



Session 8

Trailblazing Transformations: Taking Stock and Looking Ahead

Reidy Interactive Learning Series (RILS) Conference
Portsmouth, NH, September 26-27, 2024
AC Marriott Hotel

Access RILS Resources at:

<https://sites.google.com/nciea.org/rils2024>



The screenshot shows the homepage for the Reidy Interactive Learning Series (RILS) 2024. The header includes the RILS 2024 logo and navigation links for Home, About, Agenda, Location, and Presenters. The main content area features the title "The Reidy Interactive Learning Series (RILS)" and the subtitle "Consequential Uses of Assessment: Taking Stock and Looking Ahead". The event dates and location are listed as "September 26-27, 2024 Portsmouth, New Hampshire". Below this, there are four image-based navigation buttons: "About RILS" (with a photo of the Center for Assessment building), "Agenda and Resources" (with a photo of a person writing on a notepad), "Presenters" (with a photo of a group of people), and "Location" (with a photo of a building at night).



Overview

**Panelist
Presentations**



&

**Moderated
Panel**



Join by Web

PollEv.com/cassessment154

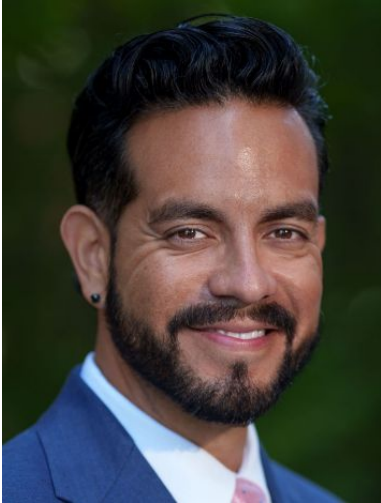
Join by Text

Send **cassessment154** and your message to **22333**

Join by QR code
Scan with your camera app



Panelists



Juan D'Brot
Center for Assessment



Ye Tong
National Board of
Medical Examiners



Susan Lyons
Lyons Assessment
Consulting

Panelist Presentations

Juan D'Brot



How Will We Know if We Get It Right?

Reidy Interactive Learning Series

Portsmouth, NH

September 27, 2024



This work is licensed under a
Creative Commons Attribution 4.0
International License.

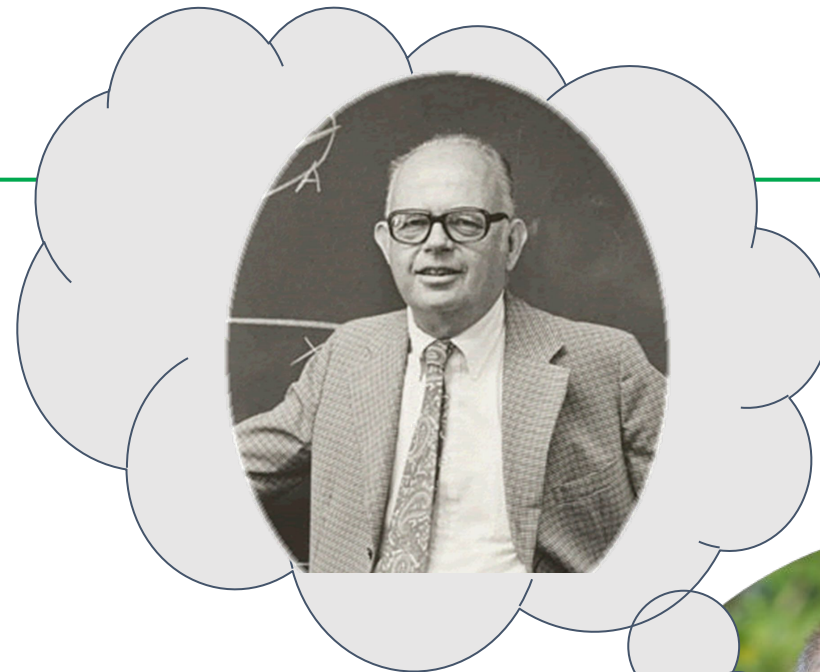
Validation!

The question posed: How will we know if we get it right?

The challenge: How does one condense the idea of validation in 12 minutes?



