"New" UROP Proposal

Title of Proposal: Improving Math Learning Outcomes for High School Students with Learning Disabilities through AI-Assisted Argumentative Writing

STATE THE PROBLEM/TOPIC

Math is an essential skill for high-demand scientific and technological careers, ultimately affecting competitive workforce development in the US. However, math is also a difficult subject, where students often lack the personalized resources and support to engage, motivate, and deepen their understanding.
This struggle is particularly pronounced for students with learning disabilities (LD). The 2019 NAEP Mathematics Report Card reports that an astonishing 75% of 12th-grade students with disabilities performed below the NAEP Basic level. Despite intrinsic factors like neurodiversity impeding math progress for students with LD, research indicates that students with LD can achieve comparable results with their non-LD peers when empowered with individualized scaffolding, feedback, and assessment (Jitendra et al., 2018; Kiru et al., 2018). To promote Diversity, Equity, and Inclusion in education, it is critical to explore effective solutions that help address the disparities and challenges that historically marginalized groups like high school students with LD face when learning math. Among existing learning sciences evidenced by learning effectiveness, argumentative writing (AW) stands out as a unique and promising avenue.

Despite the successful and wide adoption of AW in math, science, and literacy education in traditional classrooms (Graham et al., 2020; Casado-Ledesma et al., 2021; Newell et al., 2011), barriers exist to prevent an effective integration of AW for students with LD – notably the difficulty in meeting their diverse needs and evaluating their learning outcomes with AW at scale. AW is a type of writing that aims to convince others using the knowledge of persuasion. AW encourages deep engagement with content knowledge, fosters critical thinking and reasoning skills, and guides students in systematically articulating their thoughts (Ferretti et al., 2009). Prior studies investigating AW in math learning showed that it was a reliable and effective way to enhance mathematical reasoning and conceptual understanding for high school students with LD (Graham et al., 2020; Kiuhara et. al, 2023; Powell et al., 2017). Despite its benefits, the majority of high school students with LD lack the ability and proper guidance to craft convincing arguments (Powell et al., 2017). AW is also commonly evaluated using human assessment, which poses scalability challenges (Powell et al., 2017). An effective solution to this problem is Artificial Intelligence (AI), which can greatly help automate the assessment of AW and organically integrate AW instruction to help high school students with LD learn math more effectively.

There has been extensive work investigating AI in education (AIEd) to automatically direct and evaluate students' learning through adaptive learning, corrective feedback, and automated grading (Chiu et al., 2023). However, most of the current AIEd efforts focus on enhancing students' procedural fluency while downplaying the importance of deepening students' conceptual understanding, especially in the context of intelligent tutoring systems (ITS). AW, designed to facilitate students' learning in an engaging manner, can help enhance students' grasp of underlying concepts. Although there seems to be an opportunity to orchestrate AW and AI such as large language models, little is known about how we can effectively design and develop an educational tool that infuses AW with such an emerging educational technology, especially for high school students with LD who struggle with writing proficiency (National Center for Education Statistics, 2012). To address this issue, this project proposes the combination and use of AI in the form of a conversational agent and FACT, a pedagogical strategy that utilizes AW, to provide an innovative way of helping high school students with LD learn math.

RELEVANT BACKGROUND/LITERATURE REVIEW

Prior work on AW demonstrates its effectiveness in helping students in disciplines such as science, social studies, and math. In science and social studies, AW is used to help improve students' reading comprehension, content knowledge, and writing structures (Hughes, 2020). In math, the effectiveness of AW can be pointed to research done by Powell et al. (2017), who analyzed 29 studies from grades 1-12 that include the use of math writing as an intervention or assessment. Of the 29 studies, three used AW as a

metric for assessment, and only one study by Cross (2009) focused on using AW as part of the instructional process. Cross (2009) explored the effects of using AW to enhance algebraic learning in 211 averageachieving 9th-grade students who were split into four groups: those who engaged in verbal arguments, written arguments, a combination of verbal and written arguments, and a control group receiving lectureoriented instruction. Groups that were assigned a treatment condition observed considerable improvements in algebraic learning compared to the control group, especially in the group that utilized both verbal and written arguments. Scholars agree that AW is an effective way to enhance math learning in a wide variety of grade levels (Powell et al., 2017; Graham et al., 2020). However, current methods for integrating and evaluating AW in classrooms struggle to meet the diverse needs of high school students with LD and rely solely on manual human assessment, which is not scalable (Powell et al., 2017). To address these challenges, we propose leveraging AI to automate and scale the AW assessment process, improving integration and scalability.

