

# Introducing more authentic assessment in the classroom

**Jennifer Hill, Ed.D.**

Trinity College of Arts & Sciences  
Director of Assessment

**Kim Manturuk, Ph.D.**

Duke Learning Innovation  
Associate Director of Research, Evaluation, & Development



# Today's work

Define “authentic assessment”

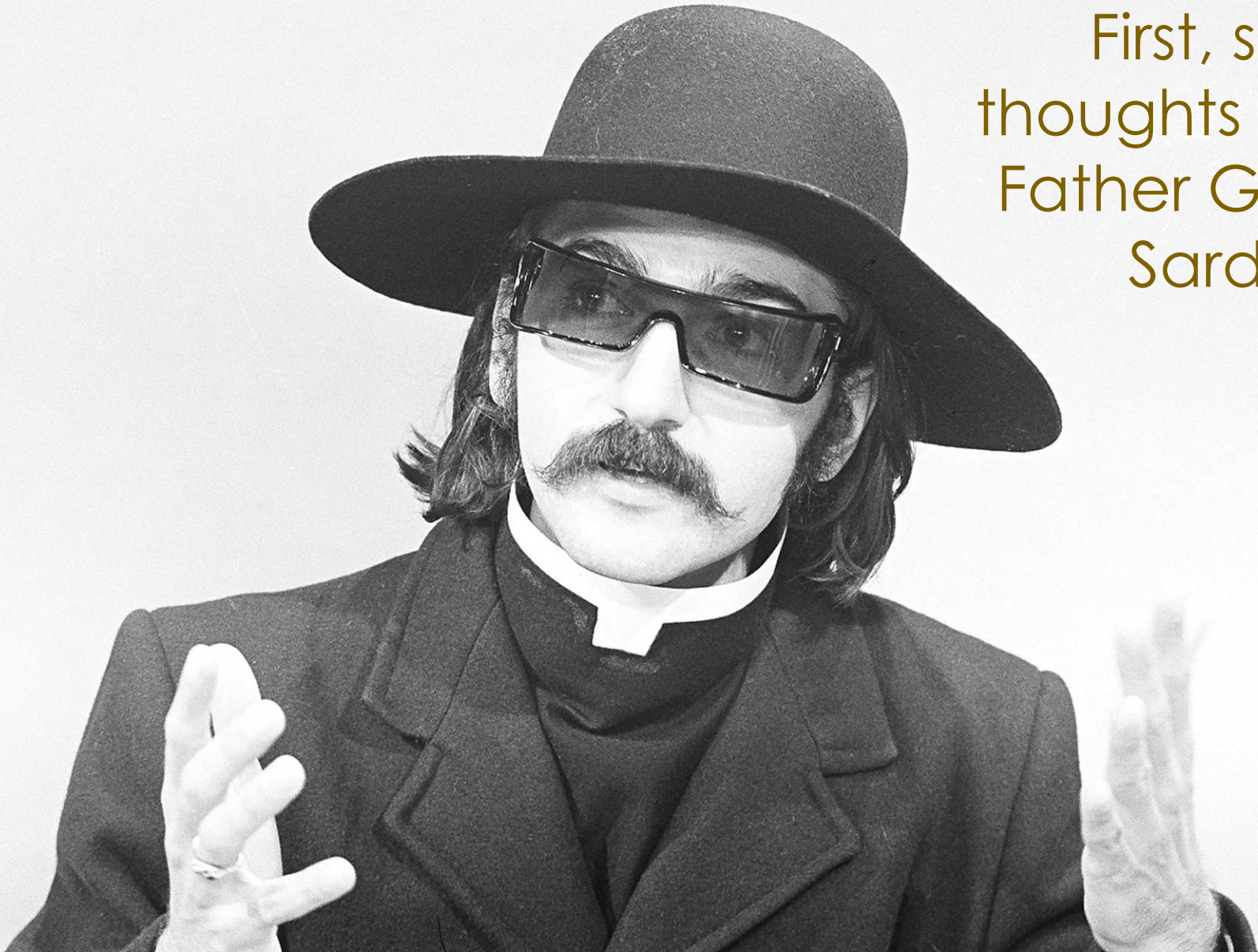
Quick diversion: Acknowledge the value of tests, and discuss the properties of well-written tests

Introduce some of the conceptual and practical challenges of high-stakes tests in higher-education, especially in remote learning

Illustrate some alternatives to exams, with a special focus on scaling “authentic assessments” in large courses

First, some  
thoughts from  
Father Guido  
Sarducci

[ clip ]

