Summary of "Towards the Future of AI-augmented Human Tutoring in Math Learning"

Danielle R. Thomas[1], Vincent Aleven[1], Richard Baraniuk[2], Emma Brunskill[3], Scott Crossley[4], Dora Demszky[3], Stephen Fancsali[5], Shivang Gupta[1], Steve Ritter[5], Simon Woodhead[5], Wanli Xing[7] and Kenneth Koedinger[1]

[1]*Carnegie Mellon University, Pittsburgh, PA 15213, USA*
[2]*Rice University, Houston, TX 77005, USA*
[3]*Stanford University, Stanford, CA 94305, USA*
[4]*Vanderbilt University, Nashville, TN 37203, USA*
[5]*Carnegie Learning, Pittsburgh, PA 15219, USA*
[6]*Eedi, London, England, UK*
[7]*University of Florida, Gainesville, FL 32611, USA*

Abstract

We summarize the proceedings of a full-day, hybrid workshop at the *International Conference of Artificial Intelligence in Education* hosted in Tokyo, Japan on July 3, 2023. The workshop, "Towards the Future of AI-augmented Human Tutoring in Math Learning," focuses on the use of artificial intelligence (AI)-assisted human tutoring in math learning. This workshop emphasizes attention to equity and improving access to high-quality learning opportunities among historically marginalized students, with a focus on obstacles to scaling. Among the six accepted papers and moderated panel discussion, we highlight the following key findings: 1) a greater general focus on identifying or diagnosing student's needs and less so on the interventions or remedies that might follow, 2) large language models are the focal point among the vast exploration of applications occuring, and 3) human mentoring remains a strong, irreplaceable influence. Challenges and takeaways from this workshop sparked interest among the AIED community in the development of human-AI hybrid tutoring systems.

Keywords

Tutoring, Personalized learning, AI-assisted tutoring

1. Introduction & Theme

The primary challenge to improving middle school math achievement is providing all students equitable access to the existing high-quality learning opportunities that we know to be effective. Students from economically disadvantaged and historically underserved backgrounds can learn just as well as their peers when given the same opportunities, but they are more likely to experience learning gaps due to a lack of access to these learning opportunities [1]. High-dosage human tutoring can produce dramatic learning gains, particularly if tutors are well-trained in providing students social-motivational support [2]. However, low-income students lack access

AIED 2023 Workshop: Towards the Future of AI-Augmented Human Tutoring in Math Learning, July 07, 2023, Tokyo, Japan
shivangg@andrew.cmu.edu (S. Gupta); koedinger@cmu.edu (K. Koedinger)

to well-trained tutors, evidenced by the 16 million low-income children on the waitlist for high-quality afterschool programs [3]. In addition, the estimated costs of $2500+ per student for individualized tutoring prohibits student access [4]. Human tutoring alone cannot meet present students' need. Sustainable and scalable tutoring infrastructures are possible through the combined synergy of artificial intelligence (AI)-assisted and human technologies that can be achieved through novel and well-engineered AI-supported tutoring models.

AI-assisted tutoring shows promise and can potentially double learning outcomes [5]. but analytics show that many students, especially from low-income backgrounds, are not getting sufficient learning opportunities. Student inaccessibility can be attributed to a variety of factors, including: not having sufficient access to the medium of using AI, such as digital devices and internet; issues facing inclusion with inadequate support of diverse student needs, such as English language learners and students with disabilities; and a lack of understanding of AI capabilities and limitations [6]. The challenges facing math learning related to access, equity, fairness, and inclusion have fostered collaborative and focused efforts on AI assisted human-technology ecosystems that increase learning opportunities for all students.

There is a concerted effort within the AIED community to increase learning opportunities among economically disadvantaged and historically underrepresented students. The COVID-19 pandemic had a severe impact on education globally. The U.S. has lost nearly twenty years of math progress among middle school students [7], with racial and economic learning gaps preventing millions of students from realizing their potential. By leveraging the power of AI, the AIED community is working to provide equitable learning opportunities and helping bridge the persistent opportunity gap in action. This workshop aimed to facilitate discussion and engagement among the AIED community regarding AI-assisted individualized learning tools to improve middle school teaching and tutoring. In particular, the workshop hosted updates on progress, findings, and challenges to AI-supported personalized instruction. We invited empirical and theoretical papers aligned with the theme particularly (but not exclusively) within the following areas of research and application:

- **AI-assisted and Human Tutoring Systems:** Insight into better understanding and supporting human, AI-assisted, and interactive learning technologies related to individualized instruction.
- **Delivery and Scale:** Efficacy of different human tutoring delivery systems (e.g., video, audio, chat) and the corresponding needed differentiated support; Different models for scaling including peer tutoring, computer tutoring, etc.
- **Training Development:** Tutor and teacher training development that recognizes diverse experiences and backgrounds, in relation to AI-assisted tutoring support structures.
- **Equity and Inclusion:** Issues facing equity and inclusion, with focus on intelligent techniques to support students from under resourced communities.
- **Ethics:** Privacy and transparency of intelligent techniques, such as using federated machine learning and explainable AI to examine data ownership and human-AI collaboration; Transferability and fairness of predictive models across educational contexts.
- **Evaluation:** Program evaluation, such as applications using large-language models or dataset development for reinforcement learning of models; Methods of measuring student growth, with possible insights into dosage; Evidence of learning outcomes.

- **Key Challenges:** Barriers, considerations, and challenges to providing human and AI-based tutoring and individualized instruction at scale.
- **Interoperability:** How do AI and human tutoring systems interact with existing technological and social systems?

The *Introduction & Theme* section (above) is described in the original call for papers. This full-day, hybrid workshop consisted of the following activities: 1) presentations of accepted papers with Q&A, 2) small-group discussions on the conference themes, 3) reports of small-group discussions, 4) a moderated panel with audience participation focused on next steps, and 5) a closing summary and discussion. The call for papers explains in greater detail the relevance, theme, workshop format, target audience, and participation details and can be found in the *International Conference of Artificial Intelligence in Education* (AIED) 2023 proceedings (volume 2): https://doi.org/10.1007/978-3-031-36336-8_3, and the workshop website: https://sites.google.com/andrew.cmu.edu/aied2023workshop/home

2. Proceedings Summary

The organizing committee received seven papers, with each submitted paper being reviewed by at least two committee members. Review of papers followed a single-blind review process, with reviewers anonymous and authors unknown. Reviewers were required to make a recommendation of either acceptance or rejection of the paper and explain their reasoning behind their decision. They assessed papers based on three criteria, using a scoring system of -1, 0, or 1; alignment with the workshop's theme, level of interest to AIED, and overall quality. Following this process, six papers were accepted into the workshop proceedings. A short summary and high-level contribution is described below, along with alignment to the theme indicated in brackets:

Orchestrating Classrooms and Tutoring with Carnegie Learning's MATHia and LiveLab
Stephen Fancasli, Michael Sandbothe, Steve Ritter
[Orchestration, Evaluation, AI-assisted & Human Tutoring Systems]
The authors describe on-going research and a "road map" for learning analytics research on detector models and software feature development to orchestrate human tutoring. The ability to provide data-driven guidance from AI-driven adaptive learning software, such as Carnegie Learning's MATHia and LiveLab, can support classroom math instructors and tutors to achieve greater efficiency and lower costs, particularly at scale.

Using Large Language Models to Provide Explanatory Feedback to Human Tutors
Jionghao Lin, Danielle R. Thomas, Feifei Han, Shivang Gupta, Wei Tan, Ngoc Dang Nguyen, Kenneth R. Koedinger
[Large Language Models (Evaluation), Training Development]
The authors describe two methods of providing real-time feedback to tutors engaging in an online lesson on how to give praise. This work-in-progress demonstrates considerable accuracy in binary classification for corrective feedback of effective and ineffective praise and showcases an enhanced approach of providing explanatory feedback using large language model-facilitated

name entity recognition. The latter of which may be able to provide tutors feedback, not only while engaging in lessons, but can potentially suggest real-time tutor moves.

Face Readers: The Frontier of Computer Vision and Math Learning
Beverly Woolf, Margrit Betke, Hao Yu, Sarah Adel Bargal, Ivon Arroyo, John Magee, Danielle Allession, William Rebelsky
[Delivery & Scale, AI-assisted & Human Tutoring Systems]
This work highlights the use of student facial expression to determine student's individual needs and provide insight to educators on delivering immediate feedback. Using the discussed Face Readers technology, the authors describe three phases of development: 1) collecting datasets and identifying salient labels of facial features; 2) building and training deep learning models; and 3) predicting student problem-solving outcomes. The author's explain how facial recognition technology to support educators in determining and responding to student's individual needs is the next frontier of AI-assisted human tutoring.