AI in the form of an ITS tends to be the most studied in AIEd due to its educational benefits, making it essential to explore within our project. In particular, the use of ITSs in education can provide personalized learning experiences, immediate student performance feedback, scalability for large-scale learning platforms, and enhanced accessibility (Chiu et al., 2023; Kulik, 2016; Xu et al., 2019). ITSs have also been shown to be effective tutors in subjects such as math (Arroyo et al., 2011). However, ITSs suffer from focusing on procedural practice while tutoring. A systematic review done by Mousavinasab et al. (2018) compiled a list of 53 studies that focused on ITSs in fields such as mathematics, computer science, and language to help students ranging from elementary to university education levels. A majority of the studies' effectiveness was based on the learner's performance, which is the learner's procedural knowledge in learning activities such as problem solving (Mousavinasab et al., 2018). Traditionally, ITSs focus on the procedural fluency of students rather than conceptual understanding when tutoring (Sedlmeier, 2001). This suggests that prior work done with ITSs often overlooks the importance of conceptual understanding, which is more beneficial to students for gaining content knowledge. To address this gap, we will implement the FACT pedagogical strategy within our conversational agent, which will act as an ITS.

To effectively incorporate AW with AI, this project proposes the use of FACT. FACT is a pedagogical strategy designed to actively engage students in problem-solving and AW through a systematic process comprising four stages: "F = Figure it out; A = Act on it; C = Compare your reasoning with a peer; T = Tie it up in an argument" (Kiuhara et. al, 2023, p. 7). FACT aids in clarifying students' conceptual misunderstandings in math through its four stages and integration of AW, which aligns with common core state standards for both mathematics and writing (Kiuhara et. al, 2023). By merging FACT with AI technology, we aim to create a conversational agent that assists high school students with LD in math learning. Because FACT is stagebased, AI can be seamlessly integrated to provide tailored support during each stage. In the "F" stage, AI can be used to extract complex math concepts from word problems and guide students in understanding them through external resources such as readings, definitions, and videos. During the "A" stage, AI can assist by providing hints and reflective feedback to help students in presenting supporting evidence and problem solving. In the "C" stage, AI will give an example solution to math problems and provide tools such as graphing that help the student compare and verify their solution against the chatbot. Finally, in the "T" stage, AI will restate the math problem and systematically review the FACT stages to help students construct and tie up their arguments. For a visual representation of how the conversational agent will implement the FACT strategy.

SPECIFIC ACTIVITIES AND TIMELINE

The project will implement the conversational agent through a Google Chrome extension directly integrated into Canvas. We will adopt a pre-post-test study design, and our goal is to invite 20 high school students with LD into a classroom setting to evaluate their math performance before and after the study. Quantitative (e.g. test results) and qualitative (e.g. chat logs) data will be collected, in which we will use mixed methods of statistical analysis to analyze students' artifacts. I have scheduled to meet with my advisor twice a week during the project timeline to track my progress and ensure that I am meeting project development deadlines, as well as to troubleshoot any issues that may arise.

May 13 - May 31: Front-end Development, Estimated Hours: 30
I will be working on the front-end development of the project to create a smooth and interactive user interface (UI) that adheres to common UI design principles (e.g. hierarchy and consistency). I will be using React as the UI framework and TypeScript for programming. We will also be using Git as a version control system and GitHub to reliably work on the development of the conversational agent.

June 3 - June 14: Back-end Development & Recruit Study Participants, Estimated Hours: 25

I will be working on the back-end development of the project which will involve a database and server. For the database, we will use a cloud-based NoSQL database service such as Firebase. TypeScript and Express.js will be utilized to develop the server, which will follow and adopt the REST paradigm. To deploy our server, we will be using a cheap AWS EC2 t2.micro instance. To incorporate AI into our project, we will be using proprietary AI services such as GPT-4, Gemini, or Claude. I will also be recruiting participants through crowdsourcing platforms such as Prolific for the usability study conducted after this phase.