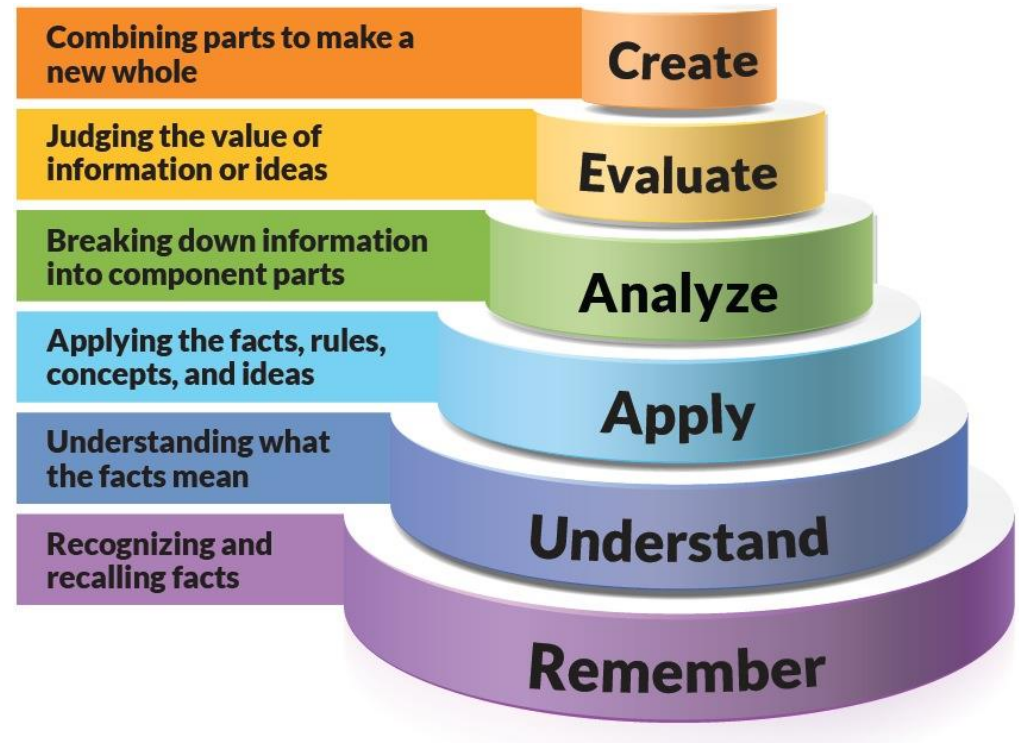




# Authentic assessment:

A task that requires application of what students have learned to a new situation, and that demands judgment to determine what information and skills are relevant and how they should be used.

# Bloom's taxonomy:

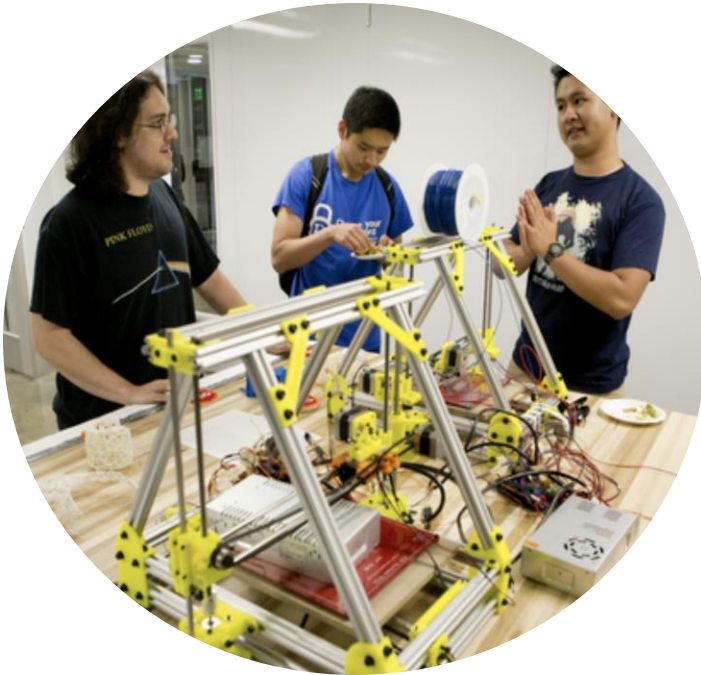


SOURCES: <https://citl.indiana.edu/teaching-resources/assessing-student-learning/authentic-assessment/index.html>  
<https://tips.uark.edu/using-blooms-taxonomy/>



# Authentic assessment:

Authentic assignments often focus on messy, complex real-world situations and their accompanying constraints. They are realistic.



- requires judgment and innovation.
- asks the student to “do” the subject.
- replicates the contexts in which adults are “tested” in the real world.
- assesses the student’s ability to efficiently and effectively use a repertoire of knowledge and skills to negotiate a complex task.
- allows opportunities to rehearse, practice, consult resources, and get feedback on and refine performances and products.

Tests introduce accountability.

They can motivate performance and improvement (Deci & Ryan, 2000).

Tests can be written with high levels of validity and reliability. They can be easy to administer, and often are efficient to grade.

Tests evaluate satisfaction of standards (e.g., Engineering education, medical education).

Tests often can introduce authenticity (application in simulated contexts).

# Why use tests in the college classroom?



Deci, E. L., & Ryan, R. M. (2000). The "what" and "why" of goal pursuits: Human needs and the self-determination of behavior. *Psychological inquiry*, 11(4), 227-268.