Comparative Analysis of GPT-4 and Human Graders in Evaluating Human Tutors Giving Praise to Students
Dollaya Hirunyasiri, Danielle R. Thomas, Jionghao Lin, Kenneth R. Koedinger, Vincent Aleven
[Large Language Models (Evaluation), Tutor Training Development]
This preliminary work showcases the potential of large language models to provide constructive feedback to tutors in practical settings. Using 30 synthetic tutor-student dialogues, the authors apply zero-shot and few-shot learning approaches to prompt GPT-4 to identify key components of praise from tutors to students. GPT-4 performs well in recognizing specific and immediate praise and underperforms in identifying sincerity. The authors express much more investigation is needed on enhancing prompt engineering, and evaluating their method using real-life tutoring dialogues.

Does ChatGPT Comprehend the Place Value in Numbers When Solving Math Problems
Jisu An, Junseok Lee, Gahgene Gweon
[Large Language Models (Evaluation), AI-assisted & Human Tutoring Systems]
In this work, authors investigate the ability of chain-of-thought and program-of-thought GPT-based models to determine if textual or numerical expressions can yield better performance in solving math word problems. The authors conclude that the concept of place value is not adequately integrated when numbers are represented as tokens using the specified GPT model and state research on training models to "understand" the concept of place value is an area of future research.

ITS Unplugged: Leapfrogging the Digital Divide for Teaching Numeracy Skills in Underserved Populations
Thomaz Edson Veloso da Silva, Geiser Chalco Challco, Luiz Rodrigues, Fabiana Maris Versuti, Rodolfo sena da Penha, Lívia Silvia Oliveira, Guilherme Corredato Guerino, Luis Felipe Cavalcanti de Amorim, Marcelo Luiz Monteiro Marinho, Valmir Macario, Diego Dermeval, Ig Ibert Bittencourt, Seiji Isotani
[AI-assisting & Human Tutoring Systems, Equity & Inclusion]

This paper introduces a pioneering pedagogical workflow that integrates an intelligent tutoring system (ITS) into school curriculum lesson plans. The ITS "unplugged" model teaches numeracy using computer vision and natural language processing techniques, without the use of internet connectivity, to capture and analyze student responses via photographs. The "unplugged" ITS model enables educators to make informed decisions based on student performance, while eliminating the need for internet connectivity—a resource not available to many students.

3. Panel Discussion Summary

A moderated panel discussion consisted of three, in-person panelists. Andrew Lan is an assistant professor in the College of Information and Computer Sciences at the University of Massachusetts Amherst. Andrew focuses on the development of human-in-the-loop machine learning methods to enable scalable, effective, and safe personalized learning. Jionghao Lin is a postdoctoral research fellow within the Human-Computer Interaction Institute at Carnegie Mellon University. Jionghao's research interests include, learning analytics, data mining, and explainable AI, particularly in relation to feedback delivery. Mutlu Cukurova is a professor of Learning and Artificial Intelligence at University College London. Mutlu's research interests include: human-AI collaboration in teaching and learning contexts; computational and statistical models of collaboration and regulation of learning behaviors; and socio-scientific and psychological challenges in the adoption of AI and analytics in education. The moderator, Danielle R. Thomas, is a systems scientist from Carnegie Mellon University focusing on the practical intersection of AI-assisted human tutoring, learning engineering, and pragmatic decision-making in real-life tutoring environments.

4. Key Findings & Take Aways

High-level findings, key issues, and commonalities from accepted papers and the panel discussion include:

There was greater general focus on identifying or diagnosing student's needs and less so on the interventions or remedies that might follow. The majority of accepted papers highlight advancing technologies related to the accuracy and efficiency of identifying student content-level (i.e., math) struggle or disengagement in learning. For example, Fancsali et al. discuss using detector models and software features to identify student cognitive and noncognitive struggles, allowing for immediate intervention by the instructor or tutor. Woolf et al. harness facial recognition technology to diagnose student's affective states for purposes of educators quickly remediating student's struggles. Lastly, da Silva et al. leverage computer vision and natural language processing to capture and analyze student's numeracy struggles via photographs—and the list goes on. However, although detecting student lack of engagement through learning analytics and detection software was a hot topic among contributors, little to no research and development work was being investigated among accepted papers on "how to" motivate students. Advancing technology to quickly identify a student lacking motivation or engagement i.e., student gazing away from the screen (Woolf et al.), increasing idle time (Fancasli et al.), student scribbling nonsensically on a numeracy problem (da Silva et al.) was of considerable interest among accepted papers, with the emphasis on “how to accurately identify and respond” and less on “how to effectively remedy or motivate.” It is important to mention that AI-in-the-loop human tutoring can only be effective in increasing learning if students are engaging with it.

