

# So you've got some data... now what? Evidence to draw Valid conclusions

AN EGAD PROJECT WORKSHOP

Brian Frank and Peter Ostafichuk  
CEEA 2021, June 20, 2021 @ 2:00 – 3:20 pm ADT



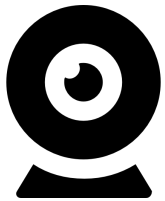
# Housekeeping details



Please mute your mic in main room (unless speaking)



CEEA is recording these sessions (but not breakouts)



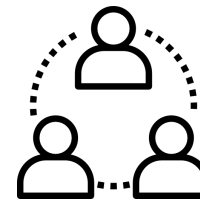
Please keep your video on (unless issues with bandwidth)



Feel free to get up to stretch or get a drink as you need



Feel free to use the chat to ask questions



This is an open, inclusive, and respectful space



Supporting Canadian engineering programs in the development of effective continuous program improvement practices

[HOME](#)

[ABOUT EGAD](#)

[BIG PICTURE](#)

[6 STEP GUIDE](#) ⌵

[RESOURCES](#) ⌵

[CONTACT](#)



# About EGAD

Who we are

What we do

## GACIP Presentations

Show 5 entries

Year	Title	Downloads
2018	GAPNet Presentation	Slides Video
2018	CEAB Update	Slides Video
2018	Aggregating Data	Slides Video
2018	EGAD Morning Workshop 1	Video
2018	EGAD Morning Workshop 2	Video

Showing 1 to 5 of 25 entries

Previous Next

## Laval

Show 10 entries

Search:

Year	Title	Downloads
2018	Using data to influence quality and make decisions	Slides Handout

