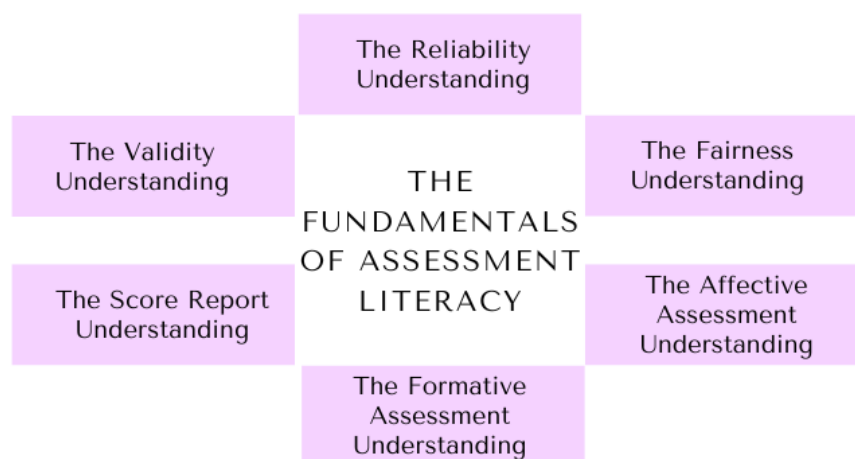
Assessments

An assessment, or evaluation of someone's status in terms of educational variables, must be effective in determining student understanding of content with both accuracy and relevance. As an educator, Dr. James Popham, well-known for his work at UCLA, conducted research and established six key points of understanding needed to develop assessment literacy as an educator (see Figure 2). First, the validity of understanding is imperative, as it ensures key concepts of understanding that students must demonstrate are targeted in the assessment (Popham, 2018). Reliability understanding ensures students have retained and comprehend desired learning goals through the collection of at least three points of evidence. Furthermore, educators are called to create and evaluate assessments with a critical eye to ensure fairness for all students and their variety of specific learning needs (Popham, 2018). Educators must understand score reporting, to ensure no misleading interpretations can be made from transparent assessment results. Additionally, a formative assessment understanding can drastically improve the performance of students, as educators and students collaboratively engage to determine current levels of understanding and make necessary adjustments to future content lessons or summative assessments (Popham, 2018). Finally, the understanding of affective assessment motivates educators to regularly assess students, while accounting for the vast impact of their emotional and

psychological needs. In summary, according to Popham (2018), assessment literacy is essential to student success in the classroom and aims to create an effective and equitable assessment system that highly values student success and authentic student understanding.

Figure 2

The Fundamentals of Assessment Literacy

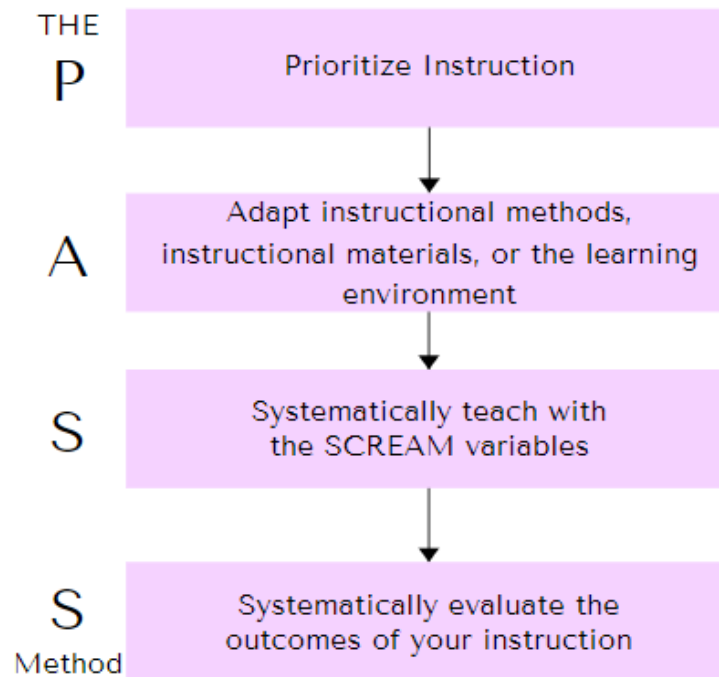


Another area of educational planning shown to be assisted by generative AI is the development of assessments. ChatGPT, or any generative AI, can help teachers create practice questions for a variety of subject assessments, generate rubrics for assessment, and suggest feedback based on student responses (Trust et al., 2023). A recent study, conducted by graduate students at East Carolina University, found that using generative AI to develop classroom assessments had three main benefits of accessibility, efficiency and assessment development. By using the software to generate an assessment, teachers can spend more time working on the important aspects which must be included in the assessment (Gregory et al., 2024). Because generative AI does not know a teacher's students, it is important for work to be done to differentiate to accommodate each student's differences. If this is not done, assessment data may

not represent what students truly know (Gregory et al., 2024). While teachers noted concern for the need to check each generated output for accuracy and feasibility to use in the classroom, they expressed confidence in their ability to complete these tasks (Gregory et al., 2024). After the assessment is completed, generative AI can support teachers in writing study guides based on a particular test. (Trust et al., 2023).

Supporting Students' Learning Needs

Meeting the learning needs of individual students is an important role in all educators' careers. Proposed by Mastropieri and Scruggs, the PASS method outlines a structure for educators to differentiate into the classroom to support learners (See Figure 3; Mastropieri & Scruggs, 2002; Scruggs & Mastropieri, 1995). To follow the method, educators must (1) Prioritize instruction, (2) adapt instructional methods, instructional materials, or the learning environment, (3) systematically teach with the SCREAM variables, and (4) systematically evaluate the outcomes of your instruction (Mastropieri & Scruggs, 2002; Scruggs & Mastropieri, 1995). The SCREAM variables are as follows: structure, clarity, redundancy, enthusiasm, appropriate rate, and maximized engagement, which is accomplished by the teacher by questioning students, as well as giving and receiving feedback (Mastropieri & Scruggs, 2002; Scruggs & Mastropieri, 1995). This method approaches differentiated instruction in an evolving, yet effective way.

Figure 3*The PASS Method*

All educational materials used in the classroom must be tailored to meet the needs of each individual learner. The Universal Design for Learning (UDL) and differentiation are additional ways for the teacher to enhance the generated lesson, as they were not included in the output provided by ChatGPT found across studies (Kehoe, 2023). Because learning should be differentiated to meet all student needs, researchers tested ChatGPT's ability to accommodate student ability. They found that ChatGPT gave suitable recommendations for students with special learning needs, when asked (Zhai, 2023). Though there is not much information on how generative AI can accommodate specific learning needs in students, more is to come.

Data

Teachers utilize data in the classroom to make a host of instructional decisions. According to Popham, the difference between summative and formative assessment types is what steps are taken after data collection and analysis (2020). Ensuring data is properly collected and analyzed is an important aspect of a teacher's role. After the collection of data, its analysis can bring to light common themes in student understanding, progression, and give insight to teacher performance (Popham, 2020). For example, data can bring to light what concepts students have mastered or need additional support in, where students are ready to go next, or how certain student groups are performing in respect to the others. Data collection is often thought of in terms of state-mandated testing; however, these tests are just one of many data points of which teachers must test, collect, analyze, and make informed decisions with.

Generative AI not only helps generate assessments but also collect data. AI can analyze data collected from students and provide adaptation suggestions to help educators differentiate for students in need of extra assistance (Pearce, 2023). Furthermore, generative AI can be utilized for tracking and reporting student performance in a way that can be shared with parents and administration alike (Zhai, 2023). Zhai's work showed the benefit and time-saving process of allowing generative AI to pull data from assessments and make suggestions in next steps; however, just as with lesson planning, these results must be critically analyzed. The researcher also found ChatGPT, the large language model used in the study, was successful in providing constructive and accurate feedback to students and their learning when given a paper to review, bringing to light another way data can be processed after the assessment is completed (Zhai, 2023). After collecting the data, generative AI can even help draft emails for teachers to communicate with their students' families (Trust et al., 2023).

Frameworks to Consider for AI in Education

The use of frameworks is imperative in an educator's role, as they use them to guide responsible decisions and effective learning in the classroom. While years of research have produced many educational frameworks for educators to follow, there are a few that stand out when considering the integration of AI in education. The emerging IDEE Framework outlines a structured approach to integrating AI into educational practices. As a relatively new area in the field, the four components - identifying desired outcomes, determining appropriate levels of automation, ensuring ethical considerations, and evaluating effectiveness - offer a systematic guide to navigating AI. The Technological Pedagogical Content Knowledge (TPACK) Framework outlines a comprehensive approach to integrate technology effectively in the classroom to support student learning and knowledge retention, similarly to the IDEE Framework. As educators work to integrate AI technologies into teaching practices, they must also develop AI literacy, a developing approach for individuals utilizing AI to interact with the technology ethically and effectively. A broad set of competencies exists for AI literacy, which will be outlined below.

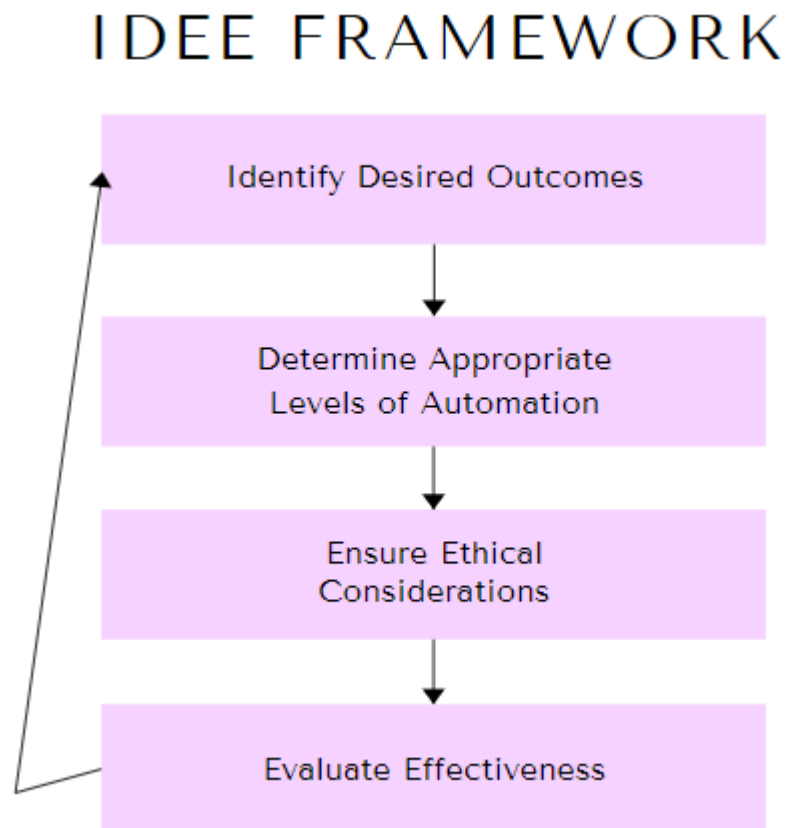
IDEE

The IDEE Framework, created by Jiahong Su & Weipeng Yang, is a structured approach for effective and ethical integration of generative AI in the field of education (see Figure 4). The framework determines four key components of AI-usage, which are to identify the desired outcomes, determine the appropriate level of automation, ensure ethical considerations, and evaluate effectiveness (Su & Weipeng, 2023). Similar to Backwards Design, educators must first identify the desired outcomes for students and align the usage of technology to enhance the

effectiveness of the lesson. The appropriate level of automation and ethical considerations are very important, as students must be both engaged and protected throughout the process. Finally, self-assessment must occur to improve before utilizing the technology again. Su & Weipeng highlight the need for the framework to be used across many settings to determine its effectiveness across a variety of contexts (Su & Weipeng, 2023, p. 363).

Figure 4

The IDEE Framework

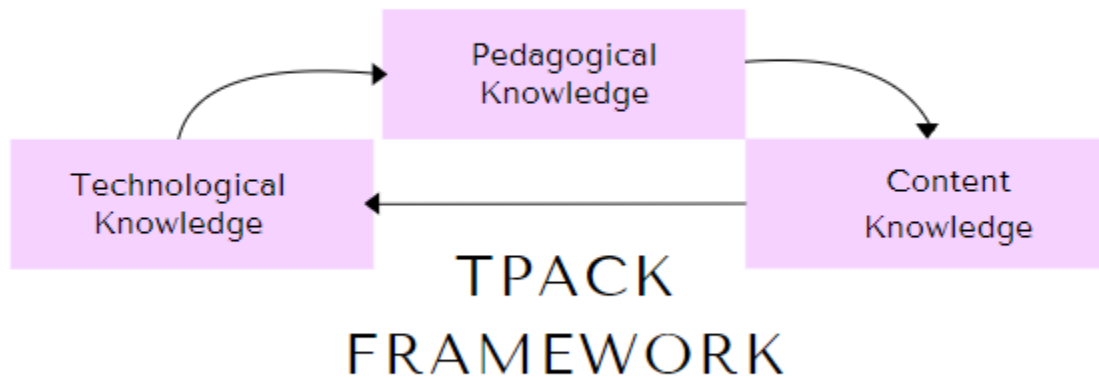


TPACK

The Technological Pedagogical Content Knowledge (TPACK) Framework outlines a comprehensive approach to integrate technology effectively in the classroom to support student learning and knowledge retention. By considering how technological knowledge, pedagogical

knowledge and content knowledge can intertwine, teachers can best meet the needs of their students and adapt to the use of AI in education. The TPACK Framework is used to ensure that all three important pieces of knowledge for students are upheld in each lesson. Targeted professional development programs centered around TPACK training have shown significant results in teacher's ability to generate AI lessons that not only meet content and pedagogical criteria, but meet AI proficiencies as well (Sun et al., 2021).

