Ethical Reflections on AI in the Development of a Teacher Dashboard

James Drimalla (University of Virginia)

Introduction

The mixed potential of artificial intelligence (AI) in education should compel developers of AI to discuss the ethical considerations of their work with teacher educators, teachers, and the broader public. As Shum and Luckin (2019), in the introduction to a special issue of the British Journal of Educational Technology on learning analytics and AI, said:

> If we do not want to see concerned students, parents and unions protesting against AI in education, we need urgently to communicate in accessible terms what the benefits of these new tools are, and equally, how seriously the community is engaging with their potential to be used to the detriment of society. (p. 2785)

Shum's and Luckin's call should prompt increased dialogue between mathematics education researchers, mathematics teacher educators (MTEs), and mathematics teachers in order to build trust in AI tools that have been responsibly developed and, ultimately, advance mathematics teachers' instruction. I am a member of a collaborative, interdisciplinary research project that uses artificial intelligence to classify instructional activities in videos of mathematics and English language arts (ELA) instruction as a means to accelerate feedback to teachers about their practice. In this manuscript, I describe (a) the AI-powered teacher dashboard being developed by our Artificial Intelligence for Advancing Instruction (AIAI) project and (b) the practical ethical considerations involved in its creation.

Description of Project and Teacher Dashboard

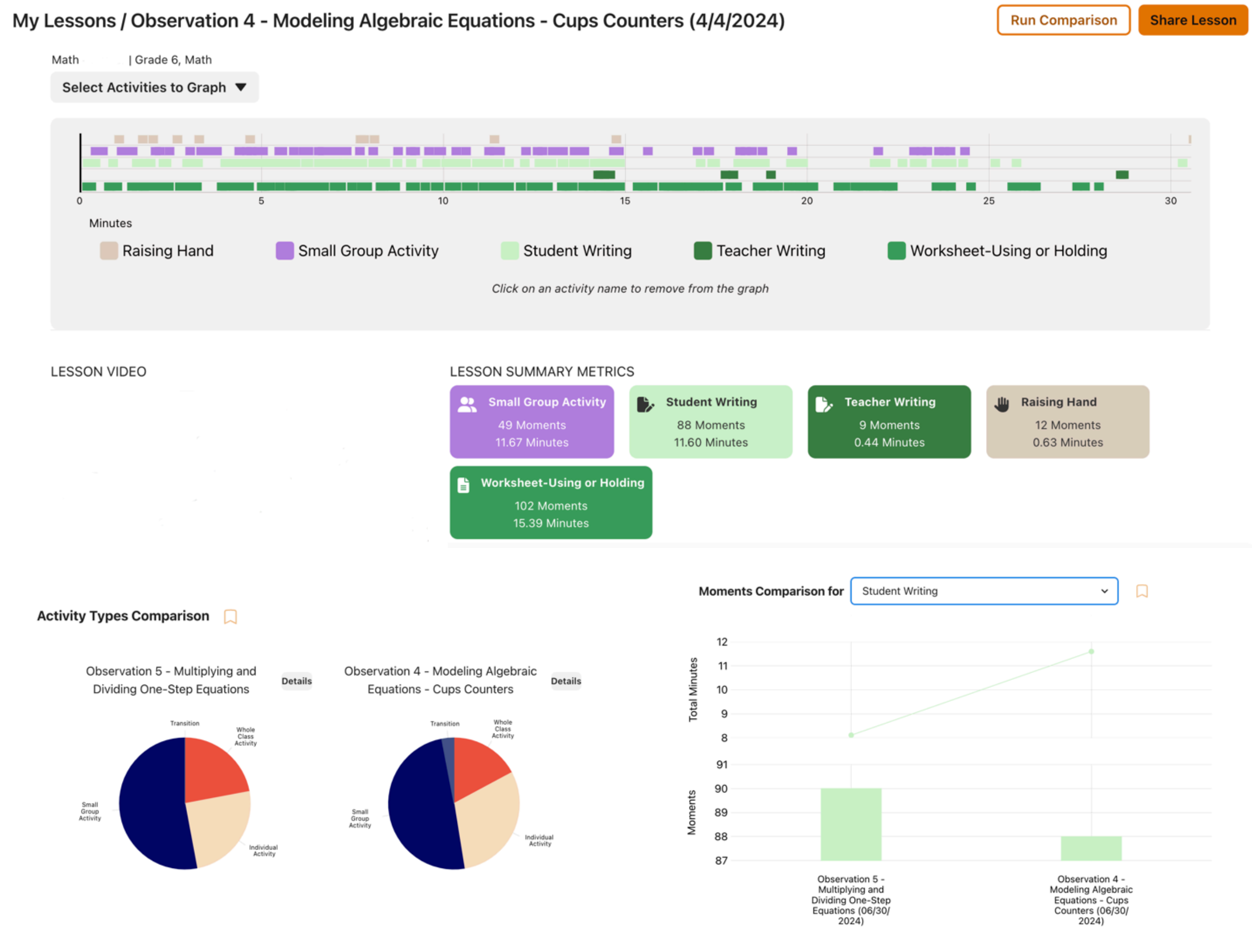
The purpose of the AIAI project is to make the process of reviewing and summarizing videos of instruction more efficient. We aim for more teachers to receive ongoing feedback on their instructional practices by automating the currently time-consuming process of manually reviewing instructional videos. We use neural networks—a type of AI that features a hierarchical learning structure that can learn and partition complex and abstract features of data—to classify components of instruction. With this structure, a neural network can carry out complicated decision functions that directly map input data to output labels and functions. At the beginning of our project, little was known about the feasibility of having neural networks discern complex activities such as representations of content, teacher questioning, and student engagement. Our project aimed to develop neural network models—with hundreds of hours of video of elementary instruction in mathematics and ELA—to identify these types of complex activities. Our results show that our neural network models can classify classroom activities with high levels of accuracy (Foster et al., 2024).

