AI FOR TEACHING AND LEARNING:

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

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”)

Context (cont.)

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/AI 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*

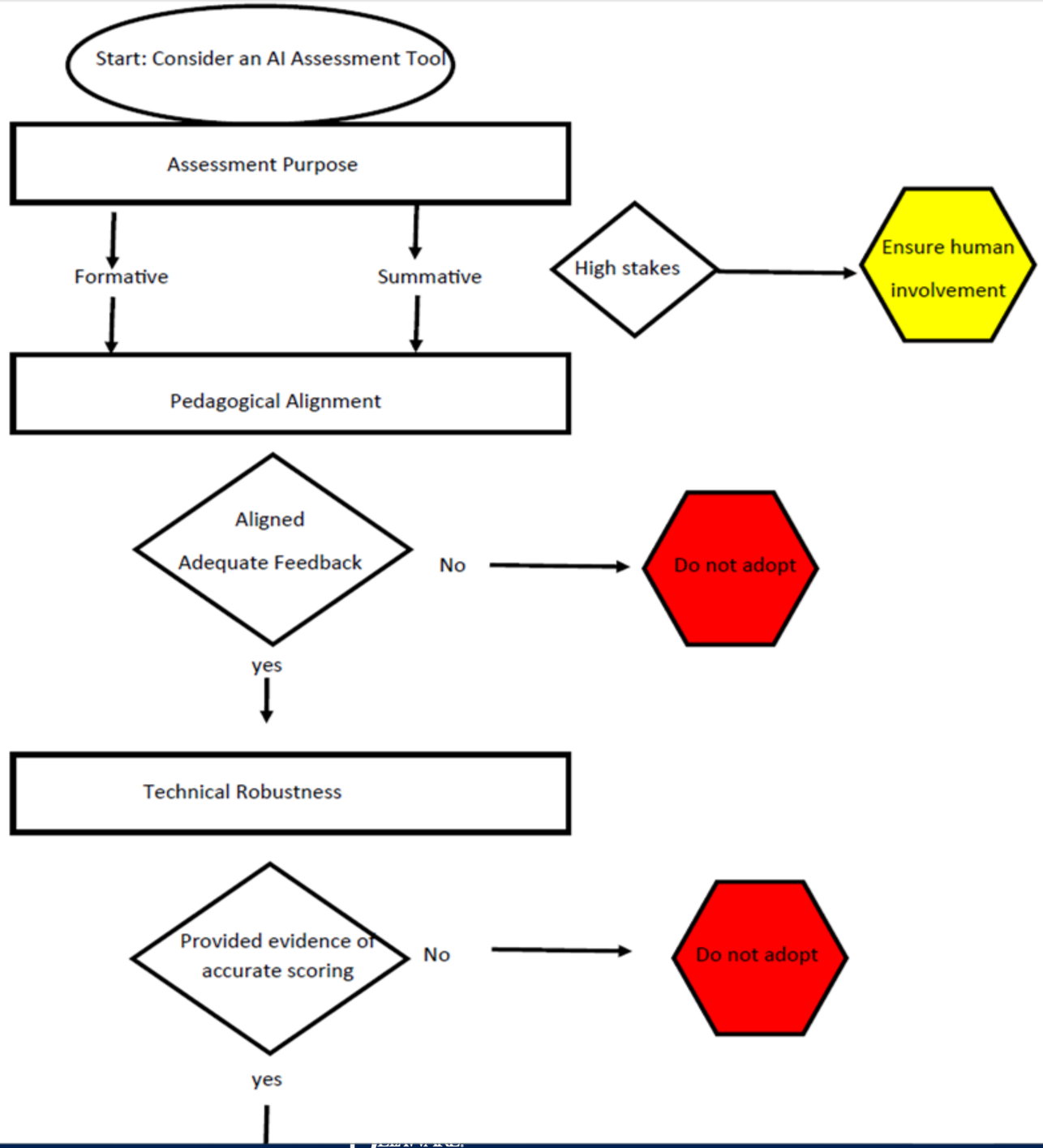
- Engage with students in decision-making processes related to AI tool adoption.
- Ensure that you hear from a multiplicity of voices. Cast a wide net and offer multiple means of sharing feedback to ensure that the piloted automated/AI assessment is working not just for the instructor, but the students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