# What does a good test look like?



<https://cft.vanderbilt.edu/guides-sub-pages/test-enhanced-learning-using-retrieval-practice-to-help-students-learn/>

<https://cft.vanderbilt.edu/guides-sub-pages/writing-good-multiple-choice-test-questions/>

The test is valid. It represents and measures the learning outcomes of your course or module.

It is reliable. If you offer it to the same students over time, or to different large groups of students, the results will be mostly consistent.

It is objective. Equally competent learners will get similar results.

It motivates learners in a healthy way.

It has consequential relevance: the results for you and for students are worth the time and effort.

(It has some norms or values to compare against.)



Test or item validity must be judged against an external criterion.

Reliability can be evaluated by looking at test scores for the student population:

Item difficulty: What percentage of students got it right?

Item discrimination

Point Biserial Index

What is the relationship between the individual item scores and the learners' overall test scores?

Reliability coefficients (Cronbach  $\alpha$ ; KR-20):

Is there internal consistency among items on a test?

Standard error of measurement: How close is the actual test score to the learner's true (but unknowable) score?

## Consulting



SSRI offers one-on-one advice to support Duke students and faculty in planning and conducting their research and works to help faculty connect with collaborators and identify opportunities to secure funding to advance their research. Please [email us](#) to request an appointment.

Research originates with a great question or idea, yet requires much more—such as testable hypotheses, useful data, coded interviews, statistical or spatial software, and the ability to interpret and disseminate findings.

Getting from great idea to feasible project can require dodging pitfalls, particularly for the novice researcher or for collaborative teams that apply a new method or tool.

SSRI is here to help! Our Connection Bar offers a consulting service staffed by advanced graduate students, postdoctoral fellows, and SSRI research staff. This team is available by appointment to consult on all stages of the research process. Team members can offer quick advice on such topics as:

<https://ssri.duke.edu/education/consulting>

# Talk to an Expert

Duke Learning Innovation > Faculty Opportunities > Talk to an Expert

## Personalized Support

Learning Innovation provides Duke faculty personalized support for teaching and classroom innovation. We work one-on-one with you to try ideas, address challenges and create projects that improve your experience as an instructor and enhance student learning.

We work with faculty on:

- Course design
- Classroom activities
- Media production
- Selecting technology and software
- Assessing learning and project effectiveness
- Online education projects

<https://learninginnovation.duke.edu/faculty-opportunities/talk-to-an-expert/>

<https://flexteaching.li.duke.edu/fall-2020/course-design-planner/>

Download worksheets at link above or:

<https://flexteaching.li.duke.edu/files/2020/06/Course-Design-Plan-F20-Self-Service-Guide.docx>

## ASSESSMENTS, ACTIVITIES & GRADING SCHEME

How will you assess whether students have met the course learning outcomes? What products or materials will students produce to demonstrate their knowledge and skills? What course activities will you use to develop student knowledge and skills in phases with feedback and time for correction? Consider formative (ungraded) and summative (graded) activities and assessments.

Remember: in online or hybrid courses, more, smaller assessments are recommended, rather than only one or two high-stakes assignments.

[illegible]

# Vanderbilt Center for Teaching

<https://cft.vanderbilt.edu/guides-sub-pages/test-enhanced-learning-using-retrieval-practice-to-help-students-learn/>

The structure of the frequency and sequence of tests also matters for retrieval and application

repeated retrieval  
enhances long-term  
retention

various testing  
formats can enhance  
learning

feedback enhances  
the benefits of  
testing

Learning (via tests) is  
not limited to rote  
memory

retrieval practice  
over one set of  
material may help  
learning of later  
material