Sun et al. noted in a recent study that research collected in the study pointed to their efforts being the first to promote K-12 AI teaching competencies using a TPACK-based PD approach (2023). Furthermore, after the completion of the study, the researchers claim the necessity of TPACK training to improve programming skills of teachers, as they found the approach improved "teacher's AI teaching competency, including their AI knowledge, AI teaching skills, and AI teaching self-efficacy" (Sun et al., 2023). Ng et al. concurred with their notion in their study, referring to the TPACK approach as a "map for understanding how to integrate AI literacy into classrooms effectively," while simultaneously enhancing learning experiences for students of all ages (2021, p. 10). To conclude, the TPACK Framework considers all needed criteria for a successful lesson and ensures the usage of technology has positive effects on student learning.

Figure 5*The TPACK Framework****AI Literacy***

AI literacy describes a developing approach for individuals utilizing AI to interact with technology ethically and effectively, using knowledge from a broad set of competencies.

According to the work of Ng et al. (2021) and Long & Magerko (2020), competencies needed to achieve AI literacy include the ability to recognize AI, understand the basic applications and uses of AI, and using ethical considerations to identify the strengths, weaknesses and societal impacts of the technology. Furthermore, individuals must apply knowledge learned about AI across a number of contexts to make human-centered considerations, such as fairness and safety online, at home, and in the workplace (Long & Magerko, 2020; Ng et al., 2021). According to Celik et al., teachers have a specific role in the development of AI literacy, which includes modeling AI training, feeding AI systems and algorithms with professional development data and information about student behaviors (2022).

In a recent study conducted by Gregory et al., graduate students worked to develop their AI literacy through assessment of ChatGPT generated assessments, where students expressed the importance of familiarizing themselves with AI to better support both themselves and their students (Gregory et al., 2024). Long & Magerko support this notion in Design Consideration 12 (Leveraging Learner's Interests), where they claim teachers should design AI literacy instruction for children around the interests of their students to bridge the knowledge gap, and the same is true for teachers using the technology (2021). AI literacy is ultimately a tool to critically evaluate, effectively communicate and collaborate, and utilize AI technologies in an ethical and productive manner, which will continue to increase in necessity as individuals transition to a world driven by the power of AI.

Large Language Models

Large language models (LLMs) are machines capable of comprehending and generating human language text, a groundbreaking advancement for artificial intelligence and humans alike in terms of access to information and technology. A vast number of LLMs exist, and each model has a unique set of strengths and weaknesses surrounding “the nature of text that is returned,” as well as how the model “relates to information from real sources” (Goodman et al., 2024 p. 10). Most well-known, Chat Generative Pre-Trained Transformer, or ChatGPT, has been praised for its ability to respond to prompts of all variety with speedy human text responses (Open AI, 2022). While ChatGPT is revered for its vast data knowledge and ability to perform a variety of tasks, limitations such as the inability to comment on recent information or accept and produce media still exist. To the contrary, MagicSchool AI, an emerging LLM, is designed to specifically meet the needs of educators. With over 60 tools, the model aims to address the needs of

educators and their students, trained only on data that pertains to those needs. Moving forward, two LLMs and considerations will be explored further.

ChatGPT

A well-recognized large language model in use is Chat Generative Pre-Trained Transformer, or ChatGPT. ChatGPT is an AI that uses learned datasets to generate human-like text responses based on a given prompt. According to Zhai, “The model is trained on a vast corpus of text data, allowing it to understand and respond to a wide range of natural language inputs. ChatGPT can be fine-tuned to perform specific tasks, such as question answering, language translation, and text generation” (Zhai, 2023, p. 2). ChatGPT was trained on a dataset in 2021, meaning the software cannot comment on any recent events happening since that time. Furthermore, as of 2023, ChatGPT cannot accept a media upload, or produce any media content, such as a YouTube video to support a lesson (Trust et al., 2023, p. 6). Therefore, users cannot rely on ChatGPT to produce content pertaining to recent events or create responses outside of those which are text-based. Because ChatGPT is a “black box,” generating automatic results, its users struggle to understand how the large language model arrives at its answers; however, researchers are currently in the process of creating a more explainable model to help its users understanding the process influencing decisions and put trust in the results (Zhai, 2023, p. 6). In the future, large language models, such as ChatGPT, will continue to develop to best suit users' needs.

MagicSchool AI

MagicSchool AI is an emerging large language model based around the specific needs of educators. The creators understand that each classroom and each student have individual needs which must be addressed and reflected in teacher lessons and materials. Therefore, they have

implemented applications in their software to account for individual needs (MagicSchool, 2024). The software has over 60 tools, including complete and partial lesson plan generators, differentiation generators, and administrative task generators (MagicSchool, 2024). There is a free version of MagicSchool AI; however, the plus version costs \$12.99 a month and gives access to all 60 features of the software (Magic School, 2024). MagicSchool 3.0, the current model of Magic School AI, has built-in features to alert users to potential biases, ensure confidential information, and point out the need for accuracy. The company has also taken action to comply with FERPA and COPPA laws (MagicSchool, 2024). MagicSchool AI offers free certification training for new users and offers level training through becoming a MagicSchool AI certified trainer (MagicSchool, 2024). Because MagicSchool AI was made by educators, for educators, the data set on which the LLM received training revolves around information educators and students would need (MagicSchool, 2024). Further, free training courses are offered through MagicSchool to ensure educators' access to the technology.

Considerations of the Use of AI in Education

The integration of generative-AI tools in the classroom has ignited discussion in the field of education which must be considered. In the following sections, the implications of incorporating AI in teaching practices is discussed, highlighting potential benefits, concerns and ethical considerations. Researchers agree that generative AI should be a starting point, having teachers work alongside AI to balance interpersonal interaction into generated content to develop lessons and assessments (Celik et al., 2022; Pearce, 2023; van den Berg & du Plessis, 2023). In the first section, benefits of using the emerging technology in education are addressed, highlighting time-conservation, personalized learning, and generated feedback. Moving forward, concerns brought to light about the usage of AI in the classroom are discussed, such as potential

biases and widening the digital divide. Finally, ethical considerations are discussed, emphasizing the growing need to secure student data and additional research on the topic. By investigating these important discussions, educators will understand the current state of research on AI in education and how to approach the technology moving forward.

Benefits

Utilizing AI in education has shown many benefits for teachers and students. By automating a host of time-consuming tasks such as creating and grading individual student assessments, creating lesson plans, and determining supporting activities or needed materials, the use of AI has a significant reduction on teacher workload (Pearce, 2023; van den Berg & du Plessis, 2023 & Zhai, 2023). According to a recent study about generation of lesson plans utilizing AI, the entire lesson planning and material generation took only 20 minutes, with ChatGPT taking only 1 minute to produce a lesson response, which allows more time to complete other necessary tasks (Kehoe, 2023). In addition, generative AI can provide personalized learning for students, ease teacher workload, create an engaging atmosphere for students, give teachers useful classroom suggestions, and encourage a technologically integrated classroom (Su & Weipeng, 2023). Furthermore, the personalized learning AI offers to students has shown positive effects on student achievement and engagement and can provide students with “highly personalized educational experiences” (Su & Weipeng, 2023). According to Kehoe, AI provides additional support to teachers through resources, feedback and structure, and by embracing these technologies, “educators can foster creativity in teaching and learning and optimize time and effort” (2023, p. 180). Also of note, van den Berg & du Plessis (2023) claim these emerging tools “leveled the playing field” (p. 2) by providing teachers from all diverse backgrounds with “educational support and equal opportunities to access resources, research

articles and lesson plans” (p. 10). Generative AI can also impact student teachers and teachers alike through submission of lessons to receive feedback from LLMs, as well as evaluating and critiquing lessons generated by AI, equipping educators with critical thinking and problem-solving skills imperative for success in the 21st century (van den Berg & du Plessis, 2023). In summary, utilizing AI in education will benefit teachers in time conservation, development of skills and enhance their ability to connect and engage with students.

Concerns

Though generative AI offers many benefits for teachers and their students, valid concerns regarding its usage persist in the educational community. The most notable concern, brought to light by Goodman et al., is the lack of suitability for generated lessons to be utilized in the classroom. In the study, he found that the investigation portion of his lesson was described in only a few sentences, and lacked the specific content and details needed, highlighting the need for critical thinking and careful review of AI-generated materials (2024). Kehoe emphasized this need, stating that though AI-generated lessons may be sufficient, more editing must be completed by the teacher due to repetition and unsuitability for certain age groups (2023). More issues with AI-generations have been noted, such as unverifiable references that are not real and inability to meet a particular word count without repetition of information (Kehoe, 2023). Additionally, users of AI must understand that AI’s knowledge base comes from the dataset on which it has been trained; therefore, LLMs have the potential to discriminate against students if the trained data are not representative of all students (Long & Magerko, 2020 & Su & Weipeng, 2023). Because software lacks the ability to convey the emotional support or human interaction needed in the teaching process, lessons and materials cannot include these necessary components and teachers must integrate them prior to teaching. (van den Berg & du Plessis, 2023). Furthermore,

generative AI may widen the digital divide because some schools have banned ChatGPT, some countries do not have access to the technology, and price plans have divided the quality of use for individuals (Trust et al., 2023). Some LLMs are unable to provide content, diversity, or cultural perspectives in generated lessons, leaving it up to the teacher to determine the content taught and materials used (van den Berg & du Plessis, 2023). In conclusion, while the rise of generative-AI has shown promising benefits for educators, it must still be approached with a critical eye, considering all potential limitations for quality education.

Ethics

The integration of AI technologies into the classroom setting raises ethical concerns for students and teachers which must be addressed. Though utilizing generative AI in the classroom has been shown to provide large benefits to the educational community, this technology cannot replace the physical teacher who is essential to student's success in learning (Kehoe, 2023). However, Pearce claims generative AI should be used in the educational field if the risks of using the technology are addressed and teachers are properly trained to use it ethically and effectively (2023). As AI is integrated into the classroom, teachers must ensure that their personal, and student, data are secured, as this technology can store sensitive and personal data (Pearce, 2023). He goes on to call for "transparent communication between educators, teachers, parents and students," to ensure all parties involved have a clear understanding of what data is being collected and how it is managed, including any risks present (Pearce, 2023, p. 152). Additional research on the ethical usage of generative AI across specific grade levels and differing LLM platforms is needed, as highlighted by Gregory et al. and Kehoe (Gregory et al., 2024; Kehoe, 2023). Similarly, Trust et al. points to a need for warning against possible violations of the Family Educational Rights and Privacy Act, due to stored copies of data including protected

student information (Trust et al., 2023). However, research in the field of education is not progressing as quickly as others due to resistance from decision makers, which may be lessened by research showing the improvements in pedagogical usage and need for additional information around ethical considerations (Celik et al., 2022). Finally, people must understand that humans and AI have similar, but extremely different capabilities; therefore, educators must make informed decisions on when to trust and use AI in terms of creativity, emotional reasoning and social interaction (Long & Magerko, 2020). Overall, educators must work with members of their educational community to navigate ethical concerns to secure the responsible use of AI.

Current Use of AI in the Field

Generative AI has recently become a point of interest for implementation in the field; however, the field still lacks information pointed towards preservice teacher perspectives and usage of AI (Celik et al., 2022). “Authors have briefly mentioned or referred to aspects of lesson planning, but no research articles that focus on GPT-generated lesson planning and the implications for teacher education could be found” (van den Berg & du Plessis, 2023, p. 2). Therefore, listed below are sections describing the current state of AI in content areas, lesson planning, assessments, data, and differentiation.

Content Areas

One of the most important parts of an educator's role is determining the lesson plans for teaching students content every day. Many frameworks have been researched and put into practice over the years, but a few are detailed below which will be used in the study at hand. First, when planning the actual lesson, many educators use backward design, a concept for planning while keeping the end goal in mind (Wiggins & McTighe, 2006). Dr. James Popham comes next, with the six key points of understanding needed to develop assessment literacy,

allowing educators to create assessments that truly evaluate student understanding (2018). Finally, the PASS framework will be used to implement differentiation into lessons, which allows for each student's learning needs to be met (Mastropieri & Scruggs, 2002; Scruggs & Mastropieri, 1995).

Methodology

This study will utilize mixed methods research to collect and analyze data sources. Mixed methods are defined as “a procedure for collecting, analyzing, and “mixing” both quantitative and qualitative methods in a single study or series of studies to understand a research problem” (Creswell & Plano Clark, 2018, p. 545). Using a combination of quantitative and qualitative data is more impactful than individually looking to each type for understanding of results. In my study, this type of research is used because one type of data collection is not capable of addressing the research questions presented, and developing alternative perspective helps provide a more in depth understanding of the results (Creswell & Guetterman, 2019). Rigorous methods will be used to increase the validity of the study. A combination of recording and transcribing conversations, using initial and advanced analysis, and using various forms of data collection will advance the quality of the study.

The four data sources will be collected simultaneously, the results will be collected and merged for comparison and interpreted or explained through a convergent design mixed methods approach (Creswell & Guetterman, 2019). The purpose of this mixed-methods approach is to use the two different methodologies to “offset the weakness of the other form,” so that “a more complete understanding of a research problem” can be discovered (Creswell & Guetterman, 2019, p. 551). Therefore, data sources are gathered simultaneously, data are analyzed, and finally compared to see if the results support or conflict with one other. This design type adds strength to

the study itself, because qualitative data provides information about context and setting, while quantitative provides information about the generalizability of information discovered.