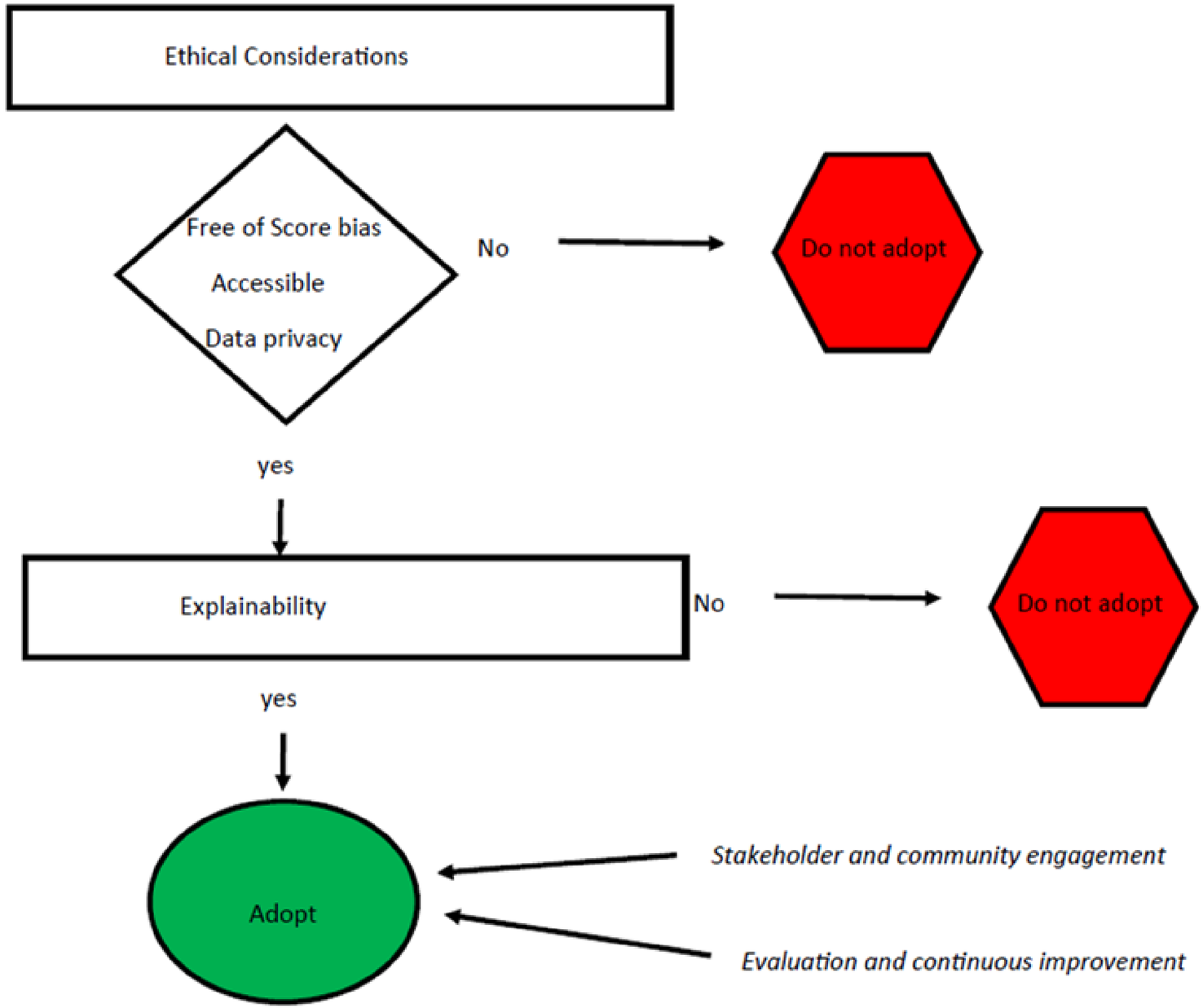
*7. Evaluation and Continuous Improvement**

- 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


Start: Consider an AI Assessment Tool
Assessment Purpose
Formative
Summative
High stakes
Ensure human involvement
Pedagogical Alignment
Aligned Adequate Feedback
No
Do not adopt
yes
Technical Robustness
Provided evidence of accurate scoring
No
Do not adopt
yes


Ethical Considerations
Free of Score bias
Accessible
Data privacy
No
Do not adopt
yes
Explainability
No
Do not adopt
yes
Adopt
Stakeholder and community engagement
Evaluation and continuous improvement

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)"

https://support.perusall.com/hc/en-us/articles/360034824694-How-is-comment-quality-defined-in-Perusall-

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

Picture a block of wood sitting motionless on a smooth wooden surface. If you give the block a shove, it slides some distance but eventually comes to rest. Depending on the smoothness of the block and the smoothness of the wooden surface, this stopping may happen sooner or it may happen later. If the two surfaces in contact are very smooth and slippery, the block slides for a longer time interval than if the surfaces are rough or sticky. This you know from everyday experience: A hockey puck slides easily on ice but not on a rough road.