Resources will be posted on EGAD Resources page



**2 Mapping**  
the Curriculum

**3 Collecting**  
the Data

**1 Defining**  
Program Objectives and Indicators

**4 Analysing**  
and Interpreting the Data

**6 Managing**  
and Implementing Change

**5 Improving**  
Curriculum and Processes

# Goals

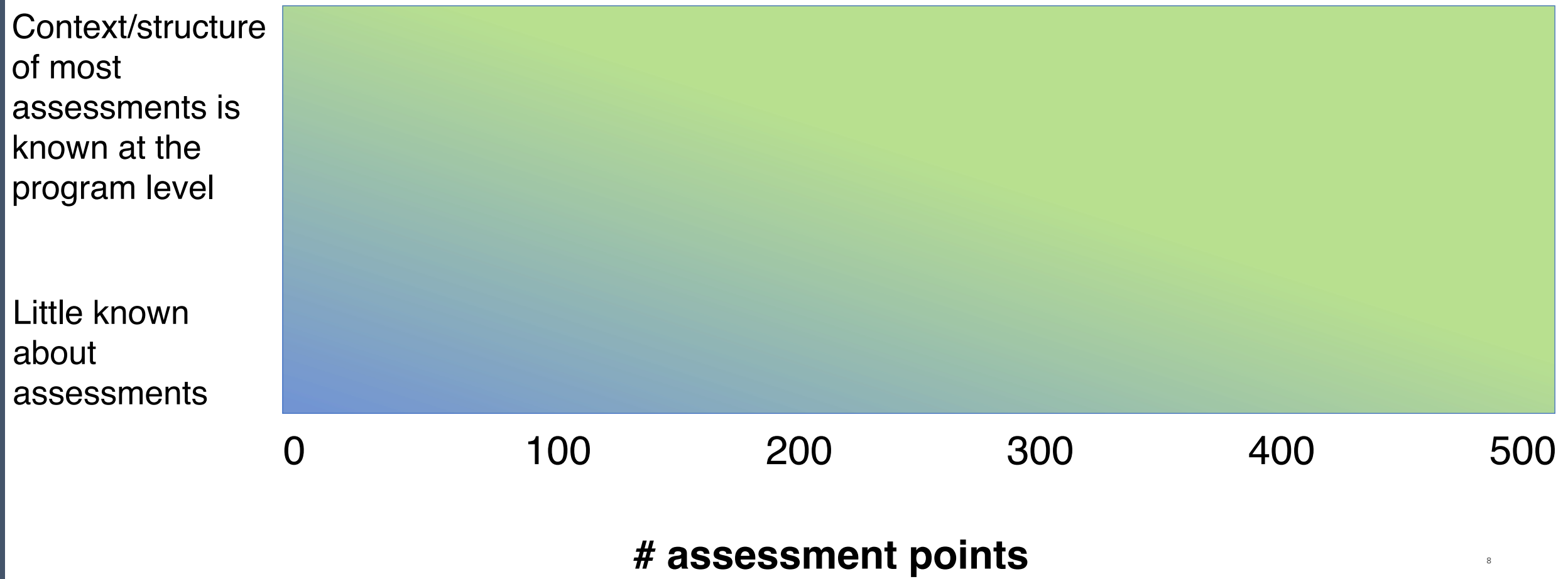
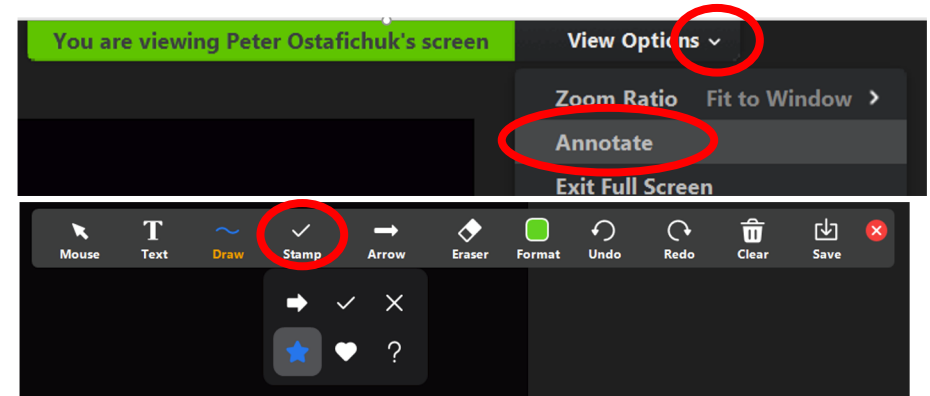
1. Use terminology to enable discussion about drawing valid conclusions from programmatic assessment.
2. Evaluate validity of conclusions drawn from an assessment measure.
3. Identify how multiple assessment measures can lead to meaningful conclusions

# Validity

“the degree to which the evidence obtained through validation supports the score interpretations and uses . . . . from a certain test administered to a certain person or group on a specific occasion”

(Standards for Educational and Psychological assessment, AERA, APA, & NCME, 2014)

Use the Zoom stamp tool: Place a stamp on the rectangle below reflecting your *overall* impression about the assessment measures in your program:





## Validity

the degree to which the evidence and theory supports the interpretations and use of assessment data

## Reliability

consistency of scores across multiple measures

# Valid conclusions require **reliable** data.

**Reliability (*consistency*) can be measured as:**

- Consistency over time,  
i.e. test-retest reliability
- Consistency between graders,  
i.e. inter-rater reliability
- Internal consistency,  
i.e. inter-item reliability

**Validity of conclusions depends on:**

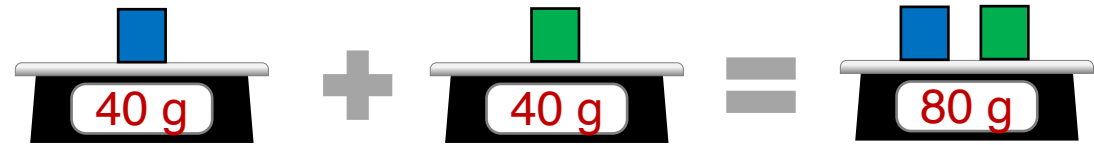
- Measuring the right things
- Using appropriate approaches to measure
- Agreement with conclusions drawn from other approaches

