Navigating the AI Landscape:
A Framework for Evaluating Assessment Tools in Higher Education

PRESENTED BY: Kevin R. Guidry
Joshua Wilson

29 November 2023
Learning goals

- Contextualize automation/AI assessment
- Engage in ethical pragmatism in assessment decisions
- Consider elements of our proposed framework when making assessment decisions
- Apply the framework to an example
Overview

1. Context - uses of AI/automation in assessment in higher education
2. Key considerations for evaluating AI assessment tools
3. Overview of flowchart
4. Example
5. Questions and discussion
Context

Automation/AI in assessment for higher ed is not entirely new

- Automation previously used solely with constrained assessment formats e.g., multiple choice questions in Scantron instruments.
- Summative examples using “Predictive AI” (vs. generative AI)
  - GRE
  - TOEFL
  - Writing placement exams (e.g., “Accuplacer”)
What is new is the “generative” component

- Formative/everyday examples using “Generative AI”
  - Grammarly
  - ChatGPT (provide feedback, enter rubric and have it assess, generate assessment items)
  - Packback
  - Perusall

Also new is the accessibility component (i.e., no code versions)

- Automation/Al for assessment is no longer just for computer scientists. It’s accessible to professors in English, music, arts, education, and it’s now being used in loosely structured assessment formats (e.g., writing, performance assessments of different types)
Big Ideas

Factors driving adoption of automation/AI assessment

- Scale of instructor’s grading/feedback load
- Independent of scale, the effort required for grading/feedback
- Technological advances and widespread availability of assessment options

Benefits

- Provide students more immediate feedback
- More consistent/objective* than instructors
- Beneficial division of labor for instructor
  - Useful analogy: Portions of assessment and feedback have been and are increasingly delegated to others e.g., TAs

Concerns

- Replacing essential instructor functions
- Inaccuracy and bias
Key Considerations
Key Considerations for Evaluating AI Assessments in HE

1. Assessment Purpose
2. Pedagogical Alignment
3. Technical Robustness*
4. Ethical Considerations*
5. Explainability
6. Community and Stakeholder Engagement*
7. Evaluation and Continuous Improvement*

* indicates new or unique considerations for an AI/automated assessment
1. Assessment Purpose

- What is the purpose of the assessment to which the AI/automated tool will be applied?
  - Summative? - applying a grade, evaluating learning outcomes
  - Formative? - assessment to support learning while learning is occurring (i.e., feedback to support improvement)

- **Big Idea:** The more stakes attached to the assessment, the more important it is for the faculty member to make an informed decision based on additional considerations.
2. Pedagogical Alignment

- Ensure that the AI tool aligns well with the learning objectives of the course or assessment.
- Check the capability of the AI tool in providing immediate, personalized, and constructive feedback to students. This is perhaps the best application of AI and automation for assessment.

**Big Idea:** The decision to adopt an automated/AI assessment tool should be because it advances instructional priorities and learning outcomes, NOT because it will save the instructor time (although that’s nice, too)
3. Technical Robustness*

- Valid inferences about students’ performance are only possible when an automated/AI tool produces scores/feedback that is (a) reliable, (b) construct relevant, and (c) used in a way that was aligned with its originally developed use case.

- Canvas integration and accessibility: Feasibility is increased when an automated/AI tool is integrated within Canvas.

- **Big Idea:** Technical qualities of the scores and technical issues related to integration are critical considerations for automated/AI assessments.
4. Ethical Considerations*

- Equitable access and bias
  - Ensure all students have equitable access to the automated/AI tool and the learning opportunities it provides.
  - Ensure that the assessment/scoring/feedback is not biased and will not unfairly advantage/disadvantage certain groups of students (i.e., algorithmic bias)

- Data privacy and security
  - Understand how student data is used (if at all) and protected and the rights that are given to the company when using the tool

- **Big Idea:** Equity, access, and data privacy must be carefully considered for automated/AI assessments.
5. Explainability

- Lack of detailed knowledge of assessment and measurement is a systematic and systemic issue
- Ensure that you are able to explain in basic terms how the AI tool operates, the data it collects, and the decisions it makes.

**Big Idea:** You’re the one that submits grades to registrar, so you have to stand by those grades and feel that those are valid. If you cannot explain in basic terms how the automated/AI system works, this is a red flag.
6. **Stakeholder and Community Engagement**

- Engage with students in decision-making processes related to AI tool adoption.
- Ensure that you hear from a multiplicity of voices. Cast a wide net and offer multiple means of sharing feedback to ensure that the piloted automated/AI assessment is working not just for the instructor, but the students as well.