This study utilizes a self-study research design, which can be defined as process in which educators review their own practice “with the purpose of improving it, making explicit and validating their professional expertise and, at the same time, contributing to the knowledge base of teacher education” (Vanassche & Kelchtermans, 2015, p. 508). Self-study research draws directly from personal experiences, which would include my experiences using AI to generate instructional materials during this study. While self-study can be completed alone, a critical friend increases the validity and trustworthiness of the study (Samaras, 2011). In addition, it abides by four main categories: critical collaborative inquiry, improved learner, transparent and systematic research process, and knowledge generation and presentation (Samaras, 2011). To improve learning, researchers must understand and improve my own work as a teacher in transparent and systematic research processes to “necessitate a disposition of openness to outside views, questions, and critiques” (Samaras, 2011, p. 11). Furthermore, I must integrate a systematic and transparent research process to ensure the effective collection of data and resources during a self-study-based research process (Samaras, 2011). The final component, knowledge generation and presentation, is relevant to “personal, professional, program and school development” (Samaras, 2011, p. 11). Therefore, self-study is not only about me, conducted by only myself, or includes only personal knowledge, but the advanced opinions of those around myself helping to improve my work. By conducting a self-study, I am allowing new ways to view professional accountability by changing demand, making outside sources more meaningful, powerful, and sustainable when self-motivated (Samaras, 2011). Using self-study, I can immediately apply the implications to my students and teaching practices. By conducting

this study, I can evaluate my own practice and help schools benefit from teachers motivated by inquiry of the efficiency of their own practices.

Another key aspect of self-study is the implementation of critical friendship. Critical friends are trusted colleagues who seek support and validation of their research to gain new perspectives in understanding and reframing their interpretations (Samaras, 2011). They provide two roles, offering critique and provide support, and using that to create a working structure with one another (Samaras, 2011). One key definition of self-study includes the implementation of collaborative critical inquiry, which is personal thinking and learning, with the addition of critical friend's viewpoints to obtain alternative perspectives (Samaras, 2011). This allows educators to form opinions that best meet students' needs on their own, instead of by popular influence. In this way, teachers can find and become the change they seek in education.

Context

The study was completed in a mid-Atlantic United States regional city, where I attend the local university. The study was conducted during my Intern 1 semester of student teaching in the university's Elementary Education Bachelor of Science program. The teacher preparation program is noted for its excellence in producing informed teachers with classroom experiences frequently, starting with the first class in the program (ECU COE, 2024a). To graduate from the program, students complete 120 semester hours, some of those dedicated to a concentration area, where students dive deeper into an area of interest, such as academically gifted, psychology, or social studies (East Carolina University, 2024). During the two-semester long internship, preservice teachers are given the opportunity to apply their prior knowledge from course experiences, develop problem-solving skills in the classroom, and obtain "professional growth by observing, planning, and teaching with a veteran, clinically trained teacher" (ECU COE, 2024b). Through early teaching experiences and the benefits of internship, preservice teachers

who graduate from the program are equipped with a toolbox of information and strategies for success in the classroom.

The classroom environment being considered during this study is in a mid-Atlantic United States rural community. The school district serves approximately 24,000 students, with the district population being made up of 175,000 individuals (North Carolina Department of Public Instruction, 2024). In this community, 58% of the population is White individuals, 36.5% Black individuals, 8% Hispanic or Latino, and the remaining percentages encompass individuals who are Asian, Hawaiian or Other Pacific Islander, or two or more races in ethnicity (United States Census Bureau, 2025). The school is home to 611 students, of which 64.6% are economically disadvantaged (North Carolina School Report Cards, 2024). The student population I am teaching in my 2nd grade classroom consists of nineteen students, of which seven are males and twelve are females. The students come from a variety of ethnic backgrounds, which include: nine white or Caucasian students, seven black or African American students, one Hispanic student, and two students of two or more races.

Participants

The study is viewed through a lense that encompasses my background experiences and culture. During my childhood, I attended a rural private elementary school, but I currently intern in public school settings. Currently, I attend a university in mid-Atlantic United States, identified as a rural or urban community. While completing the study, I am completing student teaching at a rural elementary school, serving a second-grade population of students. Working with these students, and other classes in the past, gives me insight into classroom practices and hands-on teaching opportunities.

As I completed the study, I used the following skillsets and knowledge to make informed observations and decisions. Through courses such as Differentiated Instruction in Elementary Grades, and my work in our universities Assistive Technology Lab as the student manager, I have gained valuable knowledge about working with diverse needs of individual learners. I am well-versed in the latest accommodating resources to meet student learning needs and disabilities that impair classroom abilities, and strategies to help students of various learning groups, such as high-incident learners and academically gifted students. In my outlined plan of study, I have mastered planning with the end in mind through backwards design and using the ABCD method to write strong and effective lesson objectives. Using these learned skills, I have successfully learned to write and implement lesson plans that include differentiated content for learners, content plans that accurately address all information from course standards and develop the parts of a 6-step or 5E lesson plan. In terms of AI, I am level 3 certified MagicSchool AI, meaning I can properly navigate the website, understand safe AI practices within schools, and received basic training on how to communicate with the LLM (MagicSchool, 2024). These skills and knowledge learned helped me to effectively complete the study.

Procedures

There are five phases I followed when conducting this study.

Phase 1

Prior to collecting the data, steps were taken to gather background information on the topic of interest and plan each detail of the study. A review of the literature was conducted to better understand the current implications of educator usage of AI to generate instructional materials. After gaining the proper knowledge, I worked alongside my mentor to design the study, receiving feedback and making revisions along the way. Based on the university teacher

preparation program internship, the grade level was explicitly laid out for the study. However, content areas, standards, learning objectives, and weeks to plan the unit were strategically selected to add rigor to the research taking place. Finally, before the start of material generation, I reviewed the key information taught throughout the teacher preparation program, and completed four tiers of MagicSchool AI certification trainings, ensuring my ability to productively use the technology and evaluate outputs.

Phase 2

To prepare for the start of data collection, I first wrote journal entries in response to prompts analyzing personal perceptions of using AI to generate instructional materials and my pedagogical decision about utilizing the technology. The journal entries were useful when comparing how my perceptions have changed throughout the study. Next, I participated in critical friend conversations to prepare myself for the start of data collection. In these conversations, my critical friend challenged me to think critically and enhance my studies for the best outcome. Finally, I transcribed the audio recording of the critical friend conversation to begin data collection.

Phase 3

To begin the collection of unit materials, prompts for MagicSchool AI were generated. These prompts include specific standards and literacy instruction goals for the students in a second-grade classroom. After creating the prompts, they were next implemented into MagicSchool AI, leading to an evaluation of the generated materials. The materials were scored based on a rubric reflecting all necessities for a competent rubric given to students. Based on the scoring of the rubric, lessons and materials were accepted, adapted, or trashed to start over from the beginning. If the output needed edits, the process was repeated until an acceptable outcome

was achieved. After creating acceptable lessons, assessments, and materials, weekly journals were written, recording implications of each AI generation session. These journals not only reflect my successes, but the specific faults I found in the AI generation throughout its usage. Finally, weekly critical friend conversations were conducted and transcribed to add to qualitative data that was coded to identify patterns and results.

Phase 4

After collecting data, journal entries were written based upon prompts that bring to light experiences while working through the study. These journal entries detail my developing perceptions of the AI used, and how it may benefit or hinder future teaching practices moving forward into the future. Furthermore, critical friend conversations were held and recorded to push my thinking surrounding results in the study. My critical friend used data and validated information to help guide my thinking and push me to develop ideas beyond my limited scope of practice. After these conversations, the recorded session was transcribed to assist in qualitative data collection further in the research process.

Phase 5

Finally, post data collection results were collected and analyzed. Using open and axial coding, all qualitative data sources were analyzed. Following this, quantitative data sources were analyzed utilizing descriptive statistics. Using the collection of qualitative data and quantitative data results, the data was triangulated to identify common themes and any discrepancies in the data. After the data was analyzed and compared, the results, or findings, from the data were identified and added to the paper's written format. Following the result analysis, a discussion of the findings was written and revised, followed by editing and finalizing the paper. After the completion of these steps, the study is ready for revision and publication.

Data Collection

In this study, I collected four types of data. The quantitative data source is the evaluation of AI-generated materials using an evaluation rubric and numeric scores. The three qualitative data sources are AI-generated materials, critical friend conversations, and personal journal entries. The AI-generated materials were derived from MagicSchool AI and are the focus of journal entries and conversations with a critical friend.

Data Source 1 – AI-Generated Materials

Using MagicSchool AI, a 2-week integrated unit was created for a second-grade classroom. For each lesson in the unit, a lesson plan, assessment, and instructional materials needed were created using generative AI. I created 10 lesson plans, 10 materials, one to support each lesson, and one assessment with an accompanying rubric. These lesson materials were generated using pre-planned prompts and chosen standards input into the large language model. To keep track of these collected instructional materials needed, the integrated unit is planned in a graphic organizer (see Appendix A). The responses were saved and included in the integrated unit graphic organizer after they were created. Generating these materials allowed me to better understand what the large language model can produce and how these materials can be used in the classroom.

Data Source 2 – Evaluation of AI-Generated Materials

After MagicSchool AI generated the instructional materials needed for the 2-week integrated unit, they were analyzed against the current qualifications needed for effective use in the classroom. The assessments, lesson plans, and materials must align to standards or learning objectives outlined in the specifications of the requested AI output. They were also assessed in areas of content accuracy and grade-level appropriateness for students. Furthermore, all

generated instructional materials were assessed for varied approaches to learning offered for students of different needs and learning styles. Finally, including and accurately explaining academic vocabulary was assessed in each assessment, lesson plan, or instructional material produced. Each assessment, lesson plan, and instructional material generated was evaluated against one single evaluation rubric to determine its scores in the categories above (see Appendix B). In total, 47 rubrics were used to produce the 17 AI-generated materials for the study.

Data Source 3 – Personal Journal Entries

As materials were generated and analyzed using the evaluation rubric, I recorded personal journal entries to collect qualitative insights on the study. Journal prompts were used to create a pattern of structure for pre-data collection, during data collection, and post-data collection (see Appendix C). While generating resources and analyzing outputs, I journaled my thoughts and responses to findings presented throughout the study. Throughout the study, I completed 5 journals, which contained a total of 2,611 words. These journal entries helped me to reflect on personal experiences and changes in pedagogical thinking throughout the study's progression. As I combined scores from evaluation rubrics, I compared this data to understand my perceptions of using AI to generate elementary classroom materials and make well-supported conclusions about my pedagogical decisions centering around AI usage as an educator.

Data Source 4 – Critical Friend Conversations

An important part of self-study, a critical friend was used to strengthen the validity of my research and offer a secondary insight into my findings (Samaras, 2011). During weekly critical friend check-ins, conversations were recorded and later transcribed into a document used for analysis. To guide critical friend conversations, prompts for pre-data collection, during data collection, and post-data collection were used in each meeting to assess the critical aspects of the

research process needed for qualitative analysis (see Appendix D). Prompts targeted developing ideas and challenges faced in the stage of research will help provide insight to changing perceptions of using artificial intelligence to generate classroom materials for an integrated unit. Further, the prompts assisted in determining a shift in pedagogical implications for future classroom practices. During the study, I participated in 4 conversations, in which my critical friend and I spoke 13,199 words over 87 minutes. The transcribed text from conversations answering these prompts assisted in establishing answers to my research questions supported by many other data sources.

Data Analysis

For this study, I used two forms of data analysis. For my quantitative data collected from the evaluation rubrics, I utilized descriptive statistics (i.e., frequencies, mean, median, mode, and range) to determine the effectiveness of the generated materials. For my qualitative data from the generated materials, journal entries, and critical friend conversations, I used open and axial coding to identify themes and subthemes that represent my pedagogical decisions surrounding and perceptions of the AI-generated materials. I triangulated my data across all sources to confirm the applicability of my findings.

Trustworthiness

To ensure this study's validity and trustworthiness, many rigorous methods are used throughout the study. When completing the proposed methodology for the study, I worked with a faculty mentor and researcher at the university. This mentor checked behind all aspects of the proposed study, ensuring that ethical considerations are met, and the research met all guidelines needed to ensure its validity. Research reflexivity or considering my own context in the study and how that is taken account in conducting my research has been completed prior to the

beginning of the study. In doing these things, I strive to reduce any potential biases found in the study prior to its beginning.

To maintain the study's trustworthiness, I ensured an audit trail of all consent forms and data sources are maintained. Data sources are maintained through physical journal logs, recordings, and stored materials in a drive specific to the study. When transcribing information collected in the data sources, information is de-identified to protect the privacy of all participants, in compliance with both study trustworthiness and educational practices. The drive in which all resources and data in relation to the study are stored is password-protected, or in a locked filing cabinet to ensure no tampering has occurred. To strengthen trustworthiness, the findings are presented as rich descriptions with examples from the research data to support all claims made. The combination of the various, rigorous strategies has ensured the trustworthiness of my study.

Findings

After completing the study, several themes emerge from the data analysis. Looking at quantitative analysis methods first, I give a review of AI outputs. Moving on to findings from qualitative analysis, I first address my changing perceptions of using AI to generate elementary classroom materials. Then, I address the material generation process and how the materials met alignment standards outlined in the evaluation rubric. Finally, I address my pedagogical decisions, informed from the process of completing the study and analysis of data sources.