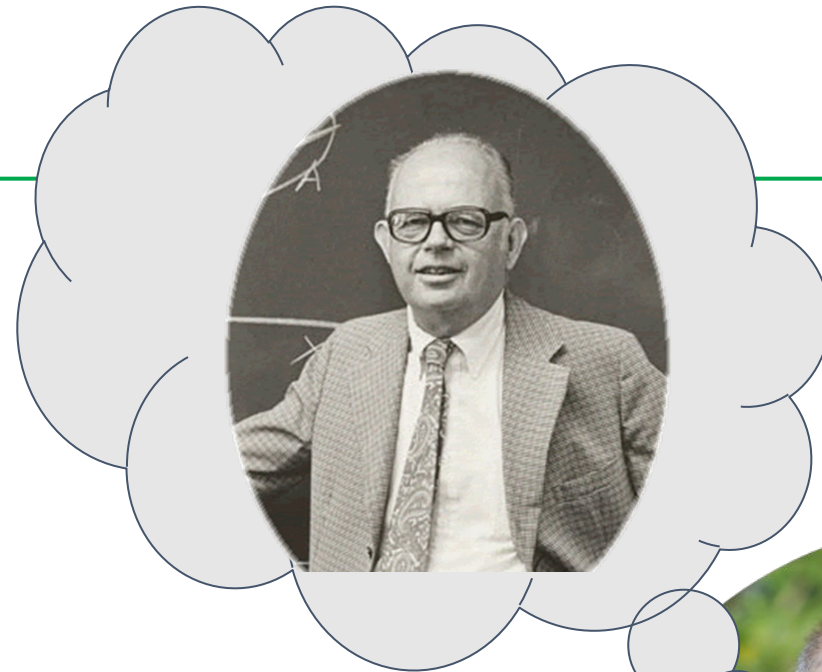
Validation



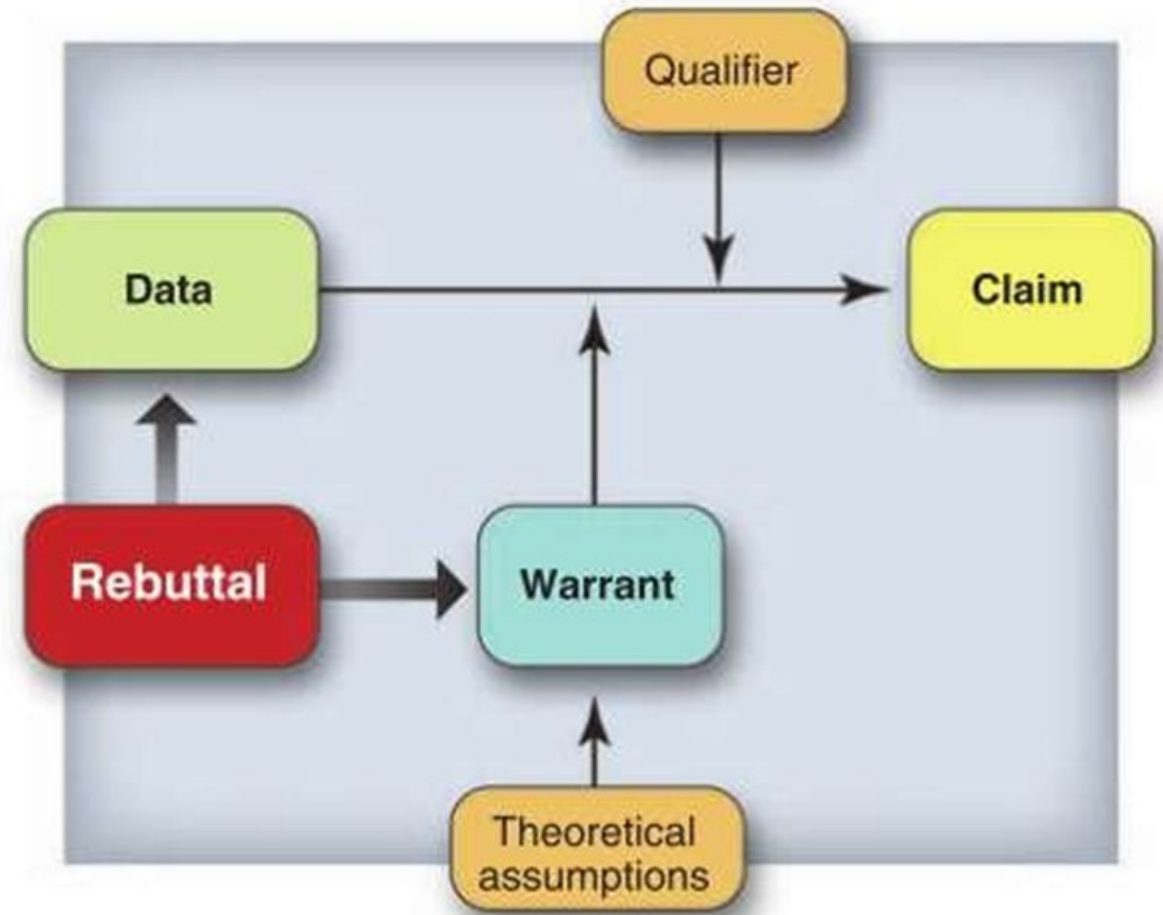
Validation

Validation is
“a lengthy, even
endless process”

(Marion, 2024, citing Cronbach, 1989, p.151)