The video data of elementary mathematics and ELA lessons used to train our neural network models was also scored with two observation tools—The Mathematics Scan (M-Scan; Berry et al., 2013) and the Protocol for Language Arts Teaching Observation (PLATO; Grossman et al., 2014)—both designed to measure and assess effective instruction. We are currently investigating the relationship between the instructional activities classified by our neural network model and scores from these observation tools. Our initial results suggest it is feasible to automate the measurement for some features of ambitious mathematics instruction using machine learning algorithms and output from our neural network models (Foster & Youngs, 2023). Hypothetically, as our research develops, we could provide mathematics teachers with estimated scores from these observation tools based entirely on a video of their instruction.

Considering these promising results and in order to make the output from our neural network model valuable to teachers, we are developing and piloting a teacher dashboard that efficiently provides feedback to teachers on their instruction (Crimmins et al., 2023). It has become more common for teachers and teacher candidates to self-record their instruction to reflect on their teaching and receive targeted feedback but the process of manually reviewing videos is incredibly time-consuming. Our teacher dashboard uses output from our neural network model to provide lesson-level analytics to teachers about their individual lessons, segments within lessons, and cumulative practices across multiple lessons (see Figure 1).

Figure 1:

Lesson-Level Analytics from AIAI Teacher Dashboard



Ethical Considerations

The ethical considerations involved in creating our teacher dashboard primarily fall in two categories: the autonomy of teachers and the potential bias of our AI outputs.

Teacher Autonomy

Shum and Luckin (2019) describe how AI tools can be misused for surveillance and control of teachers rather than used for their benefit:

> The fears are reasonable: that quantification and autonomous systems provide a new wave of power tools to track and quantify human activity in ever higher resolution—a dream for bureaucrats, marketeers and researchers—but offer little to advance everyday teaching and learning in productive directions. (p. 2785)

Cathy O'Neil (2016), in her book on weapons of math destruction (WMDs), documents a case that reveals dangers like those described by Shum and Luckin. O'Neil details how, in an attempt to optimize Washington D.C.'s schools, an assessment tool was used to weed out "bad" teachers. The tool used a model which claimed to measure the effectiveness of teachers' teaching of mathematics and language skills but was dependent on assumptions about teacher effectiveness and its relation to student test scores. The model and its assumptions, however, was a black box—most people outside the organization that developed it did not understand its inner workings—but its output was treated as an objective, impartial judgement. Some teachers lost their jobs because of low scores from the model even though there was other evidence to suggest they were quality teachers. Although the model did not use AI, it is easy to imagine how AI could have amplified the power and reach of the model and caused more harm to teachers.

For this reason, it is important for developers of AI tools to be transparent about the tool's design and communicate the values that guided the design process. Our teacher dashboard is designed to support teachers—not replace them. We intentionally seek to advance teachers' instruction—to propel it further along the path of equitable and rigorous instruction. Thus, the dashboard's design aims to maximize teacher autonomy and agency. Teachers can choose if and when they upload videos of their lessons to the dashboard and can choose who they share their lesson and the analytics with (e.g., colleagues, instructional coach, university supervisor, etc.).

