

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

André A. Rupp & Will Lorie, Senior Associates  
April 12, 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 organized into 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

<b>Opportunities</b>	Stronger pushes by business sectors for the development of these skills, leading to updating curricula and associated instructional approaches
<b>Risks</b>	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

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

### 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

<b>Opportunities</b>	More critical evaluations of how learner development is conceptualized and operationalized, leading to a more differentiated understanding of learning pathways and associated learning supports
<b>Risks</b>	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

<b>Opportunities</b>	Changes to teacher training programs as well as practices for hiring, in-service training, and ongoing recognition of skills development
<b>Risks</b>	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

<b>Opportunities</b>	Further push to change dated models of learning and assessment that are focused too intensely on knowledge-reproduction activities
<b>Risks</b>	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

<b>Opportunities</b>	Faster data-collection and analysis cycles with more differentiated analyses due to more complex tagging schemes
<b>Risks</b>	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

<b>Opportunities</b>	More efficient use of human expertise to select and adapt tasks and materials and a deeper understanding of essential task family characteristics
<b>Risks</b>	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

<b>Opportunities</b>	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
<b>Risks</b>	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

<b>Opportunities</b>	Increased use of assessments that include a broader range of tasks targeting relevant competencies in more appropriate ways
<b>Risks</b>	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

<b>Opportunities</b>	Increased use of personalized learning systems that include more complex tasks
<b>Risks</b>	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

<b>Opportunities</b>	Shift away from simplistic tasks and an increase in the development of complex automated flagging algorithms
<b>Risks</b>	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

<b>Opportunities</b>	Stronger evidence about and support for ongoing learning that can take place within and outside of classrooms, even for complex competencies
<b>Risks</b>	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

<b>Opportunities</b>	Quicker access to relevant information for stakeholders and an associated deeper understanding of how different people identify and process essential information
<b>Risks</b>	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

<b>Opportunities</b>	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
<b>Risks</b>	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

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

## Area 7: Scoring Design

### 7.1 Automated generation of sample responses for pre-trialing tasks

<b>Opportunities</b>	Reduction of resources needed for pre-trialing tasks, especially in early development stages
<b>Risks</b>	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

<b>Opportunities</b>	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
<b>Risks</b>	Insufficient evaluation of the proposed tags and possible biases for decision-making that may arise from their use

### 7.3 Automated proposition of statistical models for data analysis with integrated model-data fit information

<b>Opportunities</b>	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
<b>Risks</b>	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

<b>Opportunities</b>	More differentiated understanding of how to measure complex competencies across contexts and the limits of assessment evidence in doing so
<b>Risks</b>	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

<b>Opportunities</b>	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
<b>Risks</b>	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

<b>Opportunities</b>	Increase in the use of more interdisciplinary tasks along with tasks that cover a more comprehensive range of spaces within disciplines
<b>Risks</b>	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

<b>Opportunities</b>	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
<b>Risks</b>	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

<b>Opportunities</b>	Stronger partnerships between industry and districts that provide a wider range of accepted opportunities for more students
<b>Risks</b>	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

<b>Opportunities</b>	Improved understanding of students' needs, leading to more optimally targeted interventions and enhanced learning experiences
<b>Risks</b>	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

<b>Opportunities</b>	Increased awareness and responsibility of educators, policymakers, and AI developers in ensuring fairness, transparency, and inclusiveness in AI applications
<b>Risks</b>	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.