AI Outputs

The quantitative data collected in the study were analyzed to produce insights into the effective usage of AI in the classroom. In the sections below, findings are presented in four categories: assessments, lesson plans, materials, and overall data. In alignment with frameworks

used for the study, the assessment was created first, then the ten lessons in pairs of two, and finally the ten materials. Each AI-generated output received a score based on an evaluation rubric, drawing from important frameworks used addressed in the study. Further, each of the seventeen items generated for the unit has a tracked log of decisions made with each output, which are to: discard the output, accept it, revise it, or adapt the material myself.

Lesson Assessments

Following the Backward Design process created by Wiggins & McTighe, the summative assessment for the thematic unit was generated first (2006). The assessment directions were generated using Raina Chatbot, and the assessment rubric was created with the rubric generator tool. On average, there were 1.5 prompt revisions, as the assessment directions were revised 3 times, and the rubric required no revisions. The average initial rubric score was 12.5 out of a possible 15 points, reflecting the highest initial rubric score average out of the three types of data collected. During the process of generating assessments, I chose to adapt the output one time, revise the output twice, accept the output twice, and chose not to discard any outputs.

Lesson Plans

When generating lesson plans using Magic School AI, three different tools were used: the Lesson Plan, 5E Model Lesson, and the Science Lab. I asked each tool to create two separate lessons, which were built off one another over separate days. I used the Lesson Plan and 5E Model Lesson tools twice and the Science Lab tool once; therefore, there are ten lessons total, but only five sets of AI prompts. The Lesson Plan tool averaged 3.5 prompt revisions, while the 5E Model Lesson and Science Lab both had an average of 3 revisions (see Table 1). Additionally, the Science Lab tool rubric average score is 13.25, while the Lesson Plan tool averaged 10.8, and the 5E Model Lesson rubrics average a score of 12.17. Across the ten lessons, each lesson was

revised 3.48 times, and there was an overall total of 11 prompt revisions. The average initial rubric score was 9.8 out of a possible 15 points. During the process of generating lesson plans, I chose to adapt the output three times, revise the output eleven times, and accept the output five times. I chose to discard my very first output received, because I began to realize the extent of information should be included within a prompt.

Table 1

Lessons Output

Tool Used	# of Prompt Revisions	Alignment to Standard	Content Accuracy	Grade-Level Appropriate	Differentiation	Academic Vocabulary	Rubric Total
Lesson Plan	3.50	2.60	1.80	1.80	2.00	2.60	10.80
5E Lesson	3.00	2.73	2.37	2.37	2.67	2.03	12.17
Science Lab	3.00	2.25	2.75	2.75	2.75	2.75	13.25

Lesson Materials

When generating materials to accompany lesson plans, I used nine different tools: Mini Passages, DOK Questions, Image Generator, Common Misconceptions, Text Leveler, Informational Texts, Worksheet Generator, Clear Directions, and Math Story Word Problems. Ten materials were created, one for each lesson plan. On average, each lesson material was revised 1.28 times, and the average initial rubric score was 13.33 out of a possible 15 points (see Table 2). Looking at initial materials outputs, seven led to prompt revision, and three were accepted as is. Only one output was revised more than twice, due to running out of a free trial using a paid feature of MagicSchool AI. Looking at specific tool usage, the Math Story Problem tool was used to create two different materials. The first usage of the tool received a nine according to the rubric and was revised one time, whereas the second initial prompt used with the tool resulted in a perfect rubric score. The Mini Passages tool received a perfect score

according to the rubric on its initial prompt generation, followed by common misconceptions, which scored 14 out of 15 on the initial material generation. The Image Generator tool required 3 prompt revisions, whereas the other tools used to generate materials only required up to two revisions. Overall, the materials received the best scores according to the rubric and required the least amount of prompt revisions.

Table 2

Materials Output

Tool Used	# of Prompt Revisions	Alignment to Standard	Content Accuracy	Grade-Level Appropriate	Differentiation	Academic Vocabulary	Rubric Total
Mini Passages	0.00	3.00	3.00	3.00	3.00	3.00	15.00
DOK Questions	1.00	3.00	3.00	2.50	2.00	3.00	13.50
Image Generator	3.00	2.00	1.67	3.00	-	-	12.67
Common Misconceptions	0.00	3.00	3.00	3.00	2.00	3.00	14.00
Text Leveler	2.00	3.00	3.00	2.33	2.00	3.00	13.33
Informational Texts	2.00	2.33	2.33	3.00	1.00	2.33	10.99
Worksheet Generator	2.00	2.67	2.67	3.00	1.67	3.00	13.01
Clear Directions	1.00	-	3.00	3.00	-	2.00	14.00
Story Word Problems (2)	0.50	3.00	2.75	2.50	2.50	2.75	13.50
All Materials Average	1.28	2.78	2.71	2.81	2.24	2.79	13.33

All Generated Material

Looking at the data obtained from the generated lessons, materials, and assessment, several findings emerge. On average, initial rubrics scored 10.59 out of 15 points, and 12 of 17 initial rubrics led to a decision to revise the prompt used. On average, each item generated in the study had an average of 1.82 prompt revisions, as a total of 22 prompt revisions occurred. Aside from revisions, I adapted the item myself eight times, trashed the item one time, and accepted 17 items according to the rubric. Specifically focusing on the rubric, all items generated using

MagicSchool AI averaged a score of 2.7 out of 3 with alignment to standards, 2.48 for content accuracy, 2.54 for grade-level appropriateness, 2.26 for differentiation, and 2.6 for academic vocabulary. The mean total of all rubrics used in the study is 11.52 out of a possible 15 points, according to the rubric.

Table 3

All Generated Materials

Generated Material Type	Alignment to Standard	Content Accuracy	Grade-Level Appropriateness	Differentiation	Academic Vocabulary	Rubric Total (out of 15)
Assessment Directions (Initial Rubric)	3.00	2.00	2.00	1.00	2.00	10.00
Assessment Directions Average (All Rubrics)	3.00	2.50	2.25	2.00	2.50	12.25
Rubric Average (Initial Rubric)	3.00	3.00	3.00	-	3.00	12.00
Rubric Average (All Rubrics)	3.00	3.00	3.00	-	3.00	12.00
Lessons Average (Initial Rubric)	2.00	1.60	2.20	2.00	2.00	9.80
Lessons Average (All Rubrics)	2.62	2.33	2.38	2.62	2.52	12.48
Materials Average (Initial Rubric)	2.30	2.40	2.60	1.30	2.30	10.90
Materials Average (All Rubrics)	2.68	2.62	2.76	1.93	2.72	12.72
Mean	2.70	2.48	2.54	2.26	2.60	11.52

Changing Perceptions

In this section of the findings, I address the development of my changing perceptions of AI throughout the study. First, I address the capability of AI to be used in the classroom and its limitations in terms of the elementary classroom. Next, I address my comfortability of using AI, and its growth through trial-and-error processes. Then, I discuss my own ability to produce quality outputs using MagicSchool AI and critically evaluate them against educational standards.

Moving on, I address my findings based on the future of AI in the field of education. Finally, I addressed the topic of prompt writing, in which I spent much time during the study.

AI Capability

My perceptions of MagicSchool AI's capability shifted throughout the study. In this section, I will discuss two main topics which address the capabilities of AI to generate materials for an elementary classroom setting. First, I will address my shifting perception of how AI should be used in the classroom. During this section I will explain why I now believe AI should be used as a tool in the classroom. To conclude the section, I will discuss the various limitations of MagicSchool AI's use in this context, as well as how I overcame or worked around them.

AI is a Tool. When I first began the study, I believed AI to be nearly perfect and capable of creating a full lesson using one tool alone. However, I found that an abundance of tools is needed to create a full lesson. MagicSchool AI, in its current model, cannot produce any materials alongside the lesson it creates. For example, no assessment, content, directions, slides, etc., are included in the lesson and, therefore, must be created separately and added back in. The LLM can create these outside materials, such as assessment directions, with much greater ease and useability within a lesson. Therefore, I found MagicSchool AI to be a useful tool in creating the building blocks to insert within a lesson, rather than creating the lesson in full by itself.

Using MagicSchool AI as a tool could ease my workload as a teacher and the time spent creating supporting materials for lessons. For example, I was able to quickly produce math story problems for students which centered around the topic of our thematic unit. Taking less than 5 minutes to create a prompt, generate the problems, and review each critically, I am confident that this process took much less time than developing each story problem on my own. By strategically using MagicSchool AI to create supplemental materials, I am confident that time

spent planning lessons and materials for my students can be significantly reduced. By saving time on the creation of materials, it can lead to more time for other crucial aspects of my job in the classroom.

Limitations. While working with the LLM to compile materials for a lesson, it became clear that AI requires a human to compile information. As stated earlier, lesson materials must be created separately from the AI-generated lesson itself. Therefore, I must review the generated lesson and determine where additional materials and information are required, such as the actual content to be taught or examples to connect with the students. While this is the same process followed in stage three of the Backward Design model, it sometimes felt more challenging due to not having determined each additional piece needed myself as I worked through lesson planning. After determining what must be created, and generating it separately, I must compile the various pieces together in order to create a clear, rigorous lesson for my students. During the study, I completed this process ten times, creating just one material to add into each of the AI-generated lessons, finding that many items must be created separately and added into the lesson for it to make sense to the students. During conversation with my critical friend, I explained my frustration by saying, “I did not expect how much I was going to have to go back and forth with the AI, feeling like it was not giving me the amount of information I was wanting” (Critical Friend Conversation 3). I felt this frustration with the lack of information included within each lesson because it leads to many more steps the teacher must complete, analyze and integrate within the incomplete MagicSchool AI-generated lesson plan.

One large limitation found during the study is that AI is not perfect and cannot read my mind. Although I never truly believed the LLM could read my mind, I often found it very difficult to communicate in the prompt exactly what I was looking for in an output. While

reflecting on my experiences while generating lesson plans, I stated, “AI will never be able to do exactly what I have pictured in my head and, therefore, can make me hesitant to ask AI to fix anything else” (Journal 3). During this time, I felt frustration because asking for changes sometimes was completely ignored and I was unable to express exactly what I needed for the LLM to understand and produce it. Further, I found that AI struggles with the context of a lesson and the students it is generating materials for. While in every prompt I included the age of students being taught in the lessons, the output sometimes included information or activities not of age-appropriateness for my students. For example, one lesson asked for my “students to research online,” which is not a skill they have mastered, or are ready to, at this point in the school year.

Further, MagicSchool AI-generated materials must all be critically analyzed and revised before use in the classroom. In my final journal, as I reflected over my experiences, a main takeaway was that “educators should always look through the materials before presenting [them] to their students” (Journal 5). Aside from analyzing and revising what the LLM has produced, more work must be done before the lesson or material is classroom ready. During the study, I found that MagicSchool AI currently cannot facilitate classroom management or the accommodation of student learning needs in its outputs, which are a critical piece of every lesson taught. Therefore, I believe using the time saved by utilizing MagicSchool AI as a tool could lead to ample amounts of time for critical revision and customizing each lesson to meet the needs of every student within my classroom.

Comfortability using AI

During this portion of the findings, I will discuss my change in comfortability using AI to generate elementary classroom materials. I will first address my changing confidence levels,

which improved significantly during the study. Then, I will discuss the setbacks I faced throughout the study, and how I either overcame them or observed and documented them. Next, I will talk about my hands-on experiences using AI, and the processes which I found most effective in producing useable outputs. Finally, I will address my changing ability level as I grew in knowledge and experiences over my time learning and using the technology.

Confidence. During my time completing the study, my comfortability using MagicSchool AI drastically increased due to practice and gained experiences. Beginning the study, my confidence in myself to effectively use the LLM to create effective lessons and materials was almost non-existent. However, after conversations with my critical friend, and drawing off knowledge from my research, coursework, and professors, I began to trust my ability to critically analyze, revise, and accept materials created with MagicSchool. Looking towards the end of the study, I “feel comfortable with MagicSchool as a whole,” but feel like I am experimenting with the technology rather “than trying to become proficient at a certain tool” (Journal 4).

My confidence in using the LLM does not extend to every aspect used in the study. While I am very comfortable using MagicSchool AI to create materials and parts of a lesson, I am not confident in my ability to create an entire lesson. During a conversation with my critical friend reflecting on the study, I emphasized this by saying “[MagicSchool AI] will be better for instructional support of materials, assessments, and rubrics, versus actually planning the lesson” (Critical Friend Conversation 4). However, I can recognize that my development of prompt writing skills led to the enhancement of initial AI-generated outputs, creating a more detailed and aligned lesson plan (see Figures 6-7). My experiences in the study led me to an optimistic outlook for using AI in my everyday educator career. The level of confidence I developed in a

mere few months of using the platform has also led to a feeling of excitement in using the LLM to generate new, fun ideas to use within the classroom for students.

Figure 6

Initial AI-Generated Lesson Plan

The Impact of Culture on Technology in America

Grade Level: 2nd Grade

Subject: Social Studies

Lesson Overview:

Students will explore how various cultural groups have contributed to technological advances in America. They will learn about the importance of diverse cultures and how these contributions have shaped modern life compared to the past.

Duration: 1 hour

5E Lesson Plan

Engage (10 minutes)

- Begin with a short video or presentation showing different technologies (e.g., smartphones, cars, computers) and ask students to share how these technologies help them in their daily lives.
- Prompt questions:
 - "What is your favorite piece of technology?"
 - "How do you think life would be without these technologies?"
- Introduce the idea that many different cultures have helped create these technologies.

