A.I. Math Personalization Tool (AMPT): Empowering Students through Peer-Authored Math Content

Kole Norberg, Ethan Shafran Moltz, Husni Almoubayyed, Logan De Ley, Josh Fisher, April Murphy, Stephen E. Fancsali, Steve Ritter

Carnegie Learning

knorberg1@carnegielearning.com

Abstract

The A.I. Math Personalization Tool (AMPT) uses generative AI to directly engage students as co-authors of math problems with the goal of increasing student feelings of agency and interest in mathematics.

Introduction

Students benefit from content that makes sense to them, that reflects their interests and represents their lived experiences (Ku and Sullivan 2000; Ku et al. 2007), but educational content often includes an implicit bias that favors the interests of majority groups (Boutte et al. 2010; Davis and Martin 2008). Such content has been shown to both decrease students' sense of belonging (SOB; e.g., Cleary 2008; Guthrie et al. 2004) and increase reading difficulty (Smith et al. 2021), making it more difficult for learners unfamiliar with a topic to focus on the mathematical learning objectives (e.g., Koedinger and Nathan 2004).

The A.I. Math Personalization Tool (AMPT) provides a collaborative approach to content generation which treats students as experts on what they consider engaging and relevant. AMPT uses a sequence of prompts to OpenAI's GPT-4 Turbo (OpenAI 2023) to chat with a student about a topic of interest, construct a math word problem combining that interest with a target math concept, and then work with the student to revise the problem as needed. See Figure 1 for a full depiction. Example problems can be found at https://osf.io/pa824/.

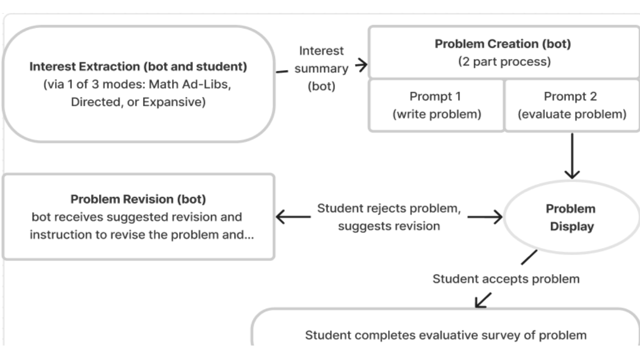

Figure 2 Prompt flow for creating a word problem using AMPT

Interest Extraction

AMPT includes a chat-based interface which is designed to prompt the student to provide details, including names, that will later be included in the math problem. Prior approaches to math personalization matched students to problems using broad topics of interest (e.g., Fancsali and Ritter 2014; Walkington and Sherman). Generative AI provides an opportunity to create hyper-personalized problems which reflect more granular interests (e.g., Fancsali et al., 2013). However, it is not clear whether hyper-personalization will yield improvements in math or SOB outcomes over those found using coarse-grained topics of interest. Further, hyper-personalization requires an open-ended conversation between the large language model (LLM) and the student. Although OpenAI provides guardrails to prevent undesired conversations, LLMs still may hallucinate, introduce bias, or deviate from their roles, especially when there are no constraints on the student's responses. Given the greater costs (i.e., tokens counts) and risks of allowing students to engage in extended conversations with Gen-AI based tools, we sought to test the relative efficacy of different types of chat-based interfaces for generating age and math level appropriate problems that students found interesting. One of the interfaces permitted open-ended conversation (Expansive) and two of them restricted it (Math Ad-Libs & Directed).

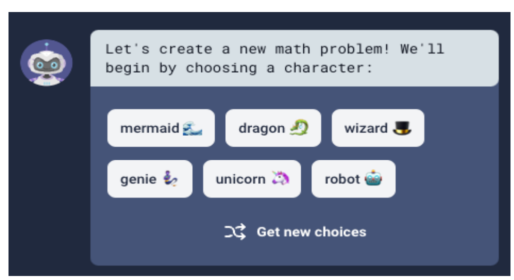

Figure 1 Math Ad-Libs and Directed Modes prompt students to pick a character, place, and activity

Math Ad-Libs and Directed Interfaces

The Math Ad-Libs and Directed interfaces are the least personalized, but still allow students to control what is included in the problem. Students are asked to select two characters, a place, and an activity. For each of these, an initial set of options is suggested, and students can request new sets up to three times. The critical difference between the Math Ad-Libs and Directed interfaces is whether GPT-4 is passed information about student's prior answers. In Math Ad-Libs, the student's response is not included in the call for GPT-4 to produce a new set of options. The resulting problem may make less narrative sense but may also be more amusing. The Directed interface includes the student's response in the call for subsequent options with the requirement that new choices should relate to prior ones. This allows the student to construct their own narrative, but through preset options.