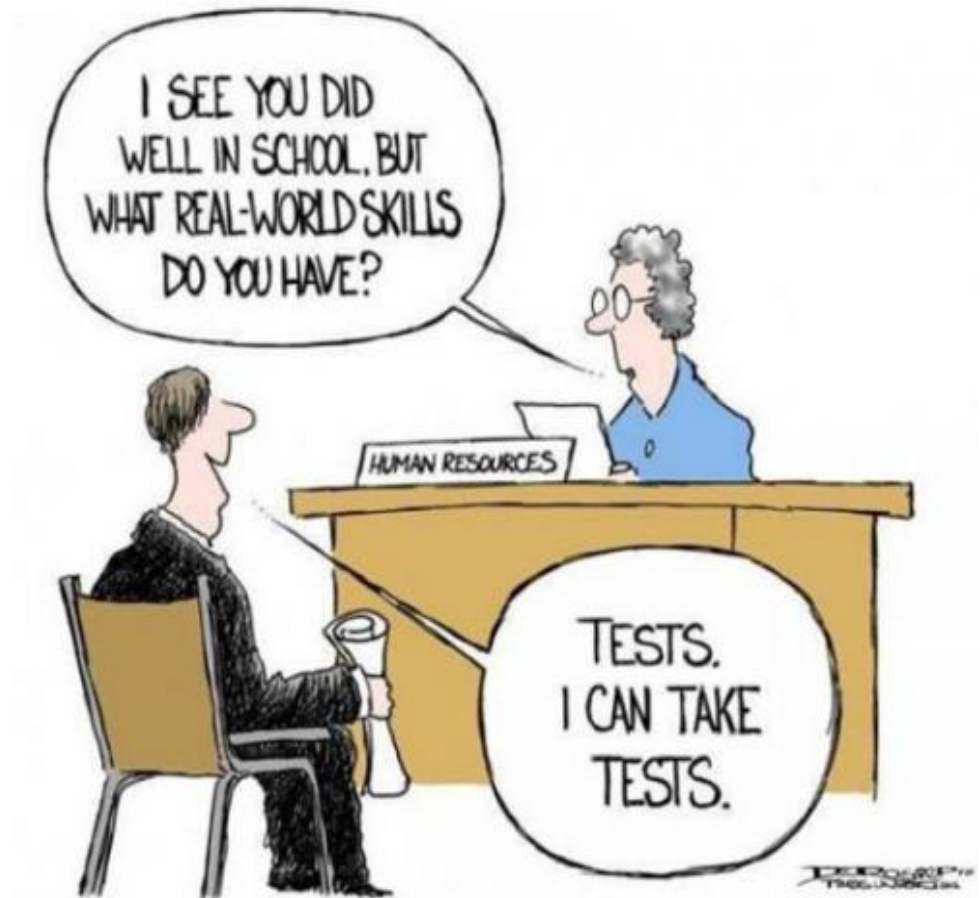
the benefits of  
(frequent) testing  
appear to extend to  
the classroom

# Hazards

A good test demands thoughtful learning outcomes first  
(e.g., application, evaluation, synthesis).

If thoughtful learning outcomes are missing or overlooked, the test cannot assess learning.

*This graduate's school taught him how to maneuver a test.*





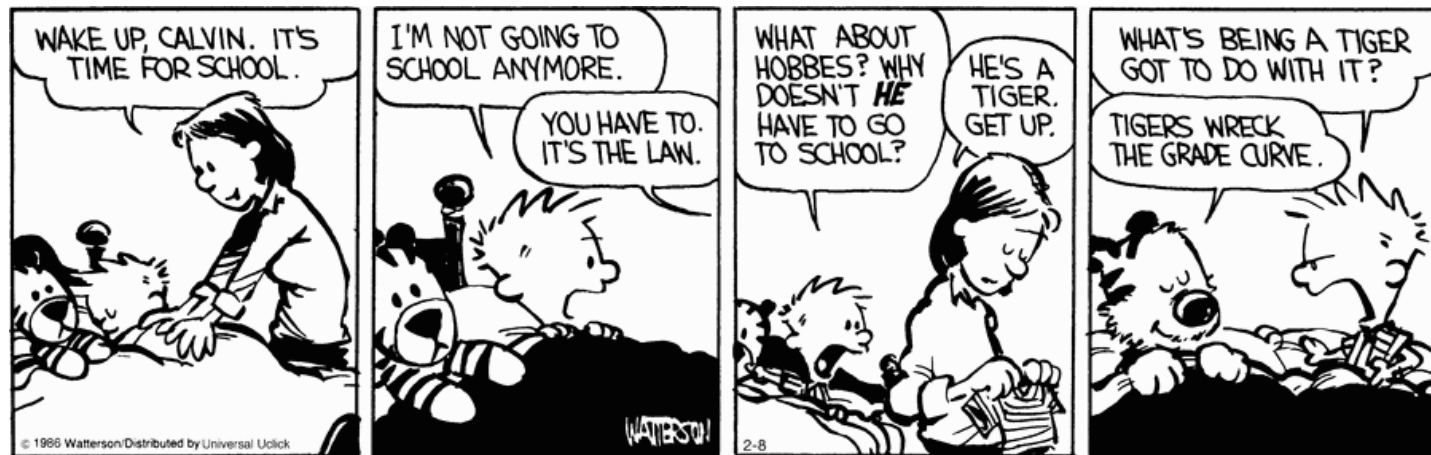
# Hazards

Reliance on grade curves puts students in competition with each other, rather than moving all students toward the learning outcomes.