Explore (15 minutes)

- Divide students into small groups and provide each group with pictures and short descriptions of technological contributions from various cultural groups (e.g., the wheel from ancient Mesopotamia, the compass from China, the internet from various innovators).
- Ask groups to discuss:
 - What is [the technology](#)?
 - Which culture created it?
 - How does this technology help people today?
- Groups will share their findings with the class.

Explain (10 minutes)

- Discuss the concept of culture and its importance in society. Define culture and give examples (food, art, music, technology).
- Explain how the contributions from different cultures have influenced American life. Highlight specific examples discussed in the Explore phase.
- Show a chart or infographic that illustrates the timeline of technological [advances](#) and the cultures involved.

Elaborate (15 minutes)

- Have students create a "Culture and Technology" poster. Each student will choose a cultural group and a technology they contributed.
- The poster should include:
 - The name of the cultural group
 - A drawing or picture of the technology
 - A few sentences explaining how this technology impacts our lives today.
- Allow students to present their posters to the class.

Evaluate (10 minutes)

- Conduct a class discussion reflecting on the lesson:
 - "Why is it important to recognize different cultures?"
 - "How do these technologies change our lives?"
- Assess student understanding through their posters and their participation in discussions.

Aligned Standards:

- **Social Studies Standard:** 2.1.1 - Explore the contributions of various cultural groups to American society.
- **Technology Standard:** 2.3.2 - Understand the role of technology in daily life and its impact on society.

Figure 7

Revised AI-Generated Lesson Plan

Final lesson plan one and two:

Exploring Native American Contributions to Technology in North Carolina
Lesson Overview
 This two-part 5E lesson plan focuses on how Native American tribes have influenced and contributed to technology in North Carolina today. Students will explore the technological advancements made by various tribes and understand the broader impact of these contributions on society.
Part One: Introduction to Native American Contributions

Engage

- **Objective:** Spark interest in Native American tribes and their technological contributions.
- **Activity:** Show students a video or slideshow highlighting Native American inventions, such as the canoe, agricultural tools, and the use of natural resources.
- **Teacher Script:** "Today, we will watch a short video about Native American tribes and some cool things they invented! Pay close attention to how their inventions helped people long ago and even today."

Explore

- **Objective:** Investigate specific tribes and their contributions.
- **Activity:** Divide students into small groups and assign each group one of the following tribes: Coharie, Eastern Band of Cherokee Indians, Haliwa-Saponi, Lumbee Tribe, Meherrin, ~~Saponi~~, Occaneechi Band of the Saponi Nation, or Waccamaw Siouan. Provide each group with differentiated texts about their assigned tribe's contributions to technology.
- **Differentiation:**
 - **Above Grade Level:** Provide additional resources, such as articles or videos on advanced technological contributions.
 - **Below Grade Level:** Use simplified texts with illustrations to support understanding.
- **Teacher Script:** "In your groups, you'll read about your tribe's inventions. Use the text to find out how they helped people. Remember to look for the word 'technology!'"

Explain

- **Objective:** Share findings with the class.
- **Activity:** Each group presents their tribe's contributions to the class, using visual aids such as posters or drawings.
- **Teacher Script:** "Now, let's share what we've learned! Each group will present their tribe and one or two important inventions. Use the word 'impact' when explaining how these inventions changed lives."

Elaborate

- **Objective:** Connect Native American contributions to current technology.
- **Activity:** Discuss how the technological advances of Native American tribes have influenced modern society. Students write a short paragraph explaining this connection.
- **Sentence Starters for Below Grade Level:**
 - "A Native American tribe that helped with technology is the. *They invented* which changed ."
- **Teacher Script:** "Think about how these inventions are still important today. Write a few sentences about one tribe and their impact."

Evaluate

- **Objective:** Assess understanding of the lesson.
- **Activity:** Collect written paragraphs and provide feedback based on the use of academic language (technology, Native American, tribe, society, impact).
- **Teacher Script:** "Hand in your paragraphs. I will check if you used the key words we talked about!"

Part Two: Rigorous Extension on Technology Contributions

Engage

- **Objective:** Review previous knowledge and prepare for deeper exploration.
- **Activity:** Quick recap of the previous lesson using a KWL chart (What I Know, What I Want to Know, What I Learned).
- **Teacher Script:** "Let's remember what we learned about our tribes! We'll fill out our KWL chart together."

Explore

- **Objective:** Conduct research on a specific technological innovation from a chosen tribe.
- **Activity:** Students select one technological contribution and research its historical context, usage, and relevance today.
- **Differentiation:**
 - **Above Grade Level:** Encourage students to explore multiple resources and create a presentation.
 - **Below Grade Level:** Provide graphic organizers to help them structure their research.
- **Teacher Script:** "Choose one invention that interests you. Find out when it was invented, how it was used, and how it is important today."

Explain

- **Objective:** Present research findings.
- **Activity:** Students create a visual presentation (poster, PowerPoint, or brochure) showcasing their findings.
- **Teacher Script:** "Now, create a visual to share your research. Include the words 'society' and 'impact' in your presentation."

Elaborate

- **Objective:** Discuss the cultural significance of technological contributions.
- **Activity:** Hold a class discussion on how these contributions reflect the culture of Native American tribes and their relevance today.
- **Teacher Script:** "Let's talk about why these inventions are important to the culture of Native American tribes. How do they show creativity and innovation?"

Evaluate

- **Objective:** Assess the understanding and application of knowledge.
- **Activity:** Students will write an essay explaining how a Native American tribe has contributed to technology today, using at least two academic vocabulary words.
- **Teacher Script:** "Write an essay about one tribe's contribution to technology. Be sure to use at least two of our key words!"

Aligned Standards

- **Social Studies Standard:** NCSS.2.1 - Understand the contributions of various cultural groups in North Carolina.
- **ELA Standard:** CCSS.ELA-LITERACY.W.2.2 - Write informative/explanatory texts to examine a topic and convey ideas.

This lesson plan is designed to foster an appreciation for cultural contributions and their impacts on society while encouraging critical thinking and collaboration among students.

Setbacks. As mentioned above, I entered the study wary of my current knowledge level and skillset. While I was doubtful of my own qualifications as a pre-service teacher, I overcame this setback as I began to generate and evaluate materials. During our second conversation, my critical friend summarized my developing feelings towards my ability to create and accept

materials. She stated, “It is not so much that you do now know how, it is that you are getting to a point where you are becoming confident in saying ‘Yes, this is solid, onward we go’” (Critical Friend Conversation 2). Further, I found it challenging to find and use best practices while generating materials. While I created a plethora of materials using the LLM, I used many tools, instead of refining my skills while using one particular tool. When properly generating and reviewing materials, critical analysis is used consistently. To create materials that correspond with a generated lesson, I had to be extremely intentional with developing prompts, ensuring that the material aligned to the lesson itself, critically reviewing and revising, and creating a flow that introduced new materials seamlessly within a lesson. From the beginning of the study, I noted my frustration with the technology because, “it can produce better information, but I am not realizing that because of my prompts” (Critical Friend Conversation 1). Though I improved drastically in my ability to write effective prompts during the study, I found prompt generation to be the most challenging piece of this process.

Hands-on Experience. I found that hands-on experiences led to the increase of my comfortability levels while using AI and of ability to efficiently create and analyze outputs. During these experiences, I used a trial-and-error process, having back and forth conversations with the LLM. When reflecting on my experiences after the study, I stated that “the back and forth, trial-and-error process takes an adjustment period,” however, “with time and practice, using AI in teaching practices is an outstanding improvement to the field” (Journal 5). I found that spending time learning how the technology responded to various wording and placement within prompts had great effects on the output received, and I adjusted my prompt writing process accordingly. The time spent learning how the technology responds was frustrating but ultimately rewarding as I experienced the desired results.

My Abilities. Prior to completing this study, I was given no guidance in how to utilize AI in my teaching career during my pre-service teacher training. Yet, during the short time of study completion, my AI literacy drastically improved. For instance, as I began to create materials to accompany each lesson, I quickly navigated to tools within the platform I knew could assist the lesson, and knew how to include needed information, ranked by priority, within the prompt. As stated earlier, my prompt writing and output analysis became easier with the growth of knowledge I gained from trial-and-error processes. Though I completed MagicSchool AI training by taking courses through the website, they did not include guidance on how to create strong prompts or utilize varying tools.

The Future of AI in Education

In a technologically driven society, I believe AI will continue to grow, especially within the field of education. After reviewing the literature to develop an understanding of where the field currently stands and completing a study of my own, I have developed two beliefs about the future of AI in education. First, I will address the capability of AI to ease teacher workload, based on my observations documented while generating classroom materials. Then, I will talk about the capability, or lack thereof, for AI to replace traditional classroom teachers in an elementary school setting.

Ease Teacher Workload. My knowledge of AI's ability to ease teacher workload expanded and strengthened throughout the study. While I believe AI will help with a reduction in teacher workload, it cannot take away crucial teacher-led tasks, such as teaching lessons and developing full, accurate lesson plans. However, while conducting the study, I found AI can assist with the automation of tasks. Simple tasks, such as creating an email to send to parents, generating some fun jokes to hook student interest into a lesson, or creating a rubric to align with

an assessment, can be automated by using Magic Tools. During a conversation with my critical friend, I reflected on my excitement to use tools specifically designed to create “that specific thing I was looking for” in half the time, if not more (Critical Friend Conversation 4).

AI Will Not Replace Teachers. Based on my experiences using AI to generate classroom materials, I do not believe AI will replace teachers. Though AI is able to produce useable materials quickly and correctly, I was not able to produce a full, useable lesson. Each lesson created by AI lacked important resources and materials, which must be generated and added into the lesson separately. For instance, lessons 9 and 10 needed an abundance of sample story word problems for examples within the lesson. However, the LLM did not provide these samples initially for the teacher, who must develop them separately. Further, each output required critical evaluation, from someone like me, who has received training on age-appropriateness of content, depth of information students must know, and alignment to standards. Therefore, even as AI continues to grow, I will always be required to critically evaluate and compile any output produced by an LLM.

Prompt Writing

Prompt writing proved to be a more rigorous process than I originally planned for, yet my skill level rapidly progressed as I created each set of instructional items. First, I will discuss the many setbacks I faced when learning to write effective prompts. Then, I will address the effective strategies used to combat these setbacks, producing usable outputs. Finally, I will discuss the revision of prompts, accomplished through critical evaluation.

Information Learned. From the first prompt written to the last, the development of my prompt writing skills drastically improved. The use of a trial-and-error strategy led to several findings which proved to be crucial in effective prompt writing. The placement of information

within the prompt is key. I learned to place the most important information at the top of my prompt, as the LLM directed most attention to matching those specifications. For example, including that I wanted two lessons, the second a rigorous extension of the first, must be placed at the very beginning of the prompt in order for the command to be followed. Further, putting information in all capital letters and asterisks drew the LLMs attention to that specific information in the prompt, even if closer to the bottom.

After testing different methods, I found that creating categories of prioritized information was the most successful, along with using specific and targeted language. When talking with my critical friend, I shared that I found “putting important information at the top and sub-details that correspond even lower in the prompt” worked best when writing long prompts (Critical Friend Conversation 3). Finally, the difference between open and directed prompts makes a large difference in the output received. When using specific Magic Tools, the prompts are targeted, asking specific information that will help with the development of the item. However, when using an open prompt, like Raina Chatbot, I had to know what information the LLM required to produce the desired output.

Setbacks. I faced many setbacks while generating lessons and materials in relation to my prompt writing. Most importantly, the lessons lacked a depth of information and detail in relation to what the prompt asked. For example, when attempting to expand upon the information initially provided, and explicitly asking for it within the prompt, the new output showed no growth. Not only did it fail to expand on the content to be taught, but it also lacked a deeper explanation of how activities would be conducted. Further, the LLM would occasionally skip over parts of the prompt. For example, when creating one of my lesson pairs, the “LLM

[skipped] over crucial information in the prompt, such as adding an entire second lesson” (Journal 5).

As I began the process of developing prompts, I often made ineffective prompt revisions, in which case the output did not reflect the changes desired. However, as I neared the end of the study, generating materials became an automated process. Though revisions were required, the recognition of what changes to make easily came to mind. Before moving into the generation of lesson materials, I stated “I have felt confident in my ability to assess the lessons, knowing it doesn’t have this or that” (Critical Friend Conversation 3). When generating lessons seven through ten, I began to copy and paste a general set of directions into the prompt, as I knew a basic set of information the LLM was looking for (see Figures 8-9). Building off prior knowledge and gaining experiences, through trial-and-error my knowledge of prompt writing grew tremendously.

Figure 8*First Prompt*

Initial Prompt:

Grade Level – 2nd

Topic, Standard, or Objective - How have various culture groups influenced and contributed to America (focus on technological advances)?

Additional Customization: The students will learn about culture, and how people of various cultures have created technology that has changed the way we live. It will show the importance of all cultures, because each group has made important tech advances. Technology allows these students to live a very different life, which can be compared with the lives of young people back in the day. The lesson will support people's contribution to society.