Expansive Interface

The Expansive interface begins by asking students broad questions about their interests. Across six exchanges, AMPT first expresses understanding of what the student said and then asks students increasingly detailed questions.

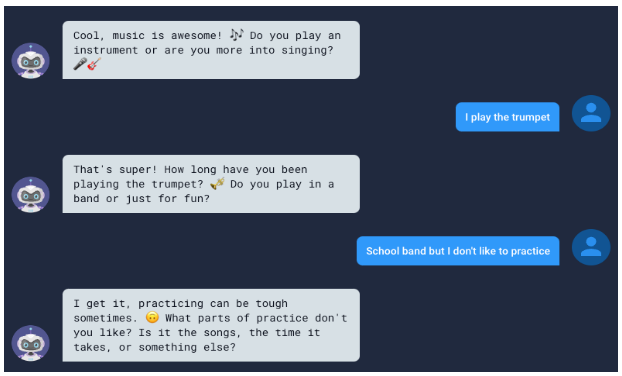

Figure 3 Expansive Mode engages the student in a conversation about their interest

There is risk with such an open exchange that the student may introduced inappropriate topics. We mitigate this risk by emphasizing the age of the conversation partner in the prompt. In testing, the interest extractor has never included age-inappropriate content in its responses and has successfully diverted student attempts to introduce sensitive topics.

Problem Creation

AMPT currently supports problem creation in the domains of algebra, probability, and ratios. Problem creation is a two-step process. In the first step, a GPT-4 generated summary of the student's interests is fed into a prompt with instructions for the AI to draft the problem. A second prompt then asks GPT-4 to evaluate the problem, ensuring the problem is age-appropriate, respectful, and includes necessary math. In a recent pilot study with 9 middle school students, all 29 student-generated questions had the appropriate mathematical attributes for the targeted problem type.

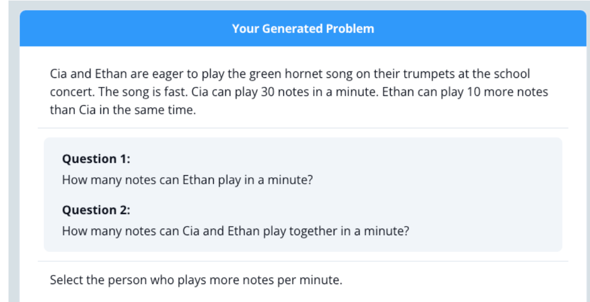

Figure 4 AMPT generated problem related to evaluating ratios

Problem Revision

After a problem is created, students are given the option to keep the problem or revise it by providing free-form directions for how the problem should change. GPT-4 receives the students' instructions along with additional instructions to leave mathematical elements of the problem intact.

Student Evaluation

Students evaluated the final math problem and interfaces on a 5-point Likert scale. During pilot testing, problems created using the Expansive interface were marginally more interesting than the other modes, $t(51) = 1.79$, $p = .08$. Further, on an end-of-study task, students showed a clear preference for typing open-ended responses versus clicking on choices from a list, $t(16) = 3.39$, $p = .004$.

Finally, in a pilot test, we evaluated sense of belonging (SOB) in mathematics (Rattan et al., 2012) before and after students interacted with AMPT. SOB showed proportional improvement of 5.64% from pre to post. These early results suggest a potential authoring effect whereby allowing students to author problems improves attitudes towards math. Results reported here are preliminary and ongoing work is in process to verify with larger student populations.