Validation

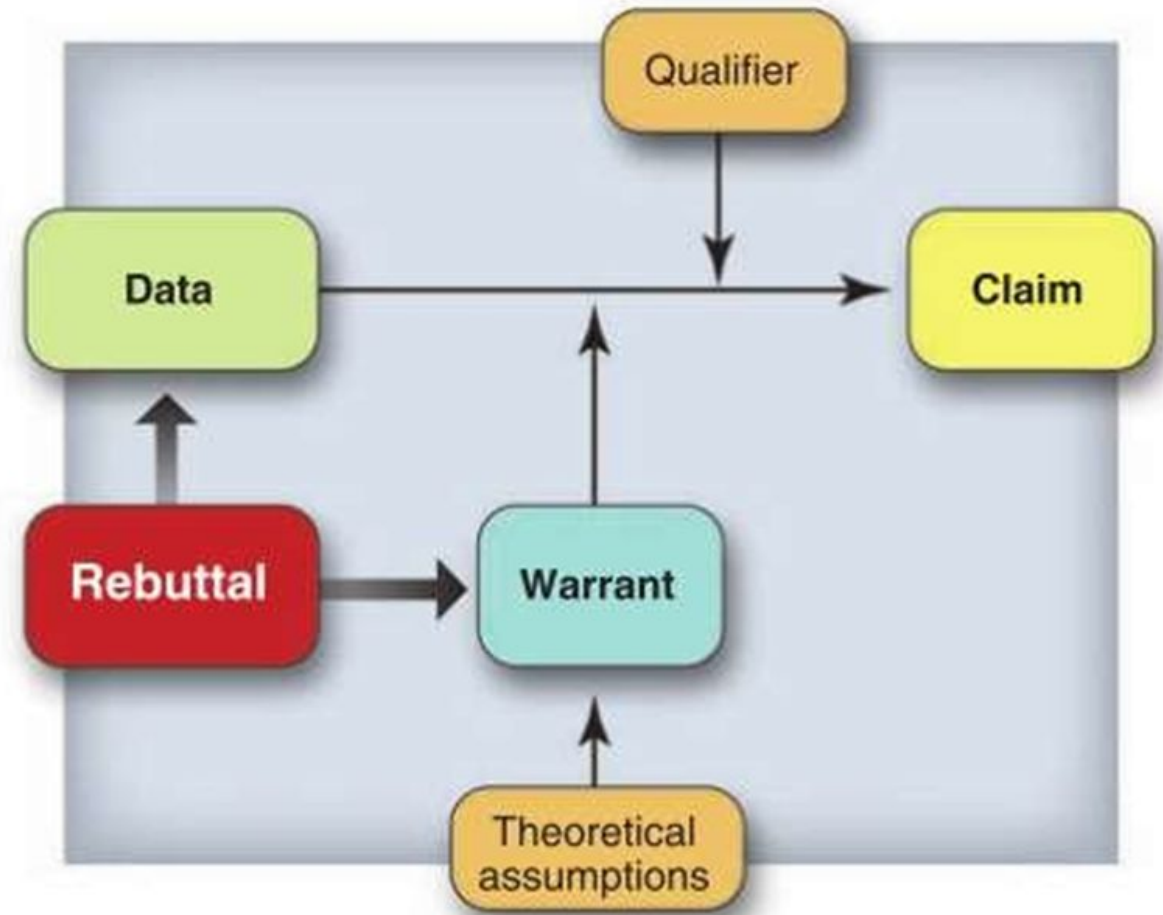


-Kane (2006, 2012) as considered in Lane & Marion (in press)



Validation

The point is not to fully understand this framework, but to know that there is a lot of work out there to inform validation efforts

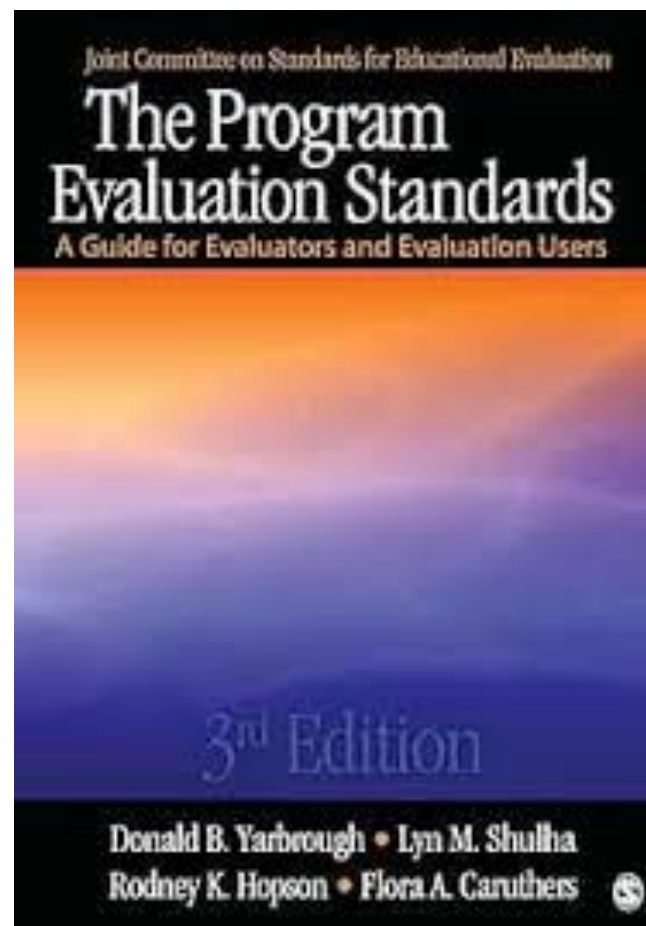


-Kane (2006, 2012) as considered in Lane & Marion (in press)