Standards set to Align:

- W.2.6 - Recall information from experiences or gather information from provided sources to answer a question.
- K2-IC-01 Compare how people live and work before and after the implementation or adoption of new computing technology.
- 2.B.1.2 Explain how belief systems of various indigenous, religious, and racial groups have influenced or contributed to culture in America

Figure 9

Final Prompt

Revised Prompt:

<p>Grade Level: 2nd</p> <p>Topic, Standard, or Objective:</p> <p>Use the following Standards:</p> <ul style="list-style-type: none"> • NC.2.OA.1 Represent and solve addition and subtraction word problems, within 100, with unknowns in all positions, by using representations and equations with a symbol for the unknown number to represent the problem • RI.2.5 Know and use various text features to locate key facts or information in a text efficiently. • PS.2.2.2 Use models to summarize the relationship between sound and how sounds are produced and detected by parts of the body that vibrate. <p>The lessons are part of a themed unit on culture. Create TWO DISTINCT, DETAILED, SEPERATE LESSONS:</p> <ul style="list-style-type: none"> - Lesson One: Using a number line and a drum, students will solve and answer single-step addition problems, with at least 80% accuracy. Start the lesson by reviewing with students what they have learned thus far in the thematic unit about culture and its impact on North Carolina today. Then, transition to talking about the drums the students created in the last lesson, to represent the importance of music in Native American culture. "A number line is a helpful tool that can make adding numbers easier! To use a number line, first draw a straight line and mark points on it, like 0, 1, 2, 3, and so on. If you want to add two numbers, like 2 and 3, start at the first number, which is 2. Then, move to the right on the number line. Each step you take to the right counts as 1. So, if you take 3 steps from 2, you will land on 5. This means that 2 plus 3 equals 5! Using a number line helps you see how numbers are added together." Tell the students we will use our drums to help us with our number lines today. Students will use a number line, and "jump" each step on the line. When they count each jump, students will beat their drum one time. EXPLAIN AND MODEL this expectation for students, in teacher script. Example: "Our problem is two plus two. So, I am going to start on the right side of the number line, at 2. Then we are going to draw our two jumps, one two. Now, let's count our jumps with our drums. One, two! Good job! So, where did we land? Four. That's right!" Practice this as a group, with table groups, and then independently. Each word problem should be focused on the context of Native American tribes that reside in North Carolina, or culture. Example: The Lumbee tribe wants to build four new tents. They already have 3. How many tents will they have all together? - Lesson Two: Using a number line and a drum, students will solve and answer single-step addition and subtraction word problems, with at least 80% accuracy. Review how students used the drum in the previous lesson. Tell the students we will use our drums to help us with our number lines today. Students will use a number line, and "jump" each step on the line. When they count each jump, students will beat their drum one time. EXPLAIN AND MODEL this expectation for students, in teacher script. Example: "Our problem is four minus two. So, I am going to start on the left side of the number line, at 4. Then we are going to draw our two jumps, one two. Now, let's count our jumps with our drums. One, two! Good job! So, where did we land? Two. That's right!" Practice this as a group, with table groups, and then independently. Each word problem should be focused on the context of Native American tribes that reside in North Carolina, or culture. Example: Dakota went hunting for his family and brought home eight deer. He shared three of his deer with Jacy. How many deer will Dakota take home to his family? <p>Add academic language into the lessons and define them in 2nd grade language.</p> <ul style="list-style-type: none"> - Academic language terms: Native American, Addition, Subtraction, Music, Number Line - Use the academic language in the lesson, have students interact with it in meaningful ways. <p>Differentiate the Assessment Questions: Give TWO assessments for EACH lesson. One should be written at a 2nd grade reading level, and one should be written at a middle-of-year 1st grade reading level.</p> <p>Additional Criteria:</p> <ul style="list-style-type: none"> - The first lesson should set students up for their second, more rigorous, lesson. - Give students a structured brain <u>break</u> during a transition in the lesson. - Model all activities for students in second grade language. - Directions for how to write the lessons: <ul style="list-style-type: none"> o Use kid-friendly language to write the lesson plans. o GIVE DETAILED, STEP-BY-STEP directions for all activities. o Give at least 5 problems for students to solve in the final assessment and explicitly state all math problems for students to solve. o Use A LOT of detail to explain what the teacher and students will be doing in each part of the lesson. o Add explicit directions in "teacher script" for all areas in the lessons where the teacher gives direction to students. o Write higher-order questions to ask students during the lesson. - Specifications for the lessons: <ul style="list-style-type: none"> o Add differentiated word problem text selections for students who are on-grade level, below grade-level, and above grade-level (INDEPENDENT PRACTICE ONLY). o The work will be recorded in their "Tribe Investigator Notebook" <p>Standards Set to Align To: North Carolina Standard Course of Study</p> <ul style="list-style-type: none"> • NC.2.OA.1 Represent and solve addition and subtraction word problems, within 100, with unknowns in all positions, by using representations and equations with a symbol for the unknown number to represent the problem • RI.2.5 Know and use various text features to locate key facts or information in a text efficiently. • PS.2.2.2 Use models to summarize the relationship between sound and how sounds are produced and detected by parts of the body that vibrate.
--

Strategies. Generating lessons and materials using MagicSchool AI proved to be a multi-step process. Following the process of backward design, the generation of the 2-week, integrated themed unit initially focused on the end goal. I first generated the assessment and rubric, followed by the lessons, and finally, the materials. Further, each lesson was developed using

Backward Design by focusing on what students should master by the end of each lesson. Aside from this strategy, trial and error proved to be the most successful when generating materials. In my typical lesson planning process, I complete one section of a lesson at a time, before moving on to another. However, with MagicSchool it is “back and forth the entire time,” working on all pieces of the lesson at once (Critical Friend Conversation 3). While at first my levels of confidence wavered, gained experiences by working back and forth with the LLM allowed me to learn effective strategies in prompt writing.

Revisions. After writing each prompt, I critically evaluated the generated output to make a decision about the next step. After the evaluation, I made an informed decision by choosing one of four options of how to proceed: accept the output as it is, revise the prompt, adapt the output myself, or trash the output and start over. Each output must receive critical evaluation to ensure its alignment with the pre-determined standards, differentiation needs, age-appropriate vocabulary, and other factors. While some outputs did not require revision, I know they still must be evaluated before being presented to students in the class. For instance, Material 10, a set of story word problems, required no revisions before being presented to the students (see Figure 10). Yet, I reviewed each problem even still, to ensure the problems made sense and the math being asked of the students was developmentally appropriate and content aligned. This was an active choice I made, to ensure the validity and reliability of information students are taught and tested on.

Figure 10

Math Story Word Problems

Native American Music: Subtraction Word Problems for 2nd Grade

Set 1 (Numbers up to 50):

73. The Cherokee tribe had 45 drums for their annual music festival. They lent some to the Lumbee tribe. Now they have 28 drums left. How many drums did they lend to the Lumbee tribe?
74. The Sappony tribe collected 50 feathers to make musical instruments. After making some rattles, they had 37 feathers left. How many feathers did they use to make the rattles?
75. There were 39 people at the Haliwa-Saponi music gathering. Some people left early. Now there are 26 people. How many people left the gathering?
76. The Coharie tribe had 42 singers in their choir. After a performance, some new members joined. Now they have 50 singers. How many new members joined the choir?
77. The Siouan tribe made 33 flutes for a music class. Some flutes were given to students. Now they have 19 flutes left. How many flutes were given to students?
78. The Catawba tribe had 47 people learning a traditional dance. Some people completed the course. Now 29 people are still learning. How many people completed the course?
79. The Waccamaw tribe collected 36 gourds to make musical instruments. After making some rattles, they had 22 gourds left. How many gourds did they use to make rattles?
80. There were 44 children in the Cherokee music class. Some children were absent today. Now there are 38 children in class. How many children were absent?
81. The Lumbee tribe had 41 traditional songs in their collection. They learned some new songs. Now they have 50 songs. How many new songs did they learn?
82. The Sappony tribe had 48 dancers for their cultural performance. Some dancers couldn't make it. Now they have 35 dancers. How many dancers couldn't make it to the performance?

Set 2 (Numbers up to 20):

83. The Cherokee tribe had 18 drums for their music class. They lent some to the Lumbee tribe. Now they have 12 drums left. How many drums did they lend to the Lumbee tribe?
84. The Sappony tribe collected 20 feathers to make musical instruments. After making some rattles, they had 14 feathers left. How many feathers did they use to make the rattles?
85. There were 15 people at the Haliwa-Saponi music gathering. Some people left early. Now there are 9 people. How many people left the gathering?
86. The Coharie tribe had 13 singers in their choir. After a performance, some new members joined. Now they have 18 singers. How many new members joined the choir?
87. The Siouan tribe made 16 flutes for a music class. Some flutes were given to students. Now they have 10 flutes left. How many flutes were given to students?
88. The Catawba tribe had 19 people learning a traditional dance. Some people completed the course. Now 12 people are still learning. How many people completed the course?
89. The Waccamaw tribe collected 17 gourds to make musical instruments. After making some rattles, they had 9 gourds left. How many gourds did they use to make rattles?
90. There were 14 children in the Cherokee music class. Some children were absent today. Now there are 11 children in class. How many children were absent?
91. The Lumbee tribe had 12 traditional songs in their collection. They learned some new songs. Now they have 18 songs. How many new songs did they learn?
92. The Sappony tribe had 20 dancers for their cultural performance. Some dancers couldn't make it. Now they have 15 dancers. How many dancers couldn't make it to the performance?

Generation and Alignment

Looking at the process of generating the unit materials, I completed it in two stages. First, I will talk about the resources provided and LLM used to generate the thematic unit components.

Then, I will explain the multi-step process used to generate each of the three item types:

assessment, lesson, and material. After each output was received, time was taken to complete thorough critical evaluations. The critical evaluation, supported by evaluation rubrics and qualitative analysis was used to determine the alignment of each output to mandated standards.

Generation Process

The change in my perceptions and determination of pedagogical decisions is a reflection of my experiences during the generation process. A combination of resources and knowledge of the LLM created a foundation for generating each of the components for the two-week unit. After first mapping out the unit, I followed the backward design framework to determine the order of item creation: first the assessment and rubric, then lessons, and finally the supporting materials. The process of generation varied for each of the item types developed.

Resources. A variety of components allowed my process of generating items using MagicSchool AI to be possible. Knowledge gained through ongoing intern experiences, developing class knowledge, and continuing use of the NCDPI and NCSCOS made my use of the technology efficient. In addition, conversations with professors, my research mentor, and critical friend conversations became an integral resource. I demonstrated the need for these resources during a conversation with my critical friend where I asked for advice in utilizing resources to help with prompt generation, in which she helped me discover five different resources I already had in my teacher toolbox to pull from. My own sense of pedagogy, technological knowledge, and experience with content allowed for accurate use of the technology and assessment of the generated materials.

The completion of MagicSchool AI training courses one, two, three, and four were a helpful resource for website navigation, allowing for me to confidentially navigate the MagicSchool website and use the LLM. While the training did not help me learn the

specification of what each tool does, or how to write effective prompts for each, they did teach me how to locate resources within the website and provided strategies for locating specific tools needed. Looking back at research, and the knowledge gained from its completion, allowed me to navigate around known issues, such as incorrect math questions. I reflected on this during a conversation with my critical friend, as I pointed out that I have seen many “studies saying that being able to submit the prompts is important,” because the use of language can lead to the quality of output received (Critical Friend Conversation 1). Finally, the specifically trained tools that MagicSchool AI offers became great resources to use throughout the study. For instance, the rubric generator was a perfect tool for the assessment rubric, and the science lab tool worked better than the lesson planner tool would have for the specified lesson goal.

MagicSchool AI. When completing the online training resources, I learned MagicSchool AI is an LLM created by educators, for educators. For this reason, each of the 60+ specific tools is trained specifically to do its designated task, giving me a vast number of ways to develop materials for lessons, assessments, or communicative materials. However, some aspects of the LLM can only be accessed through a paid subscription, which I did not secure for this study. One of those benefits is the ability to work back-and-forth with a given output. Instead, the entire prompt had to be re-written, or edited, for me to change an aspect of the output which sometimes led to unnecessary and unwanted changes. Yet, when doing this, sometimes the entire lesson would change based on one re-written line of the prompt. When explaining my frustration with this issue, I gave the example of “I want to change one line in the lesson, but I cannot edit that in the prompt, because it does not know which lesson I am referring to” (Critical Friend Conversation 3). While this issue was frustrating, I believe a paid subscription, or utilizing the Raina Chatbot tool could solve some of the problems.

Process. At the completion of the study, one strategy shone throughout the generation process: trail-and-error. Working back-and-forth with the AI allowed me to acquire LLM-specific knowledge, such as how each tool works best. Each tool I used gave a quick response, never taking more than 30 seconds to create a full lesson or material. As I prepared to create prompts, factors such as age-appropriateness, accommodation of student needs, and ethical considerations stayed at the forefront of my mind. In a journal reflection during the generation of lessons, I state “AI also struggles with context, such as age-appropriateness and social emotional student learning needs,” meaning that this area within the prompt was often revised while generating materials (Journal 3). However, the combination of these factors allowed for a successful generation of seventeen items for the planned thematic unit. After generating each item, a pre-determined process of critical evaluation began. I assessed each material against an evaluation rubric, which lead to one of the following decisions: adapt the material, revise the prompt, trash the material, or accept the material as it is.

Assessment. Using the Backward Design process, the summative unit assessment was created based around the end goal of the lesson. This allowed me to focus on key elements of the thematic unit, instead of attempting to assess mastery of ten standards at once. When using Raina Chatbot to generate the assessment directions, I believe they may have been easier to create myself. This was because I “had to think up all of the ideas for [the assessment] myself,” and then explicitly explain the assessment to AI before asking for directions (Critical Friend Conversation 2). Already understanding what I needed to explain to students in the directions, I believe I could have written those directions out in the same amount of time it took to give the chatbot all the information needed to create directions and then complete the back-and-forth process. However, the rubric generator tool produced a rubric which met all pre-determined

standards according to evaluation standards on the first attempt. The difference in effectiveness could be due to the specific training the rubric tool received, versus using an openly trained chatbot.