Potential Bias

The National Council of Teachers of Mathematics' (NCTM) position statement on AI calls for mathematics educators to (a) be involved in developing and testing AI tools and (b) be aware of AI tools trained on biased datasets (NCTM, 2023). In response to NCTM's call and others like it, our project aims to be transparent about our dataset and any potential bias. Our neural network models were trained on hundreds of hours of video of elementary instruction in mathematics and ELA in which approximately 91% of the teachers identified as White, 5% as African American, 2.5% as Asian, and just over 1% as Hispanic. 92.5% of the teachers in the videos identified as female and 7.5% identified as male. The race and ethnicity of the students at the schools where the data was collected were approximately 50% White, 21% Hispanic, 16% African American, 6% Asian, and 5% two or more races. Finally, approximately 52% of the students at the schools qualified for free or reduced-price lunch.

Considering these details, we are exploring how these teacher and student demographics may influence the results of our neural network models and machine learning algorithms. For example, in their study of the elementary teachers' development of ambitious instruction in our data, Youngs et al. (2022) said their "analyses revealed that first-year teachers in schools with higher percentages of students eligible for free or reduced-price lunch were more likely to receive low [PLATO] ratings for disciplinary demand in mathematics and ELA" (p. 12). These disparities were not unique to PLATO ratings: we found similar disparities in the ratings of various M-Scan dimensions in our data too. Although these outcomes suggest the raw data used to train our neural network models and machine learning algorithms have limitations, knowing these limitations exist empowers us to take subsequent steps to minimize biased results and be cautious in interpreting results from AI.

Conclusion

MTEs, in addition to their responsibility to introduce prospective and in-service teachers to relevant technology, need to help teachers develop wisdom about the use of AI tools. For MTEs to do that, they need developers of AI tools (and the MTEs that work with them) to be transparent about their work, their motivations, and how they are addressing ethical issues. To that end, I offer this piece as the beginning of an honest and transparent conversation among MTEs about the potential benefits and limitations of AI tools.

References

Berry, R. Q., Rimm-Kaufman, S. E., Ottmar, E. M., Walkowiak, T. A., Merritt, E. G., & Pinter, H. H. (2013). *The Mathematics Scan (M-Scan): A measure of standards-based mathematics teaching practices*. University of Virginia.

Crimmins, S. B., Foster, J. K., & Youngs, P. (2023). *Automated neural network dashboards for teacher feedback*. Paper presented at the 2023 annual meeting of the American Education Research Association, Chicago, IL. https://doi.org/10.3102/2012634

Foster, J. K. & Youngs, P. (2023). *Using Machine Learning to Score M-Scan: Is Automating the Measurement of Ambitious Mathematics Instruction Feasible?* Research report presented at National Council of Teachers of Mathematics (NCTM) Research Conference, Washington, DC.

Foster, J. K., Korban, M., Youngs, P., Watson, G. S., & Acton, S. T. (2024). Automatic classification of activities in classroom videos. *Computers and Education: Artificial Intelligence*, 100207. https://doi.org/10.1016/j.caeai.2024.100207

Grossman, P., Cohen, J., Ronfeldt, M., & Brown, L. (2014). The test matters: The relationship between classroom observation scores and teacher value added on multiple types of assessment. *Educational Researcher, 43*(6), 293–303. https://doi.org/10.3102/0013189X14544542

National Council of Teachers of Mathematics. (2023). Artificial Intelligence and Mathematics Teaching. Retrieved June 28, 2024 from : https://www.nctm.org/standards-and-positions/Position-Statements/Artificial-Intellige nce-and-Mathematics-Teaching/

O'Neil, C. (2016). *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy*. Crown.

Shum, S. J. B., & Luckin, R. (2019). Learning analytics and AI: Politics, pedagogy and practices. *British Journal of Educational Technology*, *50*(6), 2785-2793.

Youngs, P., Elreda, L. M., Anagnostopoulos, D., Cohen, J., Drake, C., & Konstantopoulos, S. (2022). The development of ambitious instruction: How beginning elementary teachers' preparation experiences are associated with their mathematics and English language arts instructional practices. *Teaching and Teacher Education, 110,* 103576. https://doi.org/10.1016/j.tate.2021.103576