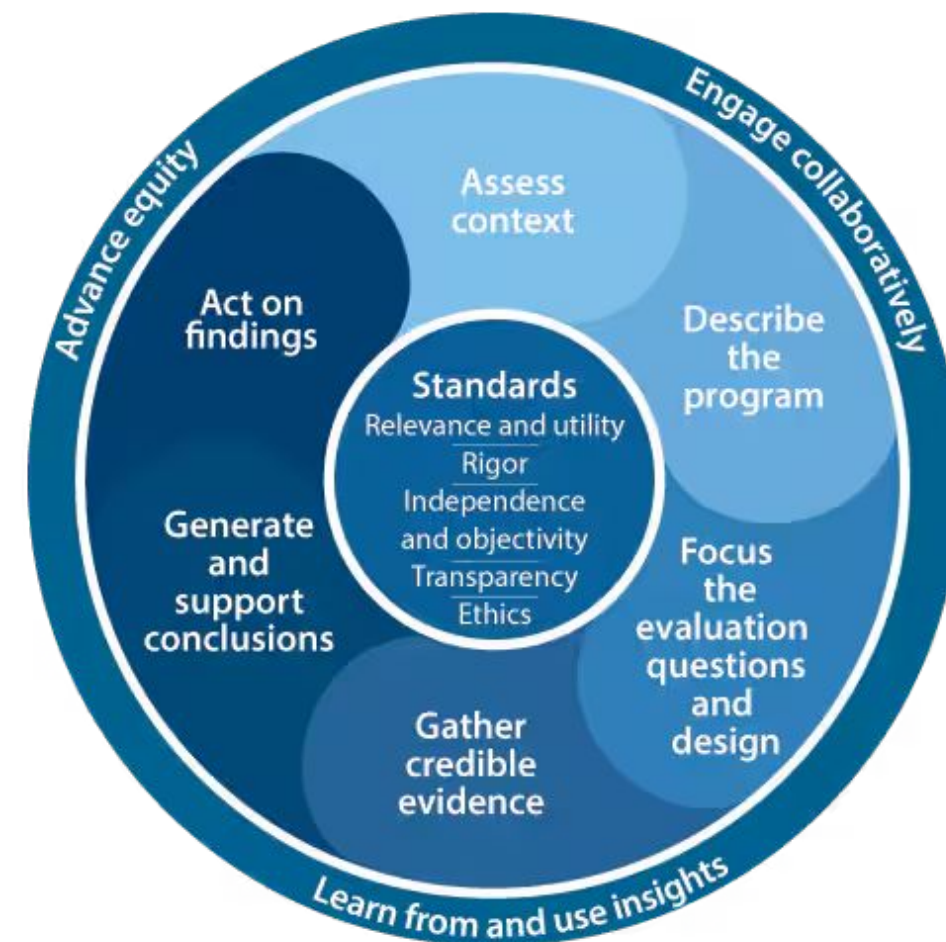


Validation as supported by Evaluation

More references!




Program Evaluation Standards



CDC Evaluation Framework

Validation as supported by Evaluation

Even more references...



Evaluation Checklist Project

Checklist of The Program Evaluation Standards Statements

Joint Committee on Standards for Educational Evaluation

The Program Evaluation Standards “identify and define evaluation question and guide evaluators and evaluation users in the pursuit of evaluation quality” (Yarbrough, Shulha, Hopson, & Caruthers, 2011). The Standards include thirty statements that define five dimensions of program evaluation quality: utility, feasibility, propriety, accuracy, and evaluation accountability. Each standard has a name and is expressed in a statement, which is then explained in more detail in *The Program Evaluation Standards* book (Yarbrough et al., 2011). The standards’ names and statements are reproduced below in checklist form with permission of the Joint Committee on Standards for Educational Evaluation (JCSEE).

The purpose of this checklist version of the Standards is to provide evaluation practitioners, clients, users, and students with an accessible overview of the Standards. We encourage users to read *The Program Evaluation Standards* in full, and then use this checklist as a quick reference.

Utility Standards

The utility standards are intended to increase the extent to which program stakeholders find evaluation processes and products valuable in meeting their needs.

- ❑ **U1 Evaluator Credibility:** Evaluations should be conducted by qualified people who establish and maintain credibility in the evaluation context.
- ❑ **U2 Attention to Stakeholders:** Evaluations should devote attention to the full range of individuals and groups invested in the program and affected by its evaluation.
- ❑ **U3 Negotiated Purposes:** Evaluation purposes should be identified and continually negotiated based on the needs of stakeholders.
- ❑ **U4 Explicit Values:** Evaluations should clarify and specify the individual and cultural values underpinning purposes, processes, and judgments.
- ❑ **U5 Relevant Information:** Evaluation information should serve the identified and emergent needs of stakeholders.

Validation as supported by Evaluation

So back to the question:

How will we know if we get it right?