Lessons. When creating the lessons for the thematic unit, it was done in pairs. Five pairs of lessons were generated, with a base lesson and extension of knowledge. The Lesson Plan and 5E Model Lesson tools each created two pairs of lessons, and the Science Lab was used for one pair. Reflecting on the creation of lessons, I was not fond of the Lesson Planner, but enjoyed the easy use of the Science Lab. While generating each lesson, I realized the difference in creating perfect lessons versus ones that are acceptable. Because I came to find that “AI cannot create a complete lesson by itself,” my focus shifted from a full, complete lesson, to one which covered each standard of the rubric and could be supported by additional instructional supports and materials (Journal 5). I developed the belief that AI cannot be expected to be perfect but can provide a lesson outline to be filled in with explicit content and materials.

Materials. The creation of materials to use within lessons was much easier than other processes. Each MagicSchool AI tool is created to perform a specific task, such as creating a short story or generating funny jokes to share with students. I generated ten materials, one for each lesson. Because of the many tools made available, I attempted to use a number of them. The criteria required for each prompt varied, making it hard to determine mastery of a tool based on one usage, or to transfer the knowledge of other prompt specifications across the platform. For example, the lesson planning tools asked for different criteria for creating a lesson plan, leading to more trial-and-error when learning what information each tool needed and how it would use that information. However, because each specific tool is trained to complete its task, the acquisition of material generation was much easier for me to master than generating lessons.

Alignment

While the MagicSchool AI was able to produce an acceptable output, the LLM often struggled with alignment. A common issue was leaving out crucial information in the lesson, such as the content to be taught. Each lesson came equipped with the bare bones needed for a lesson, without any detail or direction. As a preservice teacher, the length of lesson plans tends to be over ten pages, packed with all information and materials possibly needed. However, MagicSchool AI's lessons did not exceed two pages in length, giving an abbreviated plan. I described the plans as "brief and dry" to my critical friend, stating that the lessons are "not written in a way that is engaging the students" based on the way it is written (Critical Friend Conversation 3). For example, one set of lessons, which incorporates an art standard, begins with a read-aloud, shift to direct instruction while showing art examples, has students write about the art, and then read a passage and reflect. The lesson does not include any of the content, art examples, passages for students, or directions for learning, transitions, or assessment. While this lesson is adequate for the students, a much more engaging lesson that involved students interacting with art or creating their own would bring more student buy-in with the lesson. This concern highlights my need for AI-generated outputs to expand upon developed lesson plans with more knowledge and detail. Additionally, MagicSchool struggled to meet the goal of standards throughout the lesson or would leave out one of the integrated standards altogether. Without critical evaluation, I easily could have missed these errors and changed the rigor of the lesson plan. Therefore, I believe MagicSchool AI's alignment to the standard course of study was unmethodical.

Pedagogical Decisions

After completing the study, I reflected on my experiences and found three main areas of impact on my future pedagogical decisions. Of my three main reflections, this first is that I found a great need for AI training before I felt comfortable using varying LLMs. Although back-and-forth strategies will allow educators, like me, to become comfortable with the technology itself, foundational and ethical training will allow teachers to responsibly use the AI. Furthermore, AI is a tool to be added to my teacher toolbox, but it is incapable of replacing my role in the classroom. Moving forward, I will continue to use AI as a tool in my own pedagogical practices by creating materials to automate time-consuming tasks.

Need for AI Training

I was unable to produce an age-appropriate, content-aligned lesson without explicit training. Therefore, I believe professional development sessions would have benefitted me in learning to produce effective lessons. While I hold the belief that trial-and-error will continue to be the best strategy to learn AI's patterns and automate its usage, I also found a need for foundational training. To address this, I took time to learn how the LLM itself operates by taking all MagicSchool AI training offered on their website. In addition, my time conducting the study led to realizing the need for ethical training before using an LLM. In a conversation with my critical friend, I stated, "it is important for teachers to know from the very beginning that they cannot employ sensitive data for their students into the AI," but, "teachers would not know that [when] starting to learn how to" use the technology" (Critical Friend Conversation 4). The combination of foundational and ethical training gave me a strong sense of support to begin.

MagicSchool AI Training. MagicSchool AI offers training on their website that helps new users, like me at the start of this study, get comfortable with the layout of the website and its

features. However, the training did not teach me how to interact back and forth with the LLM. Looking back, it would have been helpful to view examples of how the LLM would respond to different pieces of prompts before starting the study. For instance, understanding the importance of prompt placement prior to using AI would have provided me with an advantage when creating a prompt. I reflect on this in a journal entry, stating that “I have not received training on how to formulate prompts that will give the output I desire,” therefore, “I believe teachers and pre-service teachers will need to receive training” (Journal 2). While I believe the back-and-forth process of using a trial-and-error strategy will continue to be the most successful learning form to gain comfortability and skill using AI, I possessed a need for the LLM to give foundational support, further than basic website navigation.

Ethical Training. Though I had a base knowledge of AI prior to taking on this study, I quickly realized the need for ethical training. One prevalent concern reflected in current literature surrounding the use of AI in the classroom is the protection of sensitive student data. I learned the ethical use of AI in the classroom follows an 80/20 approach. The 80/20 approach ensures that critical evaluation of materials is always included in the generation process by outlining that educators should allow AI to do 80 percent of the work but must always go back and complete the last 20 percent of the work themselves. Following this framework, I found that in most cases AI wrote the majority of each generated item, but I almost always found ways to improve the generated output through critical analysis. Because AI is not perfect, and can make mistakes, I found it unethical to deliver un-proofed content to students, or in this case, accept the generated output. With knowledge of the current literature in the field, I followed this approach, ensuring that critical evaluation was a key part of the generation process.

AI is a Tool

Based on my growing perceptions throughout the study, AI is best used as a tool. I learned that I cannot be reliant on AI to create everything I may need; instead, it is one of many tools in my teacher toolbox. In the classroom, I will use AI to generate parts of the lesson, like materials, instead of large lessons. While it gives me a great start in lesson ideas, MagicSchool AI currently cannot create a full lesson on its own, as noted earlier. Therefore, I found great benefit in learning how to utilize the tools AI offers to create pieces of a lesson, such as the materials. For example, MagicSchool AI easily produced story word problems, rubrics, and leveled text. In contrast, I found that using AI to create entire lesson plans became “time-consuming” while trying to use an abundance of knowledge and direction to “piece together the information needed” (Journal 3). Nonetheless, it was important that any tool used was critically analyzed before accepting the generated output.

Future AI Use

Personally, I know AI will be used as a resource in my future classroom. While it will not be used to attempt the creation of full lesson plans, it will be a helpful resource in automating small tasks, such as creating a catchy song to help students memorize content. I began to catch on to this idea early in the lesson generation process, reflecting that “using [MagicSchool AI to create] pieces [of a lesson] is showing itself to be the most effective strategy thus far (Journal 3). Because of the vast capabilities of MagicSchool AI, it is likely that it can be used daily in my pedagogical practices, to complete tasks such as creating word problems for students or leveling text selections. While the amount of AI-usage in my classroom may be limited at first, I believe as the technology develops, it will become even more impactful.

Discussion

Based on the findings, and current literature published in the field, there are four main topics of discussion. First, I discuss considerations educators must understand before utilizing AI in the classroom themselves, including the benefits, concerns, and ethical aspects. Moving on to the address the future use of AI in education, I discuss important topics including professional development needs, AI literacy, and critical analysis. Next, implications for future practice are addressed for all parties affected by educators using AI within the elementary classroom. Finally, I define the implications for future research, based upon the findings of the study.

Considerations

Reflecting on the findings of the study and current state of research in the field, there are clear draws and pushbacks for using AI in the field of education. The field of education will continue to grow, and utilizing AI in education has many perceived benefits to assist teachers, administrators, and students alike. However, the concerns surrounding AI use in the classroom are notable and must be addressed before educators begin using platforms such as MagicSchool AI. Ethical considerations play a key part in how educators should proceed when utilizing a LLM to complete instructional tasks and will be addressed below.

Benefits of AI

Due to the teacher shortage currently burdening our school system, reducing teacher burnout is of the utmost importance. As noted by Trust et al. (2023), LLMs have been found to benefit educators in the reduction of time spent creating classroom instructional materials. In my personal experiences using AI during this study, I found the quick response time of AI useful in automating simple but time-consuming tasks, such as leveling texts or creating decodable texts for students. As witnessed in my own study, and through the work of others, automating tasks

which take ample time in the classroom, such as leveling texts or creating supplemental materials with the assistance of AI has a reduction on teacher workload (Pearce, 2023; van den Berg & du Plessis, 2023 & Zhai, 2023). By providing teachers with a resource to assist in using their limited time more efficiently, AI has the potential to not only cut down on teacher overtime but, hopefully, reduce the rate of teachers experiencing burnout within the classroom.

MagicSchool AI is an LLM created by educators, for educators. For this reason, each tool is trained specifically to do its designated task. The platform offers over 60 specific tools, giving educators a vast number of ways to develop materials for lessons, or communicative materials (MagicSchool, 2024). For example, there are specific tools trained to properly level texts for students of different reading abilities. Working with these tools, the prompts ask for specified information that helps an educator understand the type of information needed by the LLM, though trial and error processes will still be needed. By providing teachers with these education-targeted tools, specifically the ones creating supporting materials, the time spent critically reviewing and revising AI-made errors can be reduced. Following the 80/20 approach, in which the generative technology completes eighty percent of the work, and the teacher completes the additional 20 percent, this back-and-forth process can lead to high-quality materials for students. Agreeing with the work of others, I believe that an output should be a starting point, from which teachers move onto “critically evaluate, refine, and modify” to ensure alignment with instructional goals and standards (Goodman et al., 2024, p. 2).

Concerns

While generative AI offers many benefits to teachers, there are some concerns which must be addressed. As pointed out by Goodman et al., one large concern surrounds the unsuitability of many AI-generated materials to be utilized in the classroom, such as the lack of

depth, content, and length of generated lesson plans (2024). Supporting his findings, I felt “the depth of information was also reflected in the number of pages,” because in these two-page, bulleted lessons, only the steps of completing the lesson are listed (Critical Friend Conversation 3). Critical components of the lesson, such as content, sample word problems, class directions, and all other instructional supports or materials must be developed separately and integrated into the generated lesson plan. Therefore, though LLMs such as MagicSchool AI can produce acceptable lesson plans in theory, they must be critically reviewed, edited, and revised before teaching to students of any age (van den Berg & du Plessis, 2023).

Further, some aspects of the LLM can only be accessed through a paid subscription, which was not secured for this study. One of those benefits is the ability to work back-and-forth with a given prompt. Because I was unable to communicate with the LLM about specific changes I wanted to make on a lesson, assessment, or material, I was required to edit the entire prompt and hope no additional changes were made to an output. However, when using a free trial of the Image Generator tool, I was able to communicate with the LLM about specific changes I wanted to make to the image, which proved to be very helpful in the process of critically reviewing and revising generated materials. In relation to prompts, one large concern is the ability for educators to write effective prompts which, therefore, lead to receiving a high-quality output from AI. As mentioned, prompt writing proved to be one of the most challenging aspects of my experiences using AI and should be a topic on which educators are trained before attempting to use the technology themselves.

Technology will continue to advance forevermore; therefore, the education system should get in on the ground floor and train their teachers and administrators alike how to effectively utilize this technology. However, prior to using the technology, teachers must receive ethical

training. Trust et. al highlights in his research that possible violations of the Family Educational Rights and Privacy Act are of risk, due to storage of protected student information (2023). To combat this issue, school administrators must offer their teachers professional development sessions for generative-AI training. These sessions should not only give teachers an overview of what AI is, and how it can be used to support them, but also how to ensure all students and their data remain protected. Although some teachers already use this technology in their classrooms, receiving effective AI-based training prior to using an LLM within the school setting must be considered. Aside from the protection of student data, the educational integrity of teaching staff must also be ensured. Without these trainings, the use of AI within the school setting is a large concern, as both educators, students and the school itself can become at risk.

Ethical Considerations

Looking towards the ethical considerations of using AI in the elementary classroom, it is imperative that educators are aware of and actively avoid such conflicts. As argued by Pearce, it is my belief that AI should be used in the field, under the condition that teachers become aware of the known risks of utilizing such technology and are trained on its ethical usage (2023). Having completed extensive research on AI ethics prior to working with MagicSchool AI, I actively chose to exclude sensitive information from prompts and critically reviewed each output received. Therefore, other educators, like me, should also be given ethical training before interacting with AI to minimize ethical risks.

Future Use of AI in Education

Teachers should have access to LLMs such as MagicSchool AI within their classroom. As a pre-service educator, I will continue to utilize generative AI to generate elementary materials for my classroom. Because I have identified AI as a tool to support the lesson planning process,

my usage of AI will center around crafting materials to support pre-planned lessons, such as leveling text selections, generating story word problems, etc. As noted in my second Critical Friend Conversation, I will continue to use the rubric generator tool because I was able to participate in “One ask, [received] one output, [and then I was] done.” Because I was able to give this tool ample information, and it simply had to put that information together, the tool provided an output that required no revisions after completing a critical evaluation. I will not be utilizing this technology to create full lesson plans, because through the study I have learned the back-and-forth processes are time consuming and often frustrating. Asking MagicSchool to create a full-length, aligned lesson plan is currently out of the LLMs scope, as the model often struggles to include relevant detail; however, using the model to spark ideas for lessons to be built upon is an acceptable practice I will use in my education career.

