Implications of Advances in Artificial Intelligence (AI) for 10 Areas of Work in Educational Assessment and Accountability

André A. Rupp & Will Lorié, Senior Associates

April 19, 2023

This list is the companion to a blog post, "Ready or Not: AI is Changing Assessment and Accountability," that highlights key implications of artificial intelligence for assessment and accountability. This table details 25 implications in 10 areas of work.

Area 1: Competency Selection

1.1 Increased attention on competencies that drive the development of AI itself, such as machine learning, computer programming, and computational thinking

Opportunities	Stronger pushes by business sectors for the development of these skills, leading to updating curricula and associated instructional approaches
Risks	Further marginalization of integrated, cross-disciplinary curricula or learning experiences under STEAM

1.2 Increased focus on critical thinking and more complex reasoning within curricula in even earlier grades to identify, elicit, evaluate, and utilize AI-generated information

Opportunities	Increased focus on the development of cross-disciplinary skills and a reconceptualization of what young learners can do under appropriate conditions
Risks	Further increase in achievement gaps due to insufficient or differential capacity and expertise to develop these skills across grades by teachers and teacher trainers

Area 2: Competency Development

2.1 Changes in the specification of learning progressions for advanced skill areas and associated performance expectations due to changes in the nature of tasks and contexts that shape learning

Opportunities	More critical evaluations of how learner development is conceptualized and operationalized, leading to a more differentiated understanding of learning pathways and associated learning supports
Risks	Over-reliance on existing learning progressions and associated research programs as the dominant "gold standard" for data collection and validation even though updated progressions and methods are required

2.2 Increased attention to the development of essential competencies in teachers and the ways teachers use these to enact future-oriented curricula

Opportunities	Changes to teacher training programs as well as practices for hiring, in-service training, and ongoing recognition of skills development
Risks	Increasing pressures on the teacher pipeline that exacerbate current challenges due to shortfalls in available educational programs, financial compensation, or policy supports

Area 3: Task Development

3.1 Increased focus on design-oriented, project-based tasks that require a richer use of relevant competencies supported by emerging technologies such as augmented reality, virtual reality, and generative interfaces

Opportunities	Further push to change dated models of learning and assessment that are focused too intensely on knowledge-reproduction activities
Risks	Further increase in gaps in learning opportunities for districts or states that do not have sufficient industry partnerships or other resources to put such curricula in place

3.2 Increased efficiencies in the automated tagging of data from verbal report studies such as think-aloud sessions and cognitive labs for large-scale assessment

Opportunities	Faster data-collection and analysis cycles with more differentiated analyses due to more complex tagging schemes
Risks	Insufficient human oversight in tagging routines, leading to uncritical use of provided transcripts

3.3 Automated generation of candidate tasks or stimulus materials for consideration in pilots or pre-trials for large-scale assessment specifically

Opportunities	More efficient use of human expertise to select and adapt tasks and materials and a deeper understanding of essential task family characteristics
Risks	Insufficient critical selection of automatically generated materials along cognitively- or metacognitively-relevant characteristics

3.4 Increased efficiencies in the automated translation and adaptation of stimuli, response options, scoring rubrics, and other materials into other languages or for accommodations

Opportunities	Easier access to learning and assessment resources for learners who speak a wider variety of native languages or who require accommodations across a wider variety of educational contexts
Risks	For assessments in particular, insufficient human expert attention to subtle translation changes that can increase biases for new learner populations and compromise comparability in higher-stakes contexts

Area 4: Test Assembly and Delivery

4.1 Increased efficiencies in automated form assembly or computer-adaptive task selection in large-scale assessment

Opportunities	Increased use of assessments that include a broader range of tasks targeting relevant competencies in more appropriate ways
Risks	Insufficient investigations into the technical quality of information gathered over time for higher-stakes, summative determinations

4.2 Increased efficiencies in task selection and delivery within personalized learning systems

Opportunities	Increased use of personalized learning systems that include more complex tasks
Risks	Over-reliance on personalized learning systems at the expense of teacher professional development that supports smart-system use

4.3 Increased consideration of issues around cheating or atypical responding for tasks that are within the realm of powerful AI performance

Opportunities	Shift away from simplistic tasks and an increase in the development of complex automated flagging algorithms
Risks	Regression towards outdated models of in-person assessment that prevent learners from using technology to solve problems

Area 5: Reporting and Data Interactions

5.1 Automated formative suggestions for learning with increased customization of student output within personalized learning systems

Opportunities	Stronger evidence about and support for ongoing learning that can take place within and outside of classrooms, even for complex competencies
Risks	Insufficient monitoring of the comprehensibility and effectiveness of automated feedback mechanisms

5.2 Automated mentoring-oriented chat agents that can interact with a student to help them understand their performance and provide relevant data and visualizations