# An analogy to test understanding...

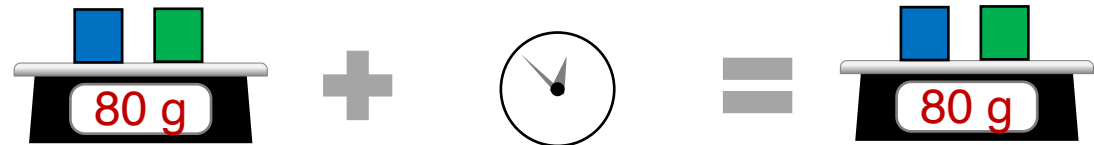
Inter-rater reliability:



Inter-item reliability:



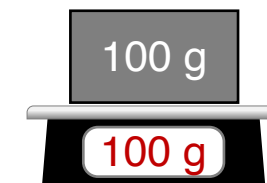
Test-retest reliability:



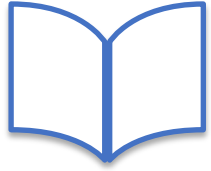
Inter-item reliability:



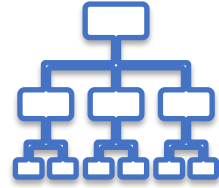
Validity:



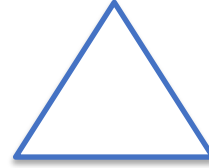
# Evidence we can use to evaluate Validity



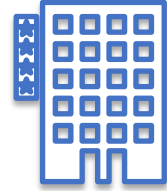
**Content** of the  
assessment measure



**Internal structure** of  
assessment measure

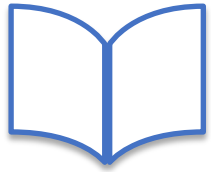


**Triangulation** with  
other measures



**Consequences** of  
**results**

# Consider evidence to evaluate Validity



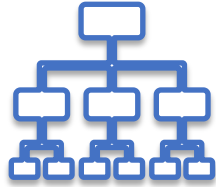
## **Content** of the assessment measure

Are questions/wording appropriate to intended purpose?

Would multiple experts agree with the alignment?

Would multiple scorers agree with scoring?

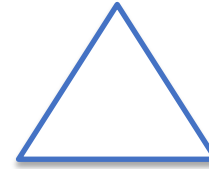
Do scores reflect other factors (lack of time, vocabulary)?



## **Internal structure** of assessment measure

Do multiple items converge/diverge as expected?

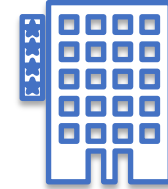
Do students with higher overall performance get tougher questions right more frequently?



## **Triangulation** with other measures

Does the evidence align/correlate with related measures?

Does the evidence diverge from unrelated measures?



## **Consequences** of results

Is the use of the assessment measure appropriate for decisions made from it?

Does the evidence from assessment relate to future intended outcomes (employer comments, alumni perspectives)?

# Task 1: Scenario (15 min) <http://bit.ly/EGAD-CEEA-2021>

**Use the framework to review the process below and identify what (a) you can use the data for, and (b) what you can't make conclusions about.**

**Scenario:** Your group is the committee charged with reviewing your department's continuous improvement process in a program with 200 students/year. You're currently examining the way your program measures student performance on the following program indicator:

*Develop and evaluate mathematical models to support solving complex engineering problems.*

Last year the evidence provided to your committee consisted of a report showing the following:

- Grades measured on two questions each from final exam questions in three courses: first year calculus, second year electronics, and third year electromagnetics. Half the class fell below expectations on one question, but everyone met expectations on the other; the other two courses showed the majority of the class meeting expectations.
- A self-rating survey of students indicated that overwhelmingly they felt they demonstrated that indicator, along with most other attributes aside from ethics & equity and lifelong learning
- Feedback from an industry focus group mentioned that students have strong mathematical skills but often struggle to know how to apply it

# Task 1

<http://bit.ly/EGAD-CEEA-2021>

## Part 1: Instructions & Planning

### Part 1: Instructions and Individual thought

Use the framework to review the Scenario below and identify  
(a) What evidence do you have to make recommendations & conclusions, and  
(b) aspects about the evidence that is weak

After you've read the scenario, get used to putting Sticky notes on the Scratchpad (next page over), and start writing down your own thoughts.