- **Big Idea:** Students, as the those most directly affected by the automated/AI assessment, should have opportunities to share their perspective and feedback. That feedback should be carefully weighed.
7. Evaluation and Continuous Improvement*

- Regularly evaluate the effectiveness of the AI tool in achieving desired learning outcomes.
- Faculty less comfortable with planning an evaluation, should consider partnering with CTAL or SOE faculty to collaborative design, implement, and analyze such an evaluation.

  **Big Idea:** Pilot → evaluate → adjust and improve → rinse and repeat
Flowchart to Aid Decision Making
Example: Perusall
Example - *Perusall*

Perusall can "grade" student comments:

"Perusall uses a machine learning algorithm that uses linguistic features of the text to create a predictive model for the score a human instructor would give....*From Perusall's Perspective, we are trying to save an instructor time by suggesting a score*. By default, will we not show students scores until you are ready to release and approve. (emphasis added)"
in the preceding two chapters, we developed a mathematical framework for describing motion along a straight line. in this chapter, we continue our study of motion by investigating inertia, a property of objects that affects their motion. the experiments we carry out in studying inertia lead us to discover one of the most fundamental laws in physics—conservation of momentum.

4.1 friction

despite the smoothness of the board and the smoothness of the wooden surface, this stopping happens so slowly that you may not notice it. if you have a smooth surface that is not very smooth and slippery, the block slides for a longer time interval than if it is on a smooth surface. this is because the block slides for a longer time interval than if it is on a smooth surface. the friction slowing down is due to the resistance to motion that one surface or object encounters when moving over another.

notice that, during the interval covered by the velocity versus-time graph, the velocity decreases as the block slides over ice, which is slowly.

the block slides slowly over ice because of friction between the two surfaces. while the friction is to the left of the block to rest with respect to each other. in this case, the wooden block and the surface it is sliding on. the less friction there is, the longer it takes for the block to come to rest.

another advantage of using such surfaces is that the track constrains the motion being along a straight line. we can then use a high-speed camera to record the cart's position at various instants, and from that information determine its speed and acceleration.

4.2 inertia

we can discover one of the most fundamental principles of physics by studying the velocities of two low-friction carts change when the carts collide. the most interesting thing about this example is that the frictional force acting on the object is that object will maintain motion after the effects of inertia.

4.3 work and energy

the work done by an object as it moves is equal to the force applied to the object times the distance the object has been moved. as an example of this, consider two objects, one which can be represented by a mass and the other as a spring. the mass is pushed into the spring, which is compressed and then allows the mass to move. the work done by the spring is then equal to the force applied to the mass times the distance the mass has moved. the energy of the system is then equal to the work done by the spring plus the potential energy stored in the spring.

4.4 mechanical energy

the mechanical energy of a system is the sum of the kinetic and potential energy of all the objects in the system. the total mechanical energy of a system is then equal to the initial mechanical energy of all the objects in the system times the total change in the system's energy.

4.5 conservation of energy

the law of conservation of energy states that the total energy in a closed system remains constant. this law is a fundamental principle of physics that applies to a wide variety of physical systems, including mechanical, electrical, and thermal systems.
Example - *Perusall*

**Allison: Meets expectations**

Allison’s comments reveal interpretation of the text and demonstrate her understanding of concepts through analogy and synthesis of multiple concepts. Her responses are thoughtful explanations with substantiated claims and/or concrete examples. She also poses a profound question that goes beyond the material covered in the text. Finally, she applies understanding of graphical representation to explain the relationship between concepts.

**Beth: Improvement needed**

While Beth asks possibly insightful questions, she does not elaborate on thought process. She demonstrates superficial reading, but no thoughtful reading or interpretation of the text. When responding to other students’ questions, she demonstrates some thought but does not really address the question posed.

**Cory: Deficient**

Cory’s comments have no real substance and do not demonstrate any thoughtful reading or interpretation of the text. His questions do not explicitly identify points of confusion. Moreover, his comments are not backed up by any reasoning or assumptions.
Example - *Perusall*

- How should a faculty member decide to use Perusall's comment scoring feature?
- What information is needed to make that decision?
- What factors are important in making that decision?
Questions and Discussion
Key questions to ask when considering AI in assessment

In addition to questions explored for every assessment decision (e.g., alignment with purpose and goals, ability to explain and take responsibility for evaluations and feedback):

- Technical robustness: Is the tool (a) reliable, (b) construct relevant, and (c) used in a way that was aligned with its originally developed use case?
- Ethical considerations: Do all students have equitable access? Is it unbiased? Are data protected?
- Stakeholder and community engagement: Have students and others been consulted to ensure the tool works for them, too?
- Evaluation and continuous improvement: Do you have plans to regularly evaluate the effectiveness of the tool and make adjustments?
Kevin R. Guidry
Associate Director of Educational Assessment
Center for Teaching and Assessment of Learning
krguidry@udel.edu

Joshua Wilson
Associate Professor, School of Education
College of Education and Human Development
joshwils@udel.edu