How can we increase student motivation to engage with these systems we are creating? The same theme, focusing on technologies to “diagnose” or “detect” students in need of support, resonated within the panel discussion with members mentioning the importance of motivating students several times. The question was posed, “Even if we are able to achieve high accuracy in detecting student lack of motivation and disengagement, then what?” There was general agreement among panelists that future work is needed on what interventions can or should be pursued based on improved diagnosis. Nevertheless, the factors surrounding the “secret sauce” to striking math interest in students, may be found within the subtle and profound importance of human relationships.

Large language models are center stage; Human mentoring remains a strong influence. Three of out six for the accepted papers focused primarily on the use of large language models: 1) for providing real-time explanatory feedback to human tutors within online lessons (Lin et al.); 2) to accurately identify criteria of effective praise among human tutors responses to students (Hirunyasiri et al.); and 3) to assess the ability of GPT-based models to comprehend place value in solving math problems (An et al.). Similarly, the majority of the panel discussion revolved around the usage of large language models for practical application, such as providing tutors real-time feedback on their tutoring and providing hints to students working through math problems. However, more abstract, thought-provoking questions were posed, such as: “Will AI, leveraging the use of large language models, eventually take over the role of a human tutor?” and “What will AI-in-the-loop tutoring look like 5 or 10 years from today?” Among both of these questions, although difficult to predict by any expert researcher, panelists were in collective agreement that the role of a human tutor, as a mentor, as a guide, or even as an academic confidante, is not going anywhere anytime soon.

5. Acknowledgments

We would like to thank the organizing committee that helped prepare for the workshop and reviewed the workshop proceedings:

Danielle R. Thomas, Ed.D., Carnegie Mellon University, drthomas@cmu.edu
Danielle is a systems scientist and faculty member at Carnegie Mellon University and research lead on the PLUS (Personalized Learning Squared) tutoring project. She is a former middle school math teacher and school administrator, founding several mentoring programs supporting young women and youth in STEM. Danielle leverages her past experiences to advance research and development of tutor training and the creation of AI-assisted tutor feedback. She has first-authored over a dozen peer-reviewed papers since 2021, focusing on AI-assisted human tutoring, learning engineering, and equity, particularly in math and STEM education.

Vincent Aleven, Ph.D., Carnegie Mellon University, aleven@cs.cmu.edu
Vincent is a Professor of Human-Computer Interaction at Carnegie Mellon University, with 30 years of experience in research of AI-based learning. His lab created Mathtutor, an AI-

based tutoring software for middle school math and the tools for AI-based software, CTAT and Tutorshop. Vincent has written over 250 publications, with he and his team winning 11 best paper awards at international conferences and has acted as PI or co-PI on 20 major research grants. Currently, Vincent is co-editor-in-chief of the *International Journal of Artificial Intelligence in Education* (IJAIED).

Richard Baraniuk, Ph.D., OpenStax, Rice University, richb@rice.edu
Richard is the C. Sidney Burrus Professor of Electrical and Computer Engineering at Rice University and the Founding Director of OpenStax. He is a Member of the National Academy of Engineering and American Academy of Arts and Sciences and a Fellow of the National Academy of Inventors, American Association for the Advancement of Science, and IEEE. For his work in open education, he has received the C. Holmes MacDonald National Outstanding Teaching Award, the Tech Museum of Innovation Laureate Award, the Internet Pioneer Award from the Berkman Center for Internet and Society at Harvard Law School, and many other prestigious awards.

Emma Brunskill, Ph.D., Stanford University, ebrun@cs.stanford.edu
Emma is an Associate Professor in the Computer Science Department at Stanford University where she aims to create AI systems that learn from a few samples to robustly make good decisions. Her work is inspired by the positive impact AI may have in education and healthcare, with interests in large language models to advance AI-assisted human tutoring. Emma is part of the Stanford AI Lab, the Stanford Statistical ML group, and AI Safety @Stanford. She has received an NSF CAREER award, Office of Naval Research Young Investigator Award, and many other awards. Emma and her lab have received multiple best paper nominations for their AI and machine learning work.

Scott Crossley, Ph.D., Vanderbilt University, scott.crossley@vanderbilt.edu
Scott is a Professor of Special Education at Vanderbilt University. His primary research focus is on natural language processing and the application of computational tools and machine learning algorithms in language learning, writing, and text comprehensibility. His main interest area is the development and use of natural language processing tools in assessing writing quality and text difficulty. He is also interested in the development of second language learner lexicons and the potential to examine lexical growth and lexical proficiency using computational algorithms.

Dora Demszky, Ph.D., Stanford University, ddemszky@stanford.edu
Dora is an Assistant Professor in Education Data Science at Stanford University. Her research focuses on measuring equity, representation, and student-centeredness in educational texts, with the goal of providing insights to educators to improve instruction. She develops measures based on natural language processing that work well for high-dimensional, unstructured data, and she applies these measures to provide feedback to educators. Dr. Demszky has received her PhD in Linguistics at Stanford.