June 17 - June 28: Usability Study & Project Refinement, Estimated Hours: 20

June 17 - June 21: After development, we will set up and conduct a usability study to evaluate the usability of our conversational agent. The study will take place in a virtual lab environment on a weekday, and we will ensure proper set-up and accommodations. We will have recruited 4 high school students with LD to take part in this study from crowdsourcing (Prolific). Usability data will be collected via the System Usability Scale, and we will also be conducting a cognitive walkthrough.

June 24 - June 28: Using the data collected from the usability study, we will refine our conversational agent to prepare it for a more complete classroom study in July. This may include changing the UI, removing/adding features, and fixing application errors.

July 1 - August 9: Classroom Study, Data Collection, & Data Analysis

July 1 - July 12: In this phase, we will be setting up a more rigorous and complete study to evaluate. We will recruit 20 participants, specifically high school students with LD, and conduct the study within a classroom setting. Laptops used for the study will be set up with all the necessary software and accommodations needed. Our study will adopt a pre-post-test design, and the tests will be conducted by randomly selecting 10 math questions from a pool of 40 math questions that all address common core state standards in linear equations. Because we will be evaluating the success of our conversational agent through a pre-posttest design, our dependent variable for this study will be the participants' math test scores. Our independent variables include the length of each session and initial laptop setup. This classroom study will take place in the first two weeks of July, and we will conduct at least three one-hour sessions throughout Monday, Wednesday, and Thursday during the two weeks. Estimated Hours: 20

July 15 - August 9: After the collection of quantitative (e.g. usage metrics, test results, correct/incorrect answers) and qualitative (e.g. user chat logs with the AI) data, the remaining four weeks will be spent on data analysis to analyze the students' artifacts and determine whether or not the project was successful in improving math performance for high school students with LD. For statistical analysis of the data, we will be using a paired t-test. We will also conduct an epistemic network analysis to gather some learning analytics. Estimated Hours: 25

Post-funding Period:
After the project, I plan to write up a report summarizing my findings and disseminate my work during the Fall 2024 Undergraduate Research Symposium. I will also be making the development work open-source through the project's GitHub repository. Additionally, I plan to seek other external conference opportunities to submit my work.

RELATIONSHIP OF WORK TO THE EXPERTISE OF THE MENTOR

Dr. XXX is an assistant professor of Instructional Design & Educational Technology (IDET) at the University of Utah College of Education. His research focuses on the design and development of nextgeneration educational tools using emerging learning technologies with fair, accountable, and transparent (FAccT) AI in education. Dr. XXX serves as key personnel in multiple multi-million-dollar federal and foundational grants on developing innovative and FAccT generative AI models in education for mathematics learning, and he currently serves as a Co-PI of the Learning Engineering Virtual Institute ($10M, G-23-2137070). He has four years of teaching and mentoring experience with undergraduate and graduate students, where he believes students learn more effectively when they have more autonomy and ownership over their learning. Since October 2023, Dr. XXX has been mentoring me through our weekly meetings on this project. Dr. XXX and I share similar research interests, and Dr. XXX has already made a significant impact on me by teaching me about AI and software development. Dr. XXX has also been kind, patient, communicative, and supportive to me, and because of our well-established mentoring relationship, I am confident in his ability to be an effective mentor for this project. Dr. XXX will dedicate 8% of the 12-month calendar year to support and guide me throughout this project.

RELATIONSHIP OF THE WORK TO YOUR FUTURE GOALS

I am currently in my freshman year of college studying computer science. As a first-generation student, my long-term goal is to attend graduate school and pursue a fulfilling, high-impact career in educational technology. In this project, I have the unique opportunity to combine my passion for education and computer science to directly investigate the impacts of AIEd in an interdisciplinary, real-world setting. I hope to further advance the promising field of AIEd in the future by continuing with my studies and research pursuits. This project will serve as the basis for my research and professional career, and by the end of the project, I will have gained valuable experience in full-stack software development, AI technologies, research design, data collection and analysis, and practice in research presentation.