*Student remarks from spring 2020 survey:*

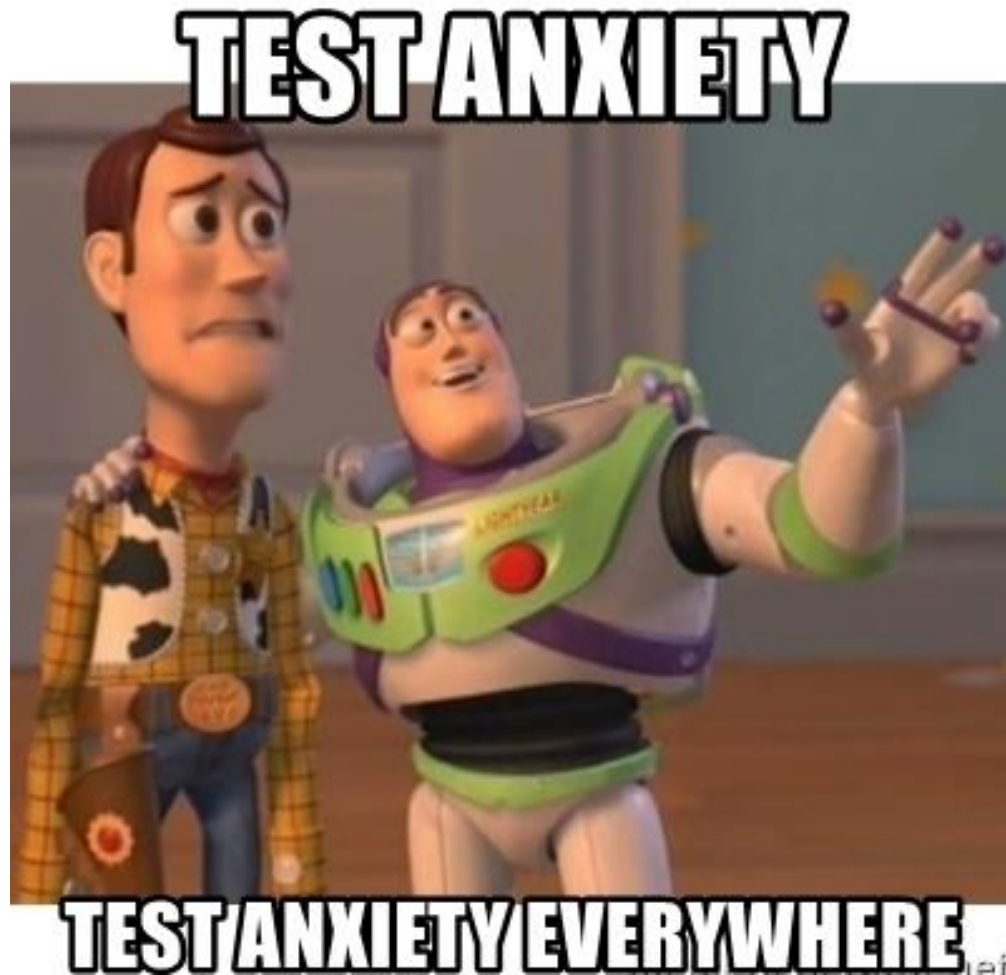
“Any curved class which requires any sort of non-essay assignment... students who were torn between abiding by the academic honesty standards and cheating to either get ahead or keep up with other work.”

“Closed-note exams with no enforcement mechanisms for honesty... reward cheaters and punish honest students, especially when the classes are graded on a curve.”



# Hazards

Because good tests take lots effort to create,  
we use them only 3-4 times during the course



# Hazards

High-stakes exams can induce anxiety and create a climate of desperation

*Student remarks on test anxiety from summer 2020:*

“Online timed exams pose unique challenges. For example, I live in a house with nine people, so during exams it was **hard to focus since I can't get any true peace and quiet.**”

“I found **the pressure of taking exams at home to be much more than taking exams in person**, especially when the professor is not understanding when it comes to timing or technology issues. I found myself being anxious where I normally wouldn't.”

“**The first exam was a terrible experience and was why I withdrew from the course.** We had so little time to complete the exam - although it wasn't insanely difficult, we all ran out of time mostly due to a virtual factor.”

“**Anxiety during timed exams made it difficult to focus**, and I fear that many students did not maintain academic integrity. “

# Hazards

Anxiety, desperation, and opportunity converge to produce academic dishonesty

*Student remarks on academic dishonesty from summer 2020:*

“[Professor] also reused old tests, which students were passing around.”

“I am very concerned about how many students are actually following the community standard... and class averages were somewhat high given the [very low] level of participation.”

“Cheating was out of control... I ended up earning several letter grades lower than what I would have attained in the normal semester due to higher grade boundaries.”

