Building Evidence to Draw Valid Conclusions

EGAD Workshop December 2019
Brian Frank
with Alex Downie, Peter Ostafichuk

EGAD

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

About EGAD

Who we are	What we do

HOME ABOUT EGAD BIG PICTURE 6 STEP GUIDE ⌄ RESOURCES ⌄ CONTACT 🔍

Institutional Examples

Workshop Materials

Publications

Glossary

Accreditation Resources

📁 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

3.1.5 Assessment results: At least one set of assessment results must be obtained for all twelve attributes over a cycle of six years or less. The **results should provide clear evidence that graduates of a program possess the above list of attributes.**

3.2.1 Improvement process: There must be processes in place that demonstrate that program outcomes are being assessed in the context of the graduate attributes, and that the **results are validated, analyzed** and applied to the further development of the program.

Last year's focus

Comparing approaches to reliability and aggregating quantitative data

From: Task-level scores assigned to **individual students** (which most of us are doing) ➡ To: **Providing data that** demonstrates that a **student cohort** possesses a graduate attribute

Workshop 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

Illustrate approach to respond to recent move to highlight limited (high quality) assessment examples

1. Knowledge Base
• Demonstrate competence in university-level mathematics.
• Interpret natural phenomena and relationships through the use of analytical and/or experimental techniques.
Apply mathematics, natural science and engineering science to engineering problems.

2. Problem Analysis
• Characterize complex engineering problems.
• Develop appropriate frameworks for solving complex engineering problems.
• Implement solutions for complex engineering problems.
• Analyze solutions to complex engineering problems to draw conclusions.

3. Investigation
• Conduct investigations to test hypotheses related to complex problems.
• Analyze and interpret data using appropriate techniques and tools.
• Synthesize information from investigations, considering sources of uncertainty and limitations to reach substantiated conclusions.
• Adhere to appropriate workplace safety protocols in all work environments.

4. Design
• Define problem, objectives and constraints.
• Compare multiple strategies for solving a problem.
• Create a product, process or system to solve a problem, that meets specified needs, and subject to appropriate iterations.
• Evaluate performance of a design, using criteria that incorporates specifications, limitations, assumptions, constraints, and other relevant factors.

5. Use of Engineering Tools
• Develop, adapt and/or extend appropriate software, equipment, models, and simulations for a range of engineering activities.
• Apply and manage appropriate techniques, apparatus, databases, models, tools, and/or processes to accomplish a task.
• Evaluate limitations and errors of instrumentation/measurement techniques/models/simulations to assess appropriateness of the results.

6. Individual & Team Work
• Take initiative to plan and complete tasks, as an individual and team member, in order to meet goals.
• Seek and integrate diverse and alternative perspectives in decision-making.
• Lead by example by assuming responsibility for some aspect of the project, seeking input from all team members.

7. Communication Skills
• Produce clear, concise, precise and well-organized written communication with language appropriate for the audience.
• Deliver formal and informal oral presentations with suitable language, content, style, timing and flow, while adapting format, content and tone to audience and purpose.
• Create figures, maps, tables and drawings to engineering report standards.
• Interpret communication from a variety of sources and respond to instructions and questions, displaying full understanding of the topic.
• Generate a traceable and defensible record of a technical project using an appropriate records system.

8. Professionalism
• Recognize that engineering is a regulated profession

9. Impact of Engineering on Society & the Environment
• Analyze the environmental impact of engineering activities, and incorporate sustainability into decision making.
• Take appropriate action to mitigate risks associated with economic, health, safety and legal aspects of engineering activities.
• Evaluate cultural, societal, and technical norms while maintaining ethical and moral position required for engineering practice in Canada.

10. Ethics & Equity
• Identify and resolve potential ethical issues using ethical principles and codes, demonstrating knowledge of professional accountability in engineering.
• Intentionally incorporate principles of fairness, access and opportunity into decision-making.
• Adhere to guidelines regarding the fair use of intellectual property and contract guidelines.

11. Economics & Project Management
• Apply economic considerations, such as capital, operating, societal and life cycle costs, to design processes.
• Effectively plan projects, including mitigating risk and managing change, to complete project on-time and on-budget.
• Evaluate and reflect on own knowledge, skills and learning.

12. Life-long Learning
• Independently acquire new knowledge and skills for ongoing personal and professional development.
• Identify, organize, and critically evaluate information from an appropriate range of sources, to meet learning needs.

Deliberate Development vs. Assessment

Deliberate Development

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

Deliberate Assessment

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

Part I: What is Validity?

Task 1: In your groups, create a response to:

What does **validity** mean?

How does **validity** differ from **reliability**?

Can a test be **valid**?

Reliability

consistency of scores
across multiple measures

Validity

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

Valid conclusions require **reliable** data.

Reliability of data relies on *consistency*, which 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 from an assessment measure depends on:

- Measuring the right things

- Using appropriate approaches to measure

- Agreement with conclusions drawn from other approaches

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)

What validity is not...

- there is no such thing as a valid assessment, only valid uses

- there are not multiple types of validity, only multiple types of evidence to support valid uses of assessment

Followup from Task 1:

Within your group:

Is everyone clear about the following? Any questions/differences in opinion?