Checklist of The Program Evaluation Standards Statements Joint Committee on Standards for Educational Evaluation

The Program Evaluation Standards “identify and define evaluation question and guide evaluators and evaluation users in the pursuit of evaluation quality” (Yarbrough, Shulha, Hopson, & Caruthers, 2011). The Standards include thirty statements that define five dimensions of program evaluation quality: utility, feasibility, propriety, accuracy, and evaluation accountability. Each standard has a name and is expressed in a statement, which is then explained in more detail in *The Program Evaluation Standards* book (Yarbrough et al., 2011). The standards’ names and statements are reproduced below in checklist form with permission of the Joint Committee on Standards for Educational Evaluation (JCSEE).

The purpose of this checklist version of the Standards is to provide evaluation practitioners, clients, users, and students with an accessible overview of the Standards. We encourage users to read *The Program Evaluation Standards* in full, and then use this checklist as a quick reference.

Utility Standards

The utility standards are intended to increase the extent to which program stakeholders find evaluation processes and products valuable in meeting their needs.

- ❑ **U1 Evaluator Credibility:** Evaluations should be conducted by qualified people who establish and maintain credibility in the evaluation context.
- ❑ **U2 Attention to Stakeholders:** Evaluations should devote attention to the full range of individuals and groups invested in the program and affected by its evaluation.
- ❑ **U3 Negotiated Purposes:** Evaluation purposes should be identified and continually negotiated based on the needs of stakeholders.
- ❑ **U4 Explicit Values:** Evaluations should clarify and specify the individual and cultural values underpinning purposes, processes, and judgments.
- ❑ **U5 Relevant Information:** Evaluation information should serve the identified and emergent needs of stakeholders.

ct activities, descriptions, interpret, or revise their

tions should attend to the

omote responsible and es and misuse.

d efficiency.

management strategies.

and responsive to the way

balance the cultural and

ficiently.

luations.

ponsive to stakeholders

d to make obligations ntexts of clients and other

conducted to protect ther stakeholders.

air in addressing

ete descriptions of findings, d violate legal and propriety

entify and address real or

d resources and comply

tion judgments

ld be

nd support

e and

rams and

llection,

esigns and

nd analyses

etely

ate scope

and a

sses and

purposes

ards to

on

other

and other

standards: A

CHECKLISTS 3

Validation as supported by Evaluation

So back to the question:

How will we know if we get it right?

Answer:

It depends. It always depends.



Checklist of The Program Evaluation Standards Statements Joint Committee on Standards for Educational Evaluation

The Program Evaluation Standards “identify and define evaluation question and guide evaluators and evaluation users in the pursuit of evaluation quality” (Yarbrough, Shulha, Hopson, & Caruthers, 2011). The Standards include thirty statements that define five dimensions of program evaluation quality: utility, feasibility, propriety, accuracy, and evaluation accountability. Each standard has a name and is expressed in a statement, which is then explained in more detail in *The Program Evaluation Standards* book (Yarbrough et al., 2011). The standards’ names and statements are reproduced below in checklist form with permission of the Joint Committee on Standards for Educational Evaluation (JCSEE).

The purpose of this checklist version of the Standards is to provide evaluation practitioners, clients, users, and students with an accessible overview of the Standards. We encourage users to read *The Program Evaluation Standards* in full, and then use this checklist as a quick reference.

Utility Standards

The utility standards are intended to increase the extent to which program stakeholders find evaluation processes and products valuable in meeting their needs.

- ❑ **U1 Evaluator Credibility:** Evaluations should be conducted by qualified people who establish and maintain credibility in the evaluation context.
- ❑ **U2 Attention to Stakeholders:** Evaluations should devote attention to the full range of individuals and groups invested in the program and affected by its evaluation.
- ❑ **U3 Negotiated Purposes:** Evaluation purposes should be identified and continually negotiated based on the needs of stakeholders.
- ❑ **U4 Explicit Values:** Evaluations should clarify and specify the individual and cultural values underpinning purposes, processes, and judgments.
- ❑ **U5 Relevant Information:** Evaluation information should serve the identified and emergent needs of stakeholders.

ct activities, descriptions, interpret, or revise their

tions should attend to the

omote responsible and es and misuse.

d efficiency.

management strategies.

and responsive to the way

balance the cultural and

ficiently.

luations.

ponsive to stakeholders

d to make obligations ntexts of clients and other

conducted to protect ther stakeholders.

air in addressing

ete descriptions of findings, d violate legal and propriety

entify and address real or

d resources and comply

tion judgments

ld be

nd support

e and

rams and

llection,

esigns and

nd analyses

etely

ate scope

and a

sses and

purposes

ards to

on

other

and other

standards: A

CHECKLISTS 3

Recognizing Complexity and Looking for Simplicity

Getting it Right

So what does “getting it right” even mean?

Getting it Right

It means starting in the right place...

Getting it Right

It means starting in the right place...