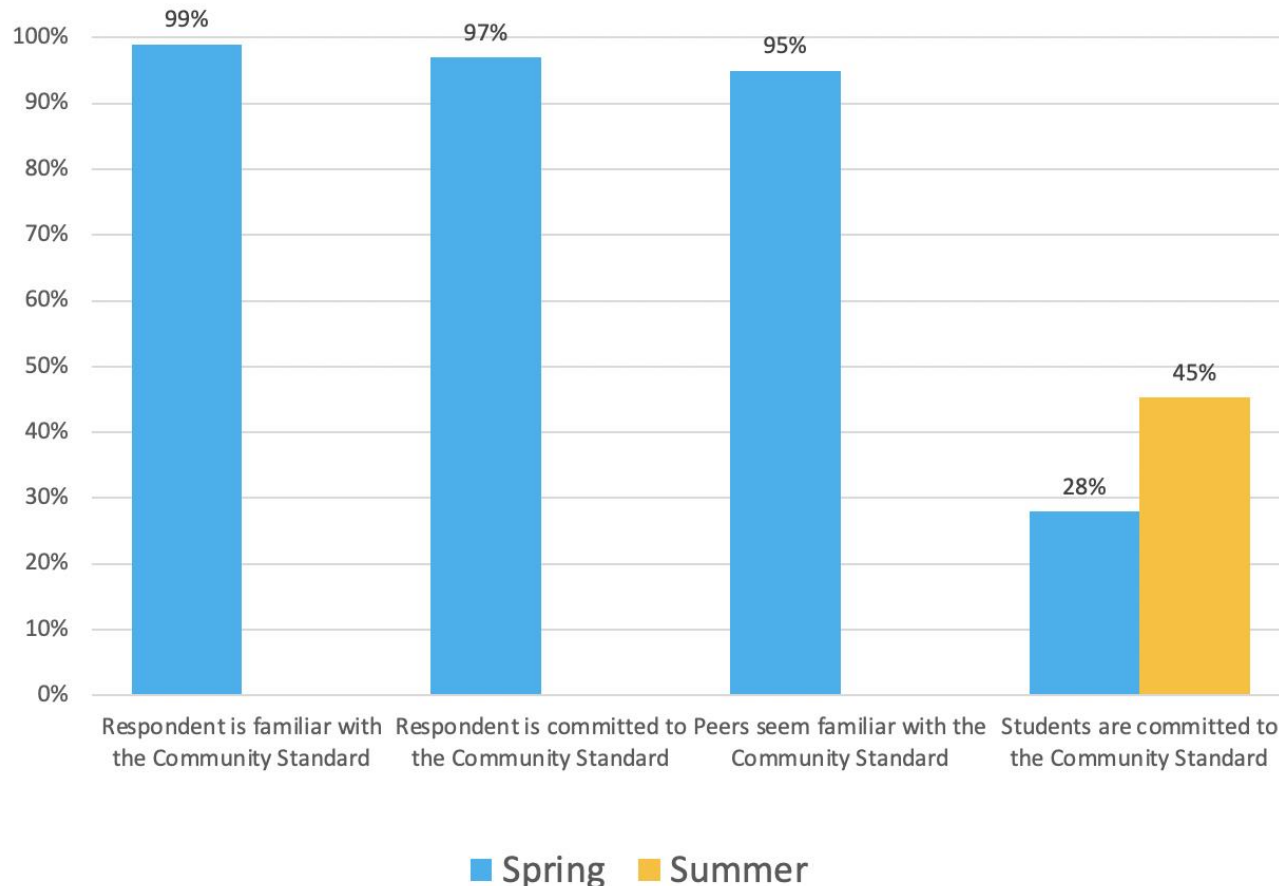
“[Professor] was not equipped with ways to prevent cheating. He made every exam "closed book" and I believe it lead to lots of cheating in the class. This gave a disadvantage to people who followed the community standard.”



# Hazards

Anxiety, desperation, and opportunity converge to produce academic dishonesty

*Student survey data from spring and summer 2020:*



# **Sustainable solutions to the problem of academic dishonesty are complex**

Clarifying, reinforcing course rules

Structuring the course well; providing many opportunities to practice and show mastery

Rethinking grade curves

Consideration of viable  
alternatives to high-stakes tests



# Alternatives?

Try Googling “authentic assessment” and see how many hits you get. A sample:

Teacher or preceptor observations

Essays or other writing

Interviews

Performances

Lab reports

Exhibitions and demonstrations

Portfolios

Journals

Simulations

Self- and peer-evaluation

Employer evaluations

# Alternatives?

Summer 2020: We asked students how effectively their graded assessment captured their learning.

	Ineffective	Only partially effective	Mostly effective	Fully effective
Essays	9	32	108	122
Exams & quizzes	33	115	229	118
Lab reports	22	39	83	56
Performances	12	21	58	35
Presentations	15	38	107	112
Computer programs or we..	10	20	76	68
Projects (general)	9	38	138	117
Projects (audio/visual)	8	23	99	76
Blogs, journals, reflections	10	31	81	98
Assessment other	2	7	3	11

← Exams were most frequent, but also rated the lowest

← Blogs, journals, and essays (above) were rated most effective

# There's an elephant in the room: Scalability





# **Seven ways to make authentic assessment more manageable**

# 1. Use scaffolded assessments

## STUDENT SCAFFOLD NEED 1

Must know how to look up sodium and potassium with their chemical symbols.

Sodium = Na

Chlorine = Cl

## STUDENT SCAFFOLD NEED 2

Must know how to figure out how many valence electrons each atom has.

Na has 1

Cl has 7

## STUDENT SCAFFOLD NEED 3

Must know how to draw an electron dot diagram for just the atoms.

Na •



## STUDENT SCAFFOLD NEED 4

Must know how to figure out whether an ionic or covalent bond will form, using electronegativity differences.