Future Directions

Encoding math problems to use in an intelligent tutoring system (ITS) has been a long-standing hurdle in allowing for user authoring of math problems (e.g., Ritter et al., 1998). GPT-4 auto-encodes problems from AMPT for use in MATHia (Carnegie Learning's ITS). This allows us to further evaluate the effect of personalization on student learning outcomes among MATHia's user base of 600k+ learners each year. Testing is underway to ensure outcomes on student generated problems match or improve on those of pre-existing problems. Further testing will implement controls to determine the role of hyper-personalization. Finally, early findings suggest using AMPT improves students' SOB. We will also test how problem authoring specifically, as compared to receiving personalized content, affects feelings of inclusion.

References

Boutte, G., Kelly-Jackson, C., & Johnson, G. L. 2010. Culturally relevant teaching in science classrooms: Addressing academic achievement, cultural competence, and critical consciousness. *International Journal of Multicultural Education*, 12(2). https://doi.org/10.18251/ijme.v12i2.343

Cleary, Sandra. 2008. Communication: A hands-on approach. Landsdowne: Juta and Company Ltd.

Davis, J., & Martin, D. B. 2008. Racism, assessment, and instructional practices: Implications for mathematics teachers of African American students. *Journal of Urban Mathematics Education* 1(1): 10-34. https://doi.org/10.21423/jume-v1i1a14

Fancsali, S. E., & Ritter, S. 2014. Context personalization, preferences, and performance in an intelligent tutoring system for middle school mathematics. In Proceedings of the fourth international conference on learning analytics and knowledge. https://doi.org/10.1145/2567574.2567615

Fancsali, S. E., Ritter, S., Stamper, J., and Nixon, T. 2013. Toward "hyper-personalized" Cognitive Tutors. In AIED 2013 Workshops Proceedings 7: 71-79.

Guthrie, J. T., Wigfield, A., & Perencevich, K. C. 2004. Scaffolding for motivation and engagement in reading. In *Motivating Reading Comprehension: Concept-Oriented Reading Instruction*: edited by J. T. Guthrie, A. Wigfield, & K. C. Perencevich, 55-86. New York: Routledge.

Koedinger, K. R., & Nathan, M. J. 2004. The real story behind story problems: Effects of representations on quantitative reasoning. *Journal of the Learning Sciences,* 13(2): 129-164. https://doi.org/10.1207/s15327809jls1302_1

Ku, H.-Y., Harter, C. A., Liu, P.-L., Thompson, L., & Cheng, Y.-C. 2007. The effects of individually personalized computer-based instructional program on solving mathematics problems. *Computers in Human Behavior*, 23(3): 1195-1210. https://doi.org/10.1016/j.chb.2004.11.017

Ku, H.-Y., & Sullivan, H. J. (2000). Personalization of mathematics word problems in Taiwan. *Educational Technology Research and Development*, 48(3): 49-60. https://doi.org/10.1007/BF02319857

OpenAI. (2023). ChatGPT. Nov 7 version. [Large Language Model].

Rattan, A., Good, C., & Dweck, C. S. 2012. "It's ok—Not everyone can be good at math": Instructors with an entity theory comfort (and demotivate) students. *Journal of experimental social psychology*, *48*(3): 731-737. https://doi.org/10.1016/j.jesp.2011.12.012

Ritter, S., Anderson, J. R., Cytrynowicz, M., & Medvedeva, O. 1998. Authoring content in the PAT algebra tutor. *Journal of Interactive Media in Education* (2).

Smith, R., Snow, P., Serry, T., & Hammond, L. 2021. The role of background knowledge in reading comprehension: A critical review. *Reading Psychology*, 42(3): 214-240. https://doi.org/10.1080/02702711.2021.1888348

Walkington, C. & Sherman, M. (2012). Using Adaptive Learning Technologies to Personalize Instruction: The Impact of Interest-Based Scenarios on Performance in Algebra. In van Aalst, J., Thompson, K., Jacobson, M. J., & Reimann, P. (Eds.), The Future of Learning: Proceedings of the 10th International Conference of the Learning Sciences (ICLS 2012) – Volume 1, Full Papers (pp. 80-87). Sydney, NSW, AUSTRALIA: International Society of the Learning Sciences. https://repository.isls.org//handle/1/2246