REFERENCES (Works Cited)

Arroyo, I., Royer, M. J., & Woolf, B. P. (2011). Using an Intelligent Tutor and Math Fluency Training to Improve Math Performance. International Journal of Artificial Intelligence in Education, 21(1-2), 135–152.
doi:10.3233/JA
I- 2011-020

Casado-Ledesma, L., Cuevas, I., & Martín, E. (2021). Learning science through argumentative synthesis writing and deliberative dialogues: A comprehensive and effective methodology in secondary education. Reading and Writing, 36(4), 965–996. doi:10.1007/s11145-021-10191-0

Chiu, T. K., Xia, Q., Zhou, X., Chai, C. S., & Cheng, M. (2023). Systematic literature review on opportunities, challenges, and future research recommendations of Artificial Intelligence in Education. Computers and Education: Artificial Intelligence, 4, 100118. doi:10.1016/j.caeai.2022.100118

Cross, D. I. (2009). Creating optimal mathematics learning environments: Combining argumentation and writing to enhance achievement. International Journal of Science and Mathematics Education, 7, 905‒930. doi:10.1007/s10763-008-9144-9

Ferretti, R. P., Lewis, W. E., & Andrews-Weckerly, S. (2009). Do Goals Affect the Structure of Students' Argumentative Writing Strategies? Journal of Educational Psychology, 101(3), 577–589. doi:10.1037/a0014702

Graham, S., Kiuhara, S. A., & MacKay, M. (2020). The Effects of Writing on Learning in Science, Social Studies, and
Mathematics: A Meta-Analysis. Review of Educational Research, 90(2), 179-226. doi:10.3102/0034654320914744

Hughes, E. M., & Lee, J. (2020). Effects of a Mathematical Writing Intervention on Middle School Students'
Performance. Reading & Writing Quarterly, 36(2), 176–192. doi: 10.1080/10573569.2019.1677537

Jitendra, A. K., Lein, A. E., Im, S., Alghamdi, A. A., Hefte, S. B., & Mouanoutoua, J. (2018). Mathematical Interventions for Secondary Students With Learning Disabilities and Mathematics Difficulties: A Meta-
Analysis. Exceptional Children, 84(2), 177-196. doi:10.1177/0014402917737467

Kiru, E. W., Doabler, C. T., Sorrells, A. M., & Cooc, N. A. (2018). A Synthesis of Technology-Mediated Mathematics
Interventions for Students With or at Risk for Mathematics Learning Disabilities. Journal of Special Education Technology, 33(2), 111-123. doi:10.1177/0162643417745835

Kiuhara, S. A., Levin, J. R., Tolbert, M., O'Keeffe, B. V., O'Neill, R. E., & Jameson, M. J. (2023). Teaching argument writing in math class: challenges and solutions to improve the performance of 4th and 5th graders with disabilities. Reading and Writing. 1-30, doi:10.1007/s11145-023-10459-7

Kulik, J. A., & Fletcher, J. D. (2016). Effectiveness of Intelligent Tutoring Systems: A Meta-Analytic Review.
Review of Educational Research, 86(1), 42-78. doi:10.3102/0034654315581420

Mousavinasab, E., Zarifsanaiey, N., Kalhori, S. R. N., Rakhshan, M., Keikha, L., & Saeedi, M. G. (2018).
Intelligent tutoring systems: A systematic review of characteristics, applications, and evaluation methods. Interactive Learning Environments, 29(1), 142–163. doi:10.1080/10494820.2018.1558257

National Center for Education Statistics (2012). The Nation's Report Card: Writing 2011 (NCES 2012–470). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

National Center for Education Statistics (2019). 2019 NAEP Mathematics Report Card (NCES 2020-090). Institute of Education Sciences, U.S. Department of Education, Washington, D.C.

Newell, G. E., Beach, R., Smith, J., & VanDerHeide, J. (2011). Teaching and Learning Argumentative Reading and Writing: A Review of Research. Reading Research Quarterly, 46(3), 273-304. doi:10.1598/RRQ.46.3.4

Powell, S. R., Hebert, M. A., Cohen, J. A., Casa, T. M., & Firmender, J. M. (2017). A synthesis of mathematics writing: Assessments, interventions, and surveys. Journal of Writing Research, 8(3), 493–530.
doi:10.17239/jowr-2017.08.03.04

Sedlmeier, P. (2001). Intelligent Tutoring Systems. International Encyclopedia of the Social & Behavioral Sciences, 7674–7678. doi:10.1016/b0-08-043076-7/01618-1

Xu, Z., Wijekumar, K., Ramirez, G., Hu, X., & Irey, R. (2019). The effectiveness of intelligent tutoring systems on K 12 students' reading comprehension: A metaanalysis. British Journal of Educational Technology, 50(6), 3119–3137. doi:10.1111/bjet.12758