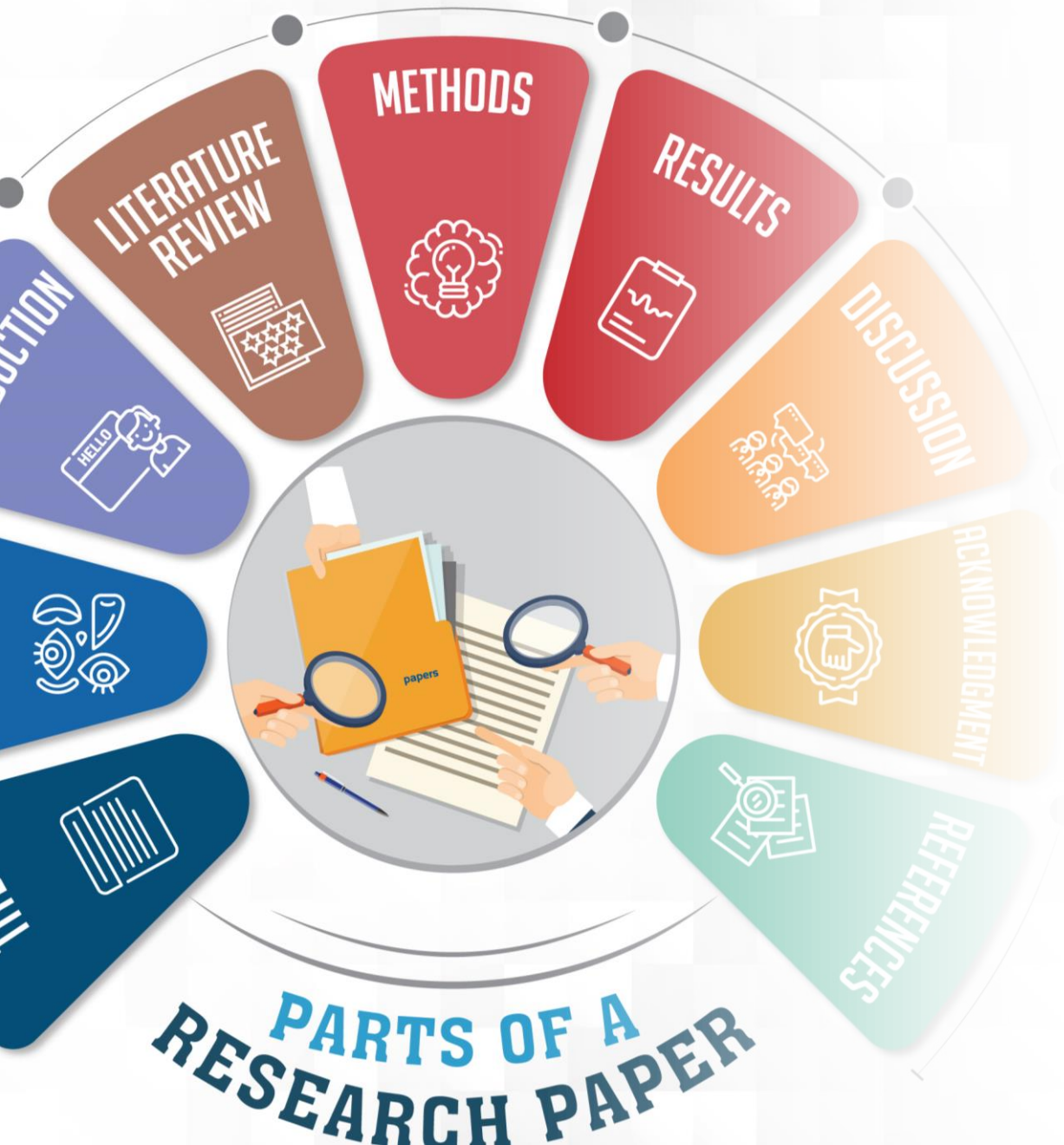
Na EN: .93  
Cl EN: 3.16

Dif: 2.23  
Bond: IONIC

## STUDENT SCAFFOLD NEED 5

Must know that an electron is traveling to Cl from Na and to show that with an arrow. (If being technical, single headed curved arrow)






## 2.

# Use mastery grading feedback structures



A group of people are gathered in a room, looking at several large posters displayed on a wall. The posters include a butterfly, a map of the United States, and a diagram of a person's body. The people are dressed in casual attire, and the room has a modern, clean aesthetic with recessed lighting. The text "3. Use peer feedback" is overlaid on the bottom left of the image.