Scenario: Your group is the committee charged with reviewing your department's continuous improvement process in a program with 200 students/year. You're currently examining the way your program measures student performance on the following program indicator:

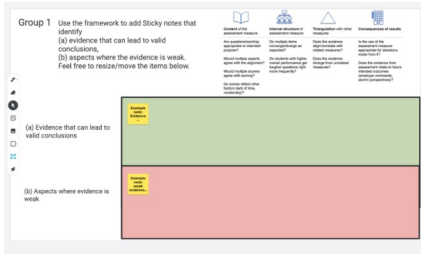
Develop and evaluate mathematical models to support solving complex engineering problems.

Last year the evidence provided to your committee consisted of a report showing the following:

1. Grades measured on two questions each from final exam questions in three courses: first year calculus, second year electronics, and third year electromagnetics. Half the class fell below expectations on one question, but everyone met expectations on the other; the other two courses showed the majority of the class meeting expectations.
2. A self-rating survey of students indicated that overwhelmingly they felt they demonstrated that indicator, along with most other indicators aside from ethics & equity, and lifelong learning.
3. Feedback from an industry focus group mentioned that recently graduated students have strong mathematical skills but often struggle to know how to apply it

### Part 2: Group work

Switch to your group's page and use the Sticky notes to respond to (a) and (b).



## Part 2: Group

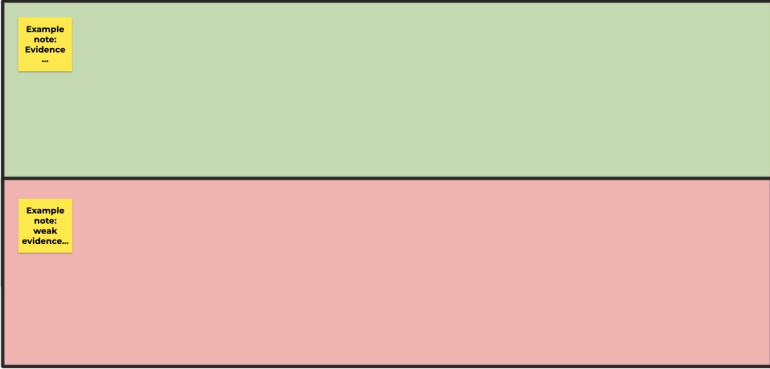
### Group 1

Use the framework to add Sticky notes that identify  
(a) evidence that can lead to valid conclusions,  
(b) aspects where the evidence is weak.  
Feel free to resize/move the items below.

Content of the assessment measure	Internal structure of assessment measure	Triangulation with other measures	Consequences of results
Are questions/wording appropriate to intended purpose?	Do multiple items converge/diverge as expected?	Does the evidence align/relate with related measures?	Is the use of the assessment measure appropriate for decisions made from it?
Would multiple experts agree with the alignment?	Do students with higher overall performance get tougher questions right more frequently?	Does the evidence diverge from unrelated measures?	Does the evidence from assessment relate to future intended outcomes (employer comments, alumni perspectives)?
Would multiple scorers agree with scoring?			Do scores reflect other factors (lack of time, vocabulary)?

(a) Evidence that can lead to valid conclusions

(b) Aspects where evidence is weak



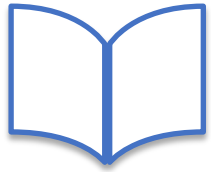
# What can we make of the data?

**(a) Data can be used for:**

**(b) Can't draw conclusions about:**



# Scenario followup



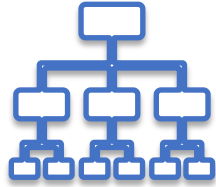
## **Content** of the assessment measure

Are questions/wording appropriate to intended purpose?

Would multiple experts agree with the alignment?

Would multiple scorers agree with scoring?

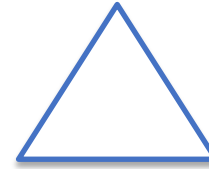
Do scores reflect other factors (lack of time, vocabulary)?



## **Internal structure** of assessment measure

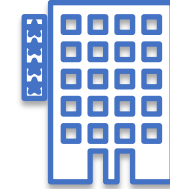
Do multiple items converge/diverge as expected?