Stephen Fancsali, Ph.D., Carnegie Learning, sfancsali@carnegielearning.com
Stephen is Director of Advanced Analytics at Carnegie Learning. With over a decade of

experience in educational data science, he specializes in statistical and causal modeling of data produced by learners as they interact with AI-driven instructional software. He works on innovative learning analytics and models of student learning underlying MATHia, LiveLab, MATHstream, and other products. Stephen has published in the *Journal of Learning Analytics* and many conference proceedings. He received a Ph.D. in Logic, Computation, and Methodology from Carnegie Mellon University.

Shivang Gupta, Carnegie Mellon University, shivangg@andrew.cmu.edu
Shiv is the Head of Product at PLUS - Personalized Learning Squared at Carnegie Mellon University. A graduate of the Masters in Educational Technology and Applied Learning Science (METALS) program at CMU, Shiv was the lead curriculum developer at First Code Academy in Hong Kong and previously worked on corporate training in the metaverse.

Steve Ritter, Ph.D., Carnegie Learning, sritter@carnegielearning.com
Steve Ritter is Founder and Chief Scientist at Carnegie Learning. Dr. Ritter earned a doctorate in cognitive psychology at Carnegie Mellon University. He was instrumental in the development of the Cognitive Tutors for math, which led to Carnegie Learning, where it forms the basis of the MATHia intelligent tutoring system. Dr. Ritter is the author of many papers on the design and evaluation of adaptive instructional systems and is recognized as an expert in the field. Dr. Ritter leads a research team devoted to using learning engineering to improve the efficacy of the company's products.

Simon Woodhead, Ph.D., Eedi, simon.woodhead@eedi.co.uk
Simon is a co-founder of Eedi and also host of the Data Science in Education meetup. He coordinates Eedi's machine learning research, which has been conducted in collaboration with Microsoft Research, and turns this into new product features. With experience leading both product development and research, he has created award-winning edtech solutions with strong data science foundations.

Wanli Xing, Ph.D., University of Florida, wanli.xing@coe.ufl.edu
Wanli is an assistant professor of educational technology at the College of Education. His research themes are: (1) explore and leverage educational big data in various forms and modalities to advance the understanding of learning processes; (2) design and develop fair, accountable and transparent learning analytics, and AI-powered learning environments; (3) create innovative strategies, frameworks, and technologies for AI, Data Science, and STEM education.

Kenneth Koedinger, Ph.D., Carnegie Mellon University, koedinger@cmu.edu
Ken is the Hillman professor of Computer Science and Psychology at Carnegie Mellon University and founder of PLUS tutoring. He is a co-founder of CarnegieLearning, Inc. that has brought Cognitive Tutor based courses to millions of students since it was formed in 1998, and leads LearnLab, the scientific arm of CMU's Simon Initiative. Through extensive research and development in human-computer tutoring, Ken has demonstrated a doubling of math learning among middle school students, with future aims at bringing similar high-quality tutoring at scale. He has authored over 300 research papers and over 60 grant proposals.

References

[1] D. R. Chine, C. Brentley, C. Thomas-Browne, J. E. Richey, A. Gul, P. F. Carvalho, L. Branstetter, K. R. Koedinger, Educational equity through combined human-ai personalization: A propensity matching evaluation, in: International Conference on Artificial Intelligence in Education, Springer, Cham, 2022, pp. 366–377.

[2] A. Nickow, P. Oreopoulos, V. Quan, The impressive effects of tutoring on prek-12 learning: A systematic review and meta-analysis of the experimental evidence, National Bureau of Economic Research (NBER), Working paper # 27476 (2020).

[3] A. Alliance, America after 3pm: Demand grows, opportunity shrinks, Washington, DC: Afterschool Alliance (2020).

[4] M. A. Kraft, G. T. Falken, A blueprint for scaling tutoring across public schools. edworking-paper no. 20-335., Annenberg Institute for School Reform at Brown University (2021).

[5] J. F. Pane, B. A. Griffin, D. F. McCaffrey, R. Karam, Effectiveness of cognitive tutor algebra i at scale, Educational Evaluation and Policy Analysis 36 (2014) 127–144.

[6] S. Vincent-Lancrin, R. Van der Vlies, Trustworthy artificial intelligence (AI) in education: Promises and challenges, OECD Education Working Papers, No. 218 (2020).

[7] U.S. Department of Education. Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP) (2022).