What does **validity** mean?

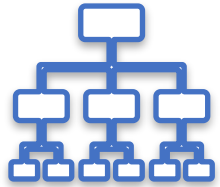How does **validity** differ from **reliability**?

Can a test be **valid**?

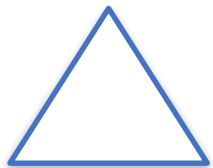# Part II: What evidence supports validity of conclusions?
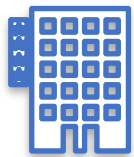
Four kinds of evidence

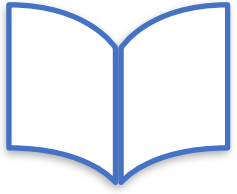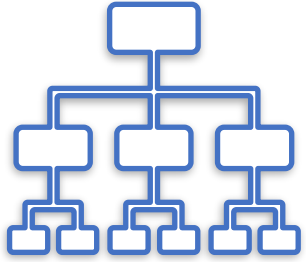|	
	Evidence about content of assessment measure
	Evidence about internal structure of assessment measure
	Evidence about relation to other assessment measures (triangulation)
	Evidence about consequences of use

1. Evidence about content ("content")

Content of the assessment aligns with the intended purpose

- Does assessment measure align with intended purpose, and only intended purpose?

- Would other experts agree the assessment measure aligns with the purpose?

E.g. student performance on written communications is related to their ability to reason and organize thoughts, not on their familiarity with technical terms they haven't seen before.

2. Evidence about internal structure ("structure")

If there are multiple items related to an outcome, do they converge/diverge as expected?

E.g. if a test is a good measure of conceptual knowledge of thermodynamics, scores on questions should be correlated to each other more than to other kinds of questions
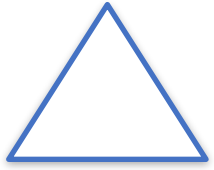
3. Evidence about relation to other assessment measures ("triangulation")

Do the results align with other measures?

E.g. Do measures of design process ability align with comments from capstone design course supervisor and observations from co-op employers?
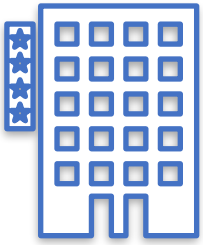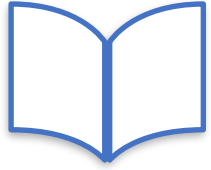
4. Evidence about consequences of assessment data ("consequences")

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)?

E.g. Do measures of design process ability align with comments from employment supervisors and alumni feedback?
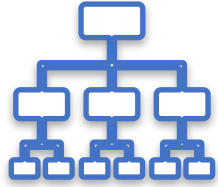
Evidence to consider related to Validity

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 the outcome, rather than other factors (student effort, 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?
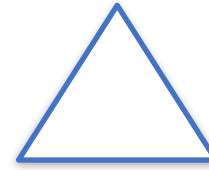
Triangulation with other measures

Does the evidence align with other measures?
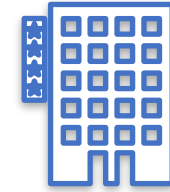
Consequences of results

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: Use the four types of evidence in the framework below to evaluate how valid conclusions could be from at least two **assessments** familiar to members of your group (i.e. from within a course or program).

1. Briefly describe the assessment measure.
2. Briefly evaluate how the four elements below impact conclusions that can be drawn from that assessment measure.

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 with other 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 from assessment relate to future intended outcomes (employer comments, alumni perspectives)?
Would multiple scorers agree with scoring?			
Do scores reflect the outcome, rather than other factors (student effort, lack of time, vocabulary)?			

Part III: Programmatic assessments

	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

Typical situation:

Large number of assessments/attribute

Unsure about validity of conclusions drawn from each one

Task 3: As a group:

1. (1 minute) Identify a rough "indicator" you'd like to draw conclusions about within a fictitious program (e.g. written communication, problem definition, ability to evaluate information, etc.).

2. (1 minute) identify one assessment measure you would use as a key source of evidence relative to that indicator

3. (15 minutes) Based on the 4 types of evidence used to establish validity of assessment measures, consider what activities would need to happen to have confidence in conclusions drawn from that measure

4. (5 minutes) How will the number of quality assessments measures developed per indicator impact ability to draw conclusions?

High Volume vs. Signature Assessments

High Volume of assessments

	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			X

Deliberate Assessment

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

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

My thoughts:
Deliberate Development vs. Assessment

Deliberate Development

	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			X

Deliberate Assessment

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

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

Building Evidence to Draw Valid Conclusions

EGAD Workshop December 2019
Brian Frank
with Alex Downie, Peter Ostafichuk

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

- Olds, B. M., Moskal, B. M., & Miller, R. L. (2005). Assessment in engineering education: Evolution, approaches and future collaborations. *Journal of Engineering Education*, **94**(1), 13–25. doi:10.1002/j.2168-9830.2005.tb00826.x

- 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

Followup

- Evaluation survey will be emailed to everyone
- Slides and resources will be posted on EGAD GACIP page, hopefully by the end of the week

Final panel