Do students with higher overall performance get tougher questions right more frequently?



## **Triangulation** with other measures

Does the evidence align with other measures?



## **Consequences of results**

Is the use of the assessment measure appropriate for decisions made from it?

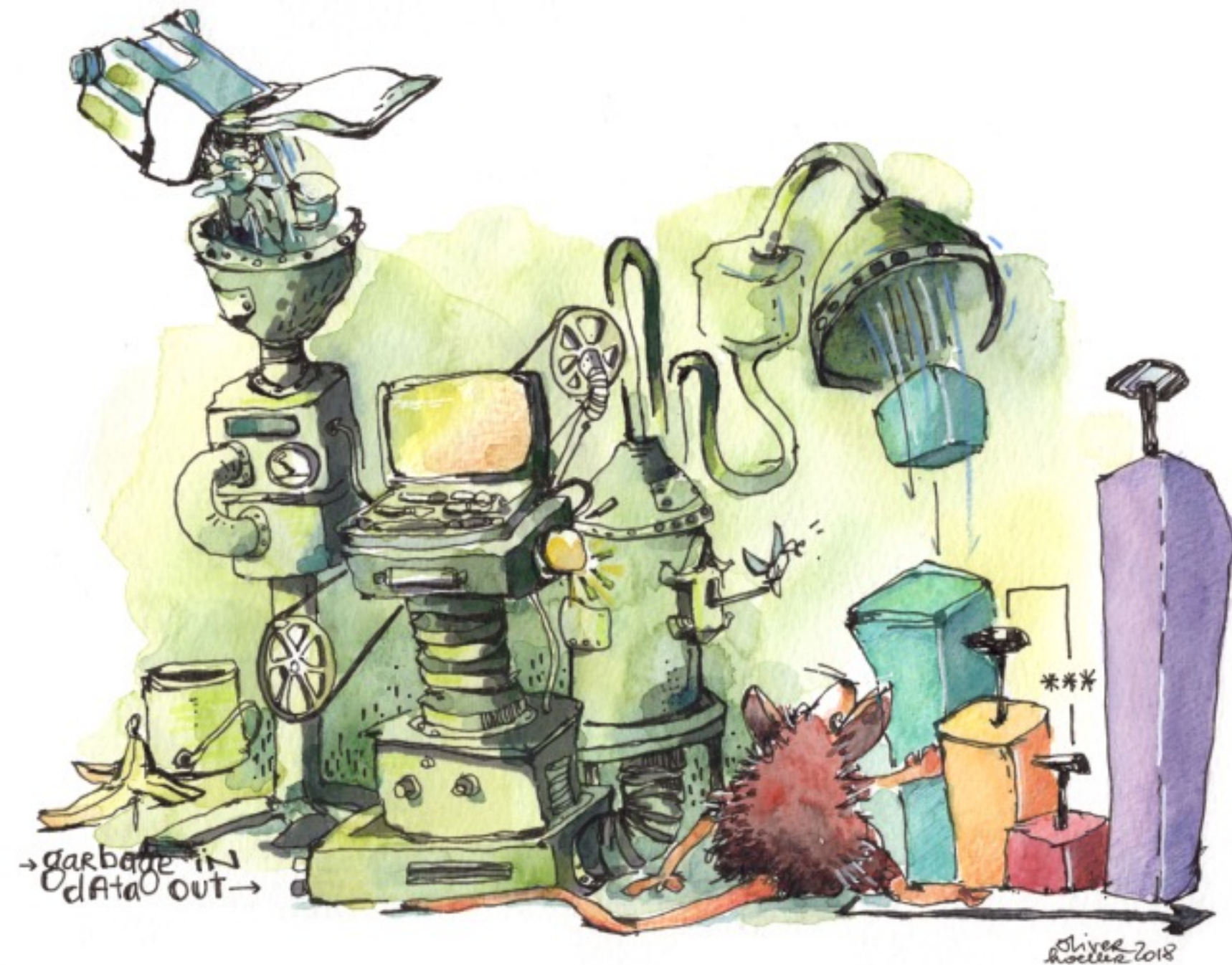
Does the evidence from assessment relate to future intended outcomes (employer comments, alumni perspectives)?