Figure 4.1 shows how the velocity of a wooden block decreases on three different surfaces. The slowing down is due to friction—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 decrease as the block slides over ice is hardly observable. The block slides easily over ice because there is very little friction between the two surfaces. The effect of friction is to bring two objects 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.

Figure 4.1 Velocity-versus-time graph for a wooden block sliding on three different surfaces. The rougher the surface, the more quickly the velocity decreases.

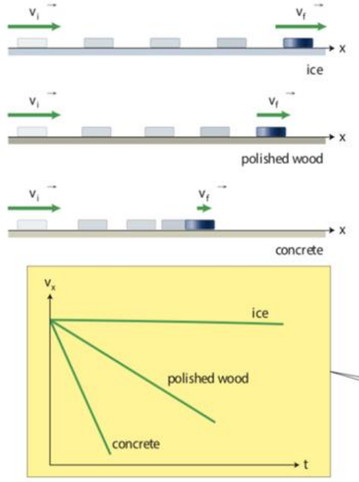


Figure 4.2 Low-friction track and carts used in the experiments described in this chapter.

You may wonder whether it is possible to make surfaces that have no friction at all, such that an object, once given a shove, continues to glide forever. There is no totally frictionless surface over which objects slide forever, but there are ways to minimize friction. You can, for instance, float an object on a cushion of air. This is most easily accomplished with a low-friction track—a track whose surface is dotted with little holes through which pressurized air blows. The air serves as a cushion on which a conveniently shaped object can float, with friction between the object and the track all but eliminated. Alternatively, one can use wheeled carts with low-friction bearings on an ordinary track. Figure 4.2 shows low-friction carts you may have encountered in your lab or class. Although there is still some friction both for low-friction tracks and for the track shown in Figure 4.2, this friction is so small that it can be ignored during an experiment. For example, if the track in Figure 4.2 is horizontal, carts move along its length without slowing down appreciably. In other words:

> In the absence of friction, objects moving along a horizontal track keep moving without slowing down.

Another advantage of using such carts is that the track constrains the motion to 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.

CONCEPTS

4.1 (a) Are the accelerations of the motions shown in Figure 4.1 constant? (b) For which surface is the acceleration largest in magnitude?

4.2 Inertia

We can discover one of the most fundamental principles of physics by studying how the velocities of two low-friction carts change when the carts collide. Let's first see what happens with two identical carts. We call these standard carts because we'll use them as a standard against which to compare the motion of other carts. First we put one standard cart on the low-friction track and make sure it doesn't move. Next we place the second cart on the track some distance from the first one and give the second cart a shove toward the first. The two carts collide, and the collision alters the velocities of both.

Cory: I remember, in high school, being amazed at how quickly carts could travel on these tracks - air would blow up through these tiny holes evenly distributed along the length of the track and the cart would essentially float on the air and consequently - the cart would move very quickly with the slightest push.

Alison: Although there is no way to create frictionless surfaces, I find it interesting that we consider experiments "in the absence of friction." In a way, this relates back to Chapter 1.5 where we talked about the importance of having too little or too much information in our representations. In some cases, the friction is so insignificant that we ignore it (simplifying our representation).

Beth: Does this only apply to solid surfaces? I feel as if a sub - stance that floats on water either has negligible or very little friction.

Cory: Why is this? I don't get it.

Alison: I believe this applies to almost every surface, although I'm not sure if water would count more as resistance than friction. Anyways, the best example I could think of would be a surf board. If people who were paddling in the same direction as the waves experienced no resistance, they would continually speed up, and eventually reach very high speeds. However, in reality if they were two stop paddling they'd slow down and only the waves would slowly push them to shore.

Beth: Is it possible to have a surface, in real life, that inflicts NO friction at all?

Beth: Doesn't air resistance factor into this at all?

Alison: The key word is "appreciably". In the absense of friction, the cart does not slow down appreciably but still would a little due to air resistance

Cory: a) yes b) concrete has the acceleration of greatest magnitude

Beth: I would think that they are not constant because if we think of the formula F=ma, the force of friction is different in every case.

Alison: As a theoretical question about inertia, if an object in motion will stay in motion, but is being affected by friction, will it slow down perpetually but remain in motion, or will it eventually stop completely due to the friction? Just curious.

Beth: With friction everything slows down to a half at one point or another. It is only if an outside force acts on the object if that object will maintain motion after the effects of inertia.

Cory: Standard carts: identical carts in mass, shape, etc. I like this notion of standard carts, it provides a good baseline to compare other motion and to understand the concepts before building on it.

Cory: Great visual representation of friction! It is interesting how this compares the velocity of things on different surfaces

Alison: The rougher the surface, the more friction between the surface and the wooden block, and thus acceleration will be greater.

Example - *Perusall*

Allison: Meets expectations

Alison'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

Thank you!