AI literacy is a tool which allows educators to critically evaluate, effectively communicate and collaborate, and use AI technologies in an ethically productive manner. Based on the principles of AI literacy, researchers and myself alike agree that teachers must work with generative AI to balance interpersonal interactions and standard-aligned content within materials (Celik et al., 2022; Pearce, 2023; van den Berg & du Plessis, 2023). Because LLMs are unable to include emotional support or human interaction within the teaching process, it is imperative they are added by the teacher, following the 80/20 approach (van den Berg & du Plessis, 2023). In the short time I completed the study, “my AI literacy increased drastically,” as I critically evaluated, modified, and added onto assessments, materials, and lessons (Critical Friend Conversation 4). Looking to the future of AI in education, I concur with studies that highlight the benefit pre-service educator’s experienced from learning how to use generative AI in their educational practices (Gregory et al., 2024).

Due to the level of critical analysis, modification, and integration of materials educators must apply to produce and use an AI-generated lesson in the classroom, I am confident that a need for certified teachers in the classroom will remain. Supporting Kehoe's (2023) assertion that AI's benefits to the educational community do not amount to the replacement of a physical teacher, I reflect on the inability of AI to foster interpersonal communication or social-emotional learning. Further, some LLMs are unable to provide the content or cultural perspectives needed to broaden student understanding of the world around them, leaving that up to the teacher to develop (van den Berg & du Plessis, 2023). As noted above, AI is not perfect and therefore must be critically analyzed. In my own experiences using MagicSchool AI, I found errors in several generated materials, such as age-inappropriate tasks and false information. Therefore, trained educators must be present in the classroom to critically review generated materials, add missing content, and provide interpersonal communication.

Implications for Future Practice

Teachers, students, administrators, and parents alike will experience the effects of efficient generative-AI usage within the classroom. Using time-saving processes, teachers will benefit from a reduction in workload and gain critical output-evaluation skills that are relevant not only to the classroom, but also to skills needed to be successful in the 21st century. As for the students receiving these classroom materials, an increase in engagement, time spent with the teacher, and personalization of content will greatly impact their ability to succeed. With a reduction in time spent creating lessons and materials, teachers will be able to properly differentiate, incorporate student interests, and broaden the content and materials students are exposed to. Additionally, administrators will be impacted by this study, as it is critical for the implementation of professional development opportunities for teachers to learn how to ethically

and professionally use generative AI in their classrooms. Finally, the community will be impacted by these changes, as students who make up the next generation will receive stronger educational training. Because teachers have more time to differentiate student needs and incorporate culturally inclusive and diverse materials, students and the community will benefit.

Implications for Future Research

As generative AI continues to develop, future research must be conducted to determine the ever-changing concerns and benefits of the technology. Currently, the field of education is working to develop and publish ample research in utilizing AI as a tool within the classroom. Furthermore, during a review of the literature in 2024, no research could be found in regard to MagicSchool AI's usage as a tool in generating classroom materials. Therefore, a need for researchers, administrators and educators alike to conduct and present research on the effectiveness of using AI to create classroom materials, and effective lesson plans, is of the utmost importance. Looking to the future, a research topic of interest would be the effectiveness of AI training, in regard to both its usage and ethical considerations. Furthermore, research comparing the effectiveness of a variety of LLMs, such as ChatGPT and MagicSchool AI, in creating quality educational materials would be a topic of note. Overall, there are many avenues of research which could be developed based off the topic of AI in education, as a dire need for research currently exists within the field.

Conclusion

Developing generative AI platforms, such as MagicSchool AI, has paved a new way for educators to conduct their classroom practices. Throughout this mixed methods self-study, I used MagicSchool AI to generate elementary classroom materials for a two-week integrated instructional unit, assessed the quality of the AI-generated materials, and developed my

pedagogical decisions based on the findings. Using my data sources, I conducted a quantitative data analysis, consisting of descriptive statistics, and a qualitative data analysis, consisting of open coding and axial coding to identify themes. The findings addressed the generation process of utilizing MagicSchool AI to generate classroom materials, professional development opportunities, best practices for using AI in the elementary classroom, and my pedagogical decisions based on the quality of the materials. Ultimately, the findings of this study point out the acquisition of MagicSchool AI into the teaching planning routine has a great potential to reduce teacher burnout, reduce time for developing materials, and increase time for other crucial aspects of their role.

References

- Celik, I., Dindar, M., Muukkonen, H., & Jarvela, S. (2022). The promises and challenges of artificial intelligence for teachers: A systematic review of research. *TechTrends*, 66, 616-630. <https://doi.org/10.1007/s11528-022-00715-y>.
- Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Thousand Oaks, CA: SAGE.
- Creswell, J. W., Guetterman, T. C. (2019). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (6th ed.). Pearson.
- East Carolina University. (2024). *Acalog ACMS*. <https://catalog.ecu.edu/index.php>
- East Carolina University College of Education (ECU COE). (2024a). *Bachelor of science in education*. <https://education.ecu.edu/elmid/elementaryeducationbs/>
- East Carolina University, College of Education (ECU COE). (2024b). *Office of clinical experiences*. <https://education.ecu.edu/oep/clinical-experiences/oce-intern/>
- Goodman, J., Handa, V., Wilson, R. E., & Bradbury, L. U. (2024) Promises and pitfalls: Using an AI chatbot as a tool in 5E lesson planning. *Innovations in Science Teacher Education*, 9(1). <https://innovations.theaste.org/promises-and-pitfalls-using-an-ai-chatbot-as-a-tool-in-5e-lesson-planning/>
- Gregory, K. H., Luterbach, K., Lin, X., & Sconyers, S. (accepted). Elementary educators' perceptions of the ethical use of ChatGPT as a classroom assessment generator. In H. Crompton & D. Burke (Eds.) *Artificial Intelligence Applications in K-12: Theories, Ethics, and Case Studies for Schools*. Routledge.

- Kehoe, F. (2023). Leveraging generative AI tools for enhanced lesson planning in initial teacher education at post-primary. *Irish Journal of Technology Enhanced Learning*, 7(2), 172-182. <https://doi.org/10.22554/ijtel.v7i2.124>
- Long, D., & Magerko, B. (2020). What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. Association for Computing Machinery. 1-16. <https://doi.org/10.1145/3313831.3376727>
- MagicSchool. (2024). *MagicSchool AI 3.0*. [Large language model] <https://www.magicschool.ai/>
- Ng, D. T. K., Leung, J. K. L., Chu, S. K. W., & Qiao, M. S. (2021). Conceptualizing AI literacy: An exploratory review. *Computers and Education: Artificial Intelligence*, 2, 100041. <https://doi.org/10.1016/j.caeai.2021.100041>
- North Carolina Department of Public Instruction. (2024). *North Carolina Public Schools Statistical Profile*. <http://apps.schools.nc.gov/ords/f?p=145:15:::NO:::>
- North Carolina School Report Cards. (2024). *Ayden Elementary School*. <https://ncreports.ondemand.sas.com/src/school?school=740304&year=2024>
- Open AI. (2022). *ChatGPT 3.5* [Large language model] <https://chat.openai.com/>
- Pearce, N. (2023). Revolutionising modern teaching with AI technology. *Education Journal Review*, 25(3), 148-154.
- Popham, W. J. (2020). *Classroom assessment: What teachers need to know* (9th ed.). Pearson.
- Popham, W. J. (2018). Wrapping up, reaching out. In *Assessment literacy for educators in a hurry* (pp. 123-137). *Association for Supervision & Curriculum Development*,
- Samaras, A. P. (2011). *Self-study teacher research: Improving your practice through collaborative inquiry*. *SAGE Publications, Inc.*

- Su, J. & Weipeng, Y. (2023). Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education*, 6(3), 355-366.
<https://doi.org/10.1177/20965311231168423>
- Sun, J., Ma, H., Zeng, Y., Han, D., & Yunbo, J. (2023). Promoting the AI teaching competency of K-12 computer science teachers: A TPACK-based professional development approach. *Education and Information Technologies* 28, 1509-1533. <https://doi.org/10.1007/s10639-022-11256-5>.
- Trust, T., Whalen, J., & Mouza, C. (2023). Editorial: ChatGPT: Challenges, opportunities, and implications for teacher education. *Contemporary Issues in Technology and Teacher Education*, 23(1), 1-23. <https://www.learntechlib.org/primary/p/222408/>
- United States Census Bureau. (2025). *Quick facts: Pitt County, North Carolina*.
<https://www.census.gov/quickfacts/fact/table/pittcountynorthcarolina/PST045223>
- van den Berg, G. & du Plessis, E. (2023). ChatGPT and generative AI: Possibilities for its contribution to lesson planning, critical thinking and openness in teacher education. *Education Sciences*, 13(10), 998-1010. <https://doi.org/10.3390/educsci13100998>
- Vanassche, E., & Kelchtermans, G. (2015). The state of the art in self-study of teacher education practices: A systematic literature review. *Journal of Curriculum Studies*, 47(4), 508–528.
<https://doi.org/10.1080/00220272.2014.995712>
- Wiggins, G., & McTighe, J. (2006). *Understanding by design* (Expanded 2nd ed.). Pearson Education, Inc.
- Zhai, X. (2023). ChatGPT for next generation science learning. *SSRN Electronic Journal*.
<https://doi.org/10.2139/ssrn.4331313>

Appendix A**Overview of AI-Generated Materials**

Week	Monday	Tuesday	Wednesday	Thursday	Friday
1	*Lesson *Material	*Lesson *Material	*Lesson *Material	*Lesson *Material	*Lesson *Material
2	*Lesson *Material	*Lesson *Material	*Lesson *Material	*Lesson *Material	*Lesson *Material *Summative Assessment

Appendix B

Evaluation Rubric

Criteria	3 Points (Excellent)	2 Points (Satisfactory)	1 Point (Needs Improvement)
Alignment to Standards	Lesson, assessment, or materials fully align with the standard/learning objective.	Lesson, assessment, or materials partially align with the standard/learning objective.	Lesson, assessment, or materials do not align with the standard/learning objective.
Content Accuracy	Information is accurate and presents a clear understanding of the topic without errors.	Information is mostly accurate with minor errors or omissions that do not impede understanding.	Information contains several inaccuracies or misconceptions that could mislead students.
Grade-Level Appropriate	Content, language, and activities are perfectly suited for 2 nd grade students' cognitive development.	Content, language, and activities are generally suitable but may not fully engage or challenge all students.	Content, language, and activities do not match 2 nd grade students' cognitive development and interests.
Differentiation	Provides varied approaches to learning that cater to different student needs and learning styles.	Some attempts at differentiation are evident, but not all student needs may be met.	Little to no differentiation is provided, making it difficult for some students to access the material.
Academic Vocabulary	Includes and accurately explains all relevant academic vocabulary, enhancing student understanding.	Includes most of the relevant academic vocabulary, but explanations may lack clarity or detail.	Academic vocabulary is either not included or poorly explained, hindering student comprehension.

Appendix C

Journal Entries Protocol

Pre-data collection prompt(s):

- ◆ What are my perceptions about AI?
- ◆ How comfortable am I with utilizing AI to generate information?
- ◆ What are my drawbacks about using AI?
- ◆ What are my perceptions about using AI to generate classroom instructional materials?
- ◆ What do I predict my pedagogical decisions surrounding using AI will be?
- ◆ What challenges do I foresee while using AI?

During data collection prompt(s):

- ◆ How comfortable am I with utilizing AI to generate information?
- ◆ What drawbacks or challenges am I facing while using AI?
- ◆ What are my developing perceptions about using AI to generate classroom instructional materials?
- ◆ How is AI changing my pedagogical decisions surrounding technology in the classroom?
- ◆ What benefits am I experiencing while using AI in my teaching practices?

Post data collection prompt(s):

- ◆ How did my pedagogical decisions surrounding using AI to generate instructional materials change?
- ◆ How comfortable am I with utilizing AI to generate information?
- ◆ What considerations must I be cautious about when using AI?
- ◆ How can I use what I have learned to shape my future teaching experiences?
- ◆ What are my perceptions about using AI to generate classroom instructional materials?
- ◆ How have my perceptions about using AI changed throughout this study?

Appendix D

Critical Friend Conversations Protocol

Pre-data collection prompt(s):

- ◆ What are my perceptions of using AI to generate instructional materials over a two-week integrated unit?
- ◆ How can my critical friend support me in my study? What advice might she have for my next steps?
- ◆ What do I predict my pedagogical decisions surrounding using AI will be?
- ◆ What challenges do I foresee while using AI?
- ◆ How will my personal biases and viewpoints shape my decisions about AI?

During data collection prompt(s):

- ◆ How is the information I am learning shaping my perceptions of AI? AI in the classroom?
- ◆ What questions do I have that my critical friend can provide guidance on?
- ◆ How do my personal biases and viewpoints shape my decisions about AI?
- ◆ What pedagogical decisions surrounding using AI are changing?
- ◆ What challenges am I facing while completing this study?

Post data collection prompt(s):

- ◆ How have my perceptions about using AI changed throughout this study?
- ◆ How did my personal biases and viewpoints shape my decisions about AI?
- ◆ What are my pedagogical decisions surrounding using AI to generate elementary instructional materials?
- ◆ How will the new information I learned shape future decisions?
- ◆ What information is important for me to share with others about my findings?