# Common assessment situation

	KB	PA	IN
Course 1	X	X	
Course 2		X	
Course 3	X		X
Course 4		X	
Course 5		X	X
Course 6	X	X	
Course 7			
Course 8	X		X

1. Large number of assessments/attribute
2. Nature and alignment of each one is unknown when drawing conclusions.

12 GA X 3 ind/GA X 2 measures/yr X 4 yrs = 288 assessment measures



“It is not possible to carry out meaningful statistical analysis of data that is fundamentally inaccurate.”

Oliver  
Hoeller 2018

# High Volume vs. High Quality Assessments

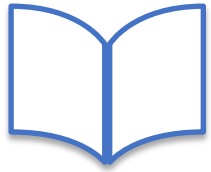
## High Assessment Volume

	KB	PA	IN
Course 1	X	X	X
Course 2	X	X	
Course 3	X		X
Course 4		X	
Course 5		X	X
Course 6	X	X	
Course 7	X		X

## High Assessment Quality

	KB	PA	IN
Course 1	Signature exam question	Signature exam question	
Course 2			
Course 3			Signature lab report
Course 4			
Course 5			
Course 6	Signature design report	Signature design report	
Course 7			Signature lab report

# Scenario followup



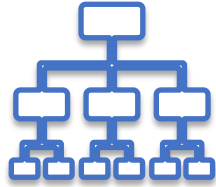
## **Content** of the assessment measure

Are questions/wording appropriate to intended purpose?

Would multiple experts agree with the alignment?

Would multiple scorers agree with scoring?

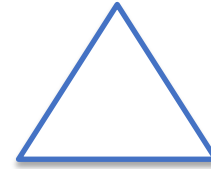
Do scores reflect other factors (lack of time, vocabulary)?



## **Internal structure** of assessment measure

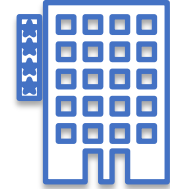
Do multiple items converge/diverge as expected?

Do students with higher overall performance get tougher questions right more frequently?



## **Triangulation** with other measures

Does the evidence align with other measures?



## **Consequences of results**

Is the use of the assessment measure appropriate for decisions made from it?

Does the evidence from assessment relate to future intended outcomes (employer comments, alumni perspectives)?

# Task 2: Scenario

(20 minutes) Use the Evidence framework to recommend how to assess the following indicator such that you can have confidence in conclusions:

Develop and evaluate mathematical models to support solving complex engineering problems.

Click to add text

Your program has about 40 indicators, and the department has asked you to use this indicator as an example of how to assess other similar indicators.

Feel free to use your Jamboard. Be prepared to give a summary afterwards.

# Task 2: Scenario debrief

(20 minutes) Use the Evidence framework to recommend how to assess the following indicator such that you can have confidence in conclusions:

Develop and evaluate mathematical models to support solving complex engineering problems.

Your program has about 40 indicators, and the department has asked you to use this indicator as an example of how to assess other similar indicators.



# Recommendations

1. Maintain richness of data for program-level decisions
2. Triangulate
3. Involve multiple stakeholder groups in assessment and interpretation
4. Use small assessments for student learning, and to inform course-level improvements; more significant well-crafted assignments for program-level data

# Resources

- Kerrie Anna Douglas, Şenay Purzer (2015), “Validity: Meaning and Relevancy in Assessment for Engineering Education Research”, JEE, <https://doi.org/10.1002/jee.20070>
- Standards for Educational and Psychological Testing, AERA/APA/NCME
- J. Pierce et al. (2019), “When I say.... programmatic assessment in postgraduate medical education”, <https://onlinelibrary.wiley.com/doi/epdf/10.1111/medu.13949>

# So you've got some data... now what? Evidence to draw Valid conclusions

AN EGAD PROJECT WORKSHOP

Brian Frank and Peter Ostafichuk  
CEEA 2021, June 20, 2021 @ 2:00 – 3:20 pm EDT