**3.**

**Use peer feedback**

# 4.

## Use rubrics

### Writing Rubric

Course: \_\_\_\_\_

Student: \_\_\_\_\_

Unacceptable	Acceptable	Target	Excellent
Does not develop ideas clearly, uneven and lacks overall organization, unclear introduction or conclusion	Develops and organizes ideas in paragraphs that are not necessarily connected. Some overall organization, but some ideas seem illogical and/or unrelated, unfocused introduction or conclusions	Develops unified and coherent ideas within paragraphs with generally adequate transitions; clear overall organization relating most ideas together, good introduction and conclusion.	Develops ideas clearly, organizes them into paragraphs and paragraphs with effective transitions. Clear and specific introduction and conclusion
Contains words that are not sentence fragments, inadequate sentence structure, errors are distracting	Word forms and sentence structures are adequate to convey basic meaning. Errors cause noticeable distraction	Word forms are correct, sentence structure is effective. Presence of a few errors is not distracting.	Employs word forms correctly, develops concise sentences, English sentence structure, variety of sentence structures effectively.
Contains frequent errors in spelling and grammar that interfere with readability	Frequent errors in spelling and grammar distract the reader	While there may be minor errors, the writing follows normal conventions of spelling and grammar throughout and has been carefully proofread	The writing is free in terms of spelling and grammar
Contains errors in organization, confusion, personal and evidence, logical flawed	Presents ideas in general terms, support for ideas is inconsistent, some distinctions need clarification, reasoning unclear	Supports most ideas with effective examples, references, and details, makes key distinctions	Explores ideas thoroughly, supports points with evidence, balance of subjective and objective evidence, effectively makes distinctions
Purpose and focus of writing are not clear to the reader	The writer's decisions about focus, organization, style, and content sometimes interfere with the purpose of the writing.	The writer has made good decisions about focus, organization, style, and content so as to achieve the purpose of the writing.	The writer's decisions about focus, organization, style, and content fully achieve the purpose and keep the reader at the center of the writing.

# Speaking of rubrics,

---

**Picking the right rubric type for your assessment will save a lot of time when grading**

# Breakfast in Bed: Holistic Rubric

Score	Description
4	All food is perfectly cooked, presentation surpasses expectations, and recipient is kept exceptionally comfortable throughout the meal.
3	Food is cooked correctly, the meal is presented in a clean and well-organized manner, and the recipient is kept comfortable throughout the meal.
2	Some food is cooked poorly, some aspects of presentation are sloppy or unclean, or the recipient is uncomfortable at times.
1	Most of the food is cooked poorly, the presentation is sloppy or unclean, and the recipient is uncomfortable most of the time.

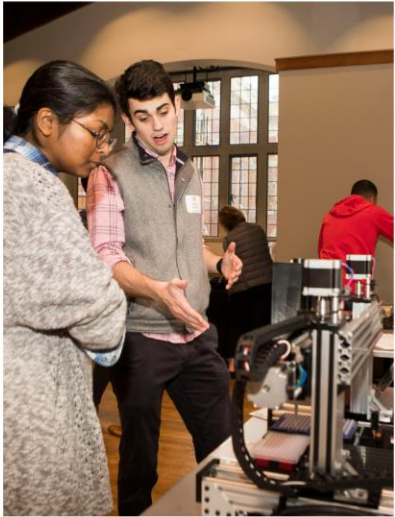
# Breakfast in Bed: Analytic Rubric

	Beginning 1	Developing 2	Accomplished 3	Exemplary 4	Score
<b>Food</b>	Most food is colder or warmer than it should be, is under- or over-seasoned, or is under- or overcooked.	Some food is colder or warmer than it should be, is under- or over-seasoned, or is under- or overcooked.	All food is at the correct temperature, adequately seasoned, and cooked to the eater's preference.	All food is perfectly cooked and seasoned to the eater's preference. Additional condiments are offered.	
<b>Presentation</b>	More than one item (tray, napkin, or silverware) are dirty or missing.	Tray, napkin or silverware may be dirty or missing.	Food is served on a clean tray, with napkin and silverware. Some decorative additions may be present.	Food is served on a clean tray, with napkin and silverware. Several decorative touches are added.	
<b>Comfort</b>	Wake-up is abrupt, little to no help with seating, and the recipient is rushed and crowded during the meal.	Wake-up is somewhat abrupt, recipient may struggle with seat adjustment, or there may be some rushing or crowding during eating.	Recipient is woken gently, assisted in seat adjustment, and given reasonable time and space to eat.	Recipient is woken gently and lovingly, assisted until seating is just right, and given abundant time and space to eat.	

## Breakfast in Bed: Single-Point Rubric

Concerns <i>Areas that Need Work</i>	Criteria <i>Standards for This Performance</i>	Advanced <i>Evidence of Exceeding Standards</i>
	<b>Food:</b> All food is at the correct temperature, adequately seasoned, and cooked to the eater's preference.	
	<b>Presentation:</b> Food is served on a clean tray, with napkin and silverware. Some decorative additions may be present.	
	<b>Comfort:</b> Recipient is woken gently, assisted in seat adjustment, and given reasonable time and space to eat.	





5.

**Let students design their  
assessment pathways**





# 6. Take advantage of technology

**Ready to get started?**

**[learninginnovation@duke.edu](mailto:learninginnovation@duke.edu)**

---