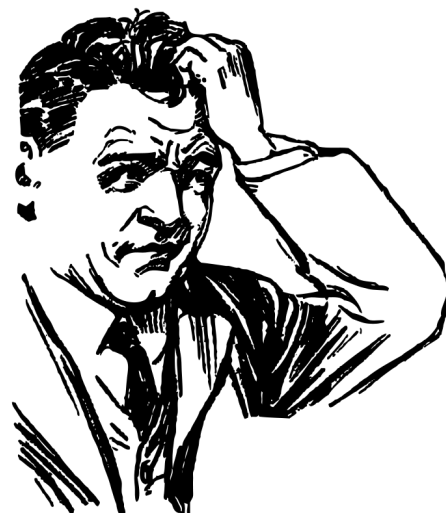
The “right” place could include:

- The vision
- The goal
- The problem
- The initiative
- The validation

Getting it Right



VS.



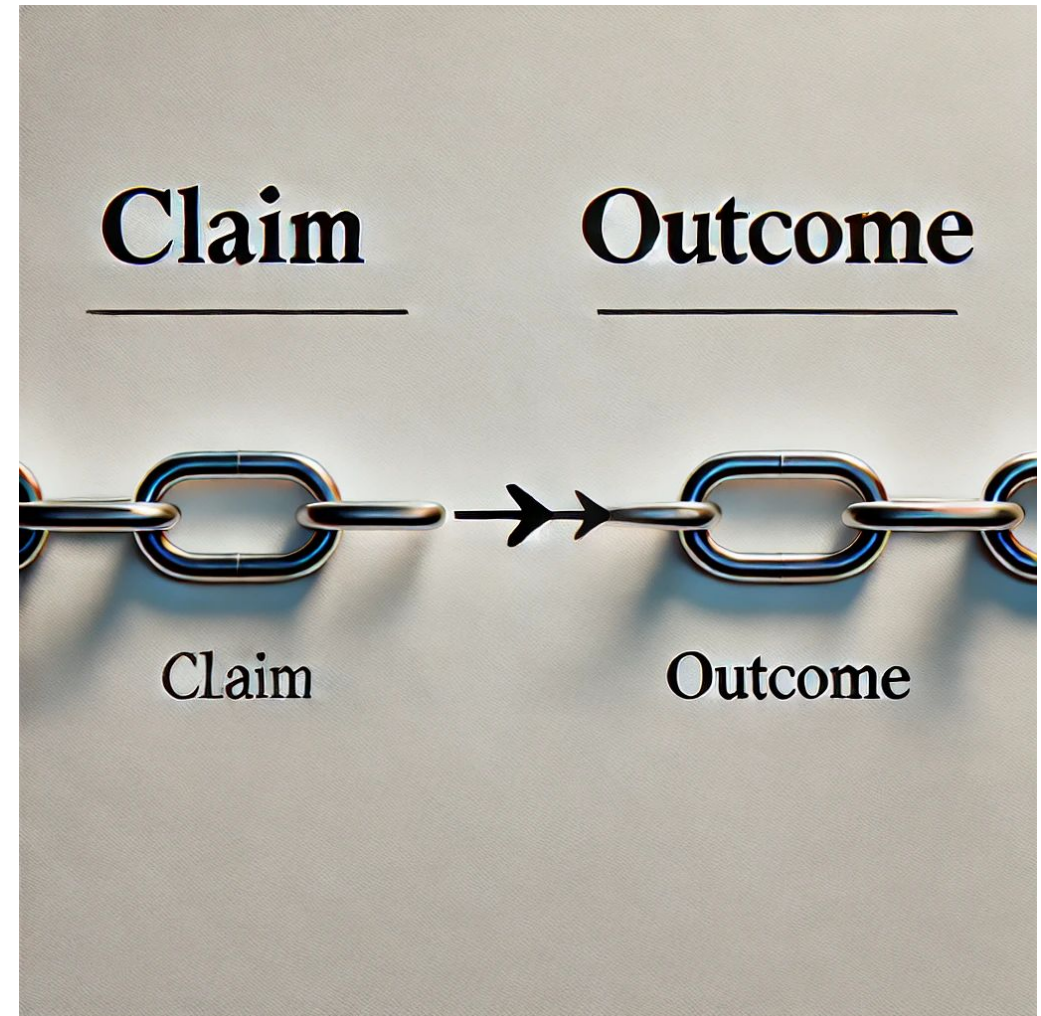
Getting it Right

- The “right place” could be a number of places
- The reality is that the decisions are intertwined, are interconnected, and interdependent.
- Pull on one strand and everything starts to unravel...



(I think) The real questions are...

- How confident are you that each link in your chain (argument) is solid?
- Does it get you to your intended outcome?



(I think) The real questions are...

- How confident are you that each link in your chain (argument) is solid?
- Does it get you to your intended outcome?
- What are the consequences that exist beyond the intended outcome? → How do people's behavior change?