Opportunities	Quicker access to relevant information for stakeholders and an associated deeper understanding of how different people identify and process essential information
Risks	Insufficient guidance by systems coupled with an over-reliance on such guidance that leads to misuse of the information for instructional and/or self-directed decision-making

Area 6: Data Engineering

6.1 Increased focus on the streamlined collection of data from multiple sources within a coherent data architecture

Opportunities	Stronger recognition of the importance of coherent data systems and the ability to make rich and accurate inferences due to relevant data being available promptly
Risks	Insufficient consideration of the complexity and expertise required to design such systems that lead to challenges in the sustainability of the implementation

6.2 Increased focus on hiring staff with advanced data engineering and data science skills for system design and maintenance

Opportunities	Stronger opportunities for cross-team collaboration that integrates data engineering and visualization experts into design and implementation cycles to create more effective solutions
Risks	Unproductive siloing within teams despite the increase in staff

Area 7: Scoring Design

7.1 Automated generation of sample responses for pre-trialing tasks

Opportunities	Reduction of resources needed for pre-trialing tasks, especially in early development stages
Risks	Insufficient training of algorithms resulting in skewed exemplar responses that, in turn, lead to inappropriate judgments about tasks and expectations about performance

7.2 Automated tagging of data from logfiles/process data to support the design of relevant indicators

Opportunities	Increased efficiencies in the design of relevant indicators and the inclusion of a broader set of evidence from performance on a more comprehensive set of tasks
Risks	Insufficient evaluation of the proposed tags and possible biases for decision-making that may arise from their use

7.3 Automated exploration and/or generation of statistical models for data analysis with integrated model-data fit information

Opportunities	Increased efficiencies in the evaluation of a broader range of models that allow for a more differentiated representation of student competencies across common statistical frameworks
Risks	Insufficient exploration of issues that may lead to model-data misfit and a continued overreliance on simple, unidimensional approaches for scaling across too many contexts

Area 8: Validation Approaches

8.1 Increased need for data from other assessments and measures of learning outcomes that capture similar novel constructs

Opportunities	More differentiated understanding of how to measure complex competencies across contexts and the limits of assessment evidence in doing so
Risks	Over-reliance on coarse metrics from the contributing assessments, such as correlations that do not sufficiently leverage and tease out relevant information for the competencies in question

8.2 New conceptualizations, operationalizations, and methodological approaches for investigating issues of differential functioning, bias, fairness, and related aspects

Opportunities	More robust systems-thinking approach to the conceptualization of these issues that considers opportunities to learn, sociocultural factors, and technical characteristics in conjunction, among others
Risks	Over-reliance on standard methodologies to new data streams without sufficient critical reflection on the limits of inferential insight in context

8.3 Increased need to differentiate learner-generated and machine-generated artifacts in higher stakes settings

Opportunities	Increase in the use of more interdisciplinary tasks along with tasks that cover a more comprehensive range of spaces within disciplines
Risks	Over-emphasis on cheating as a central issue in learning and assessment and an associated increase in in-person or even paper-based testing that corrupts the assessment goals

Area 9: Accountability Approaches

9.1 Inclusion of new indicators that capture the development of new relevant competencies in local and federal systems

Opportunities	More robust signaling and rewarding of the development of these competencies, leading to systemic changes that create more differentiated pathways for college- and career readiness across a broader range of institutions
Risks	Over-reliance on aggregate indicators for novel competencies instead of multi-faceted indicator systems and dashboards that more appropriately represent learner competencies

9.2 Increased value of dual credit courses, apprenticeships, and professional certification in high school

Opportunities	Stronger partnerships between industry and districts that provide a wider range of accepted opportunities for more students
Risks	Increase in opportunity gaps for students from communities with impoverished extracurricular opportunities

Area 10: Ethical Considerations and Data Privacy

10.1 Balanced data collection, privacy, and security while implementing AI-driven personalized learning and assessment systems

Opportunities	Improved understanding of students' needs, leading to more optimally targeted interventions and enhanced learning experiences
Risks	Potential misuse of personal information, infringement of privacy rights, and security breaches

10.2 Greater attention to the ethical implications of AI-driven learning and assessment systems

Opportunities	Increased awareness and responsibility among educators, policymakers, and AI developers in ensuring fairness, transparency, and inclusiveness in AI applications
Risks	Unintended biases, unfairness, or discrimination arising from AI systems, exacerbating existing inequalities

Fun Fact:

The text for Area 10, subareas 10.1 and 10.2, and associated opportunities and risks is lightly edited *ChatGPT* output, operating under the GPT-4 model. On March 20, 2023, we prompted the model with all the preceding text in the table and asked it to complete the text for a tenth area.