A Process to Support Validation: The Role of Program Evaluation

How to Validate

Proposed (!) Rapid Conceptual Framework

- Built on the following assumptions
 - This is not pseudo- or quasi-evaluation (see Stufflebeam & Coryn, 2014)
 - The outcomes and problem are well-specified
 - There is a focus on impact beyond the primary user, user group, or outcome
 - *The behaviors of people and the impact of that behavior must be addressed*

What is Program Evaluation

Program evaluation is a systematic method for collecting, analyzing, and using information to answer questions about projects, policies and programs, particularly about their effectiveness and efficiency. (Shackman, [2020](#))

But what is a program? It helps to answer the following questions:

- What is the intervention or thing trying to do?
- What should the outcomes be?
- How is the intervention or thing trying to get there?
- How and why does the intervention or program lead to the intended outcome(s)?

Tangent: Limitations with Pseudo-Evaluation

Evaluation that seems like thorough and objective evaluation but lacks the rigor, transparency, and comprehensiveness needed for credible or valid results. This is important to recognize, and it's everywhere.

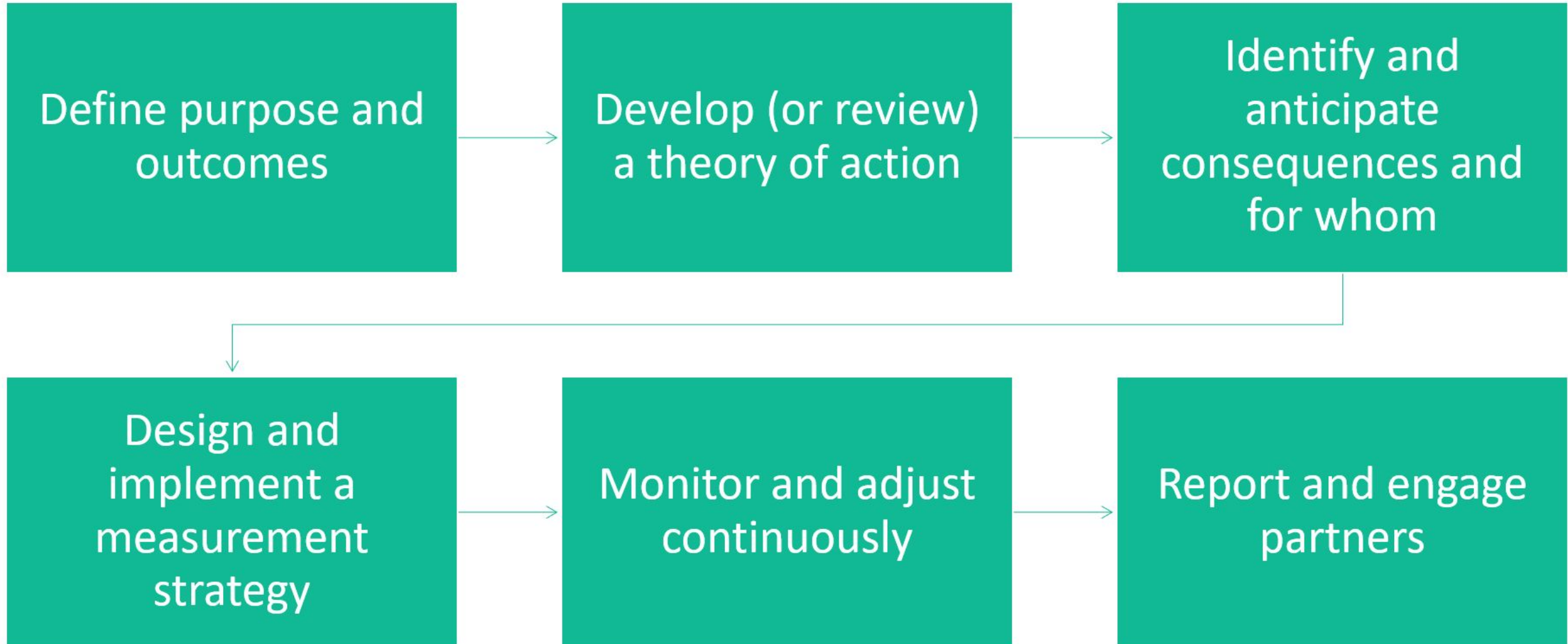
Characteristics:

1. Cherry-picked Data
2. Lack of Clear Objectives or Questions
3. Absence of Partner Input
4. Focus only on Positive Outcomes
5. Lack of Transparency or Methodological Rigor
6. Disregard of Long-Term or Unintended Consequences

Limitations of Decent Evaluation

- **Missing the Bigger Picture:** it is difficult to capture the complex realities of how systems operate in practice (aka the real-world)
- **Overemphasis on Success vs. Failure:** Immediate outcomes often prevail, limiting insight into hidden issues
- **Limited Insight into User Behavior:** Often focuses on whether the system works, not on the behavioral changes of primary users, let alone secondary users.

A Process to Support a Consideration of Consequences



A Process to Help Consider Consequences

Step	Actions	Questions for Consideration	Response
1. Define Purpose and Outcomes	<ul style="list-style-type: none"> - Establish the purpose of the evaluation. - Identify desired outcomes. 	<ul style="list-style-type: none"> - What are the main goals of the evaluation? - What specific outcomes do you want to achieve? - Why is this evaluation necessary? 	
2. Develop or Review a Theory of Action	<ul style="list-style-type: none"> - Develop or review the theory of action. - Identify key mechanisms. 	<ul style="list-style-type: none"> - How does the system work to achieve desired outcomes? - What assumptions are being made about how users will interact with the system? - What are the key drivers of success? 	
3. Identify and Anticipate Consequences	<ul style="list-style-type: none"> - Anticipate intended consequences. - Consider unintended consequences. 	<ul style="list-style-type: none"> - What are the expected outcomes and behaviors? - What unintended consequences could arise? - Who might be affected in unexpected ways? 	

A Process to Help Consider Co

I'm also making the assumption that like good program or intervention designers, we are beginning with the end (see validation plan) in mind, hence the first step including scoping out the evaluation.

Step	Actions	Questions for Consideration	
1. Define Purpose and Outcomes	<ul style="list-style-type: none"> - Establish the purpose of the evaluation. - Identify desired outcomes. 	<ul style="list-style-type: none"> - What are the main goals of the evaluation? - What specific outcomes do you want to measure? - Why is this evaluation necessary? 	
2. Develop or Review a Theory of Action	<ul style="list-style-type: none"> - Develop or review the theory of action. - Identify key mechanisms. 	<ul style="list-style-type: none"> - How does the system work to achieve desired outcomes? - What assumptions are being made about how users will interact with the system? - What are the key drivers of success? 	
3. Identify and Anticipate Consequences	<ul style="list-style-type: none"> - Anticipate intended consequences. - Consider unintended consequences. 	<ul style="list-style-type: none"> - What are the expected outcomes and behaviors? - What unintended consequences could arise? - Who might be affected in unexpected ways? 	

A Process to Help Consider Consequences

Step	Actions	Questions for Consideration	Response
4. Design and Implement Measurement Strategy	<ul style="list-style-type: none"> - Develop or identify metrics to measure outcomes and consequences. - Use mixed methods to gather data. 	<ul style="list-style-type: none"> - What metrics will you use to track both outcomes and consequences? - How will you ensure data collection is comprehensive? - How often will data be collected? 	
5. Monitor and Adjust Continuously	<ul style="list-style-type: none"> - Set up feedback loops for ongoing monitoring. - Analyze data to drive adjustments. 	<ul style="list-style-type: none"> - How will you monitor system performance in real time? - What mechanisms will you put in place to make adjustments? - How will feedback influence system refinement? 	
6. Report and Engage Partners and Users	<ul style="list-style-type: none"> - Report findings transparently. - Engage stakeholders in continuous improvement. 	<ul style="list-style-type: none"> - How will you communicate findings clearly and transparently? - How will partners and users be involved in refining the system? - What collaboration strategies will ensure system improvement? 	

A Process to Help Consider Consequences

This resource is intended to be a relatively quick series of questions to respond to that can be used as a self-reflection or partner check-in. It is focused on assessment systems, but can be applied to any program, system, or intervention.



You can find this resource here: <https://tinyurl.com/3bem4rza>

Additional resources: [Validation blog](#) & D'Brot and Brandt (in press) 3-part paper series available on our website soon.

Evaluation: The Bare Minimum

1. Build out the program logic model.

- TOAs are important (and often implicit)
- Logic models force specificity: activities, necessary resources, parties responsible, and outputs → outcomes.

2. Connect evidence (measures) to each activity in the logic model.

- Translates outputs into measurable data elements
- Evidence may be direct observations of counts, completed tasks, or getting individuals to attend some training. In other cases, the evidence is a little more difficult to capture, like surveys, interviews, or document reviews.
- The coherent connection between the evidence (i.e., data elements) and activities is most important to make judgments during the next step.

3. Collect data and determine evidence quality.

- Collecting high-quality evidence helps monitor incremental progress
- Evidence justifies that intermediate activities yield intended outcomes.
- Examining the intermediate steps can help identify breakdowns and opportunities for redeployment or course correction.

Activity Directions

Activity Directions

We invite you to think about 1 of 3 scenarios at your table:

- **Scenario 1:** Interim assessments
- **Scenario 2:** Large-scale assessments
- **Scenario 3:** Performance tasks

You will be given a scenario and offered a few considerations focused on equity (e.g., outcomes, resources, opportunities, growth). You will then be respond to a few reflection questions.

Scenario 1: Interim Assessments

<https://tinyurl.com/4nyc9cp6>



Scenario 2: Large-Scale Assessments

<https://tinyurl.com/yzak5djh>



Scenario 3: Performance Assessments

<https://tinyurl.com/4br95bn6>



Ye Tong

Assessments in Medical Education

Ye Tong, Ph.D.

September 27, 2024



Medical School Assessment Journey

MCAT

USMLE

Medical
Specialty

Continued
Certification

Other Types of Assessments in Medicine



School developed exams



Clinical exams



Knowledge-based subject exams



“21 century skills” type assessments



Milestone



In-Training exams

Consequences

MCAT

USMLE

Medical
Specialty

Continued
Certification

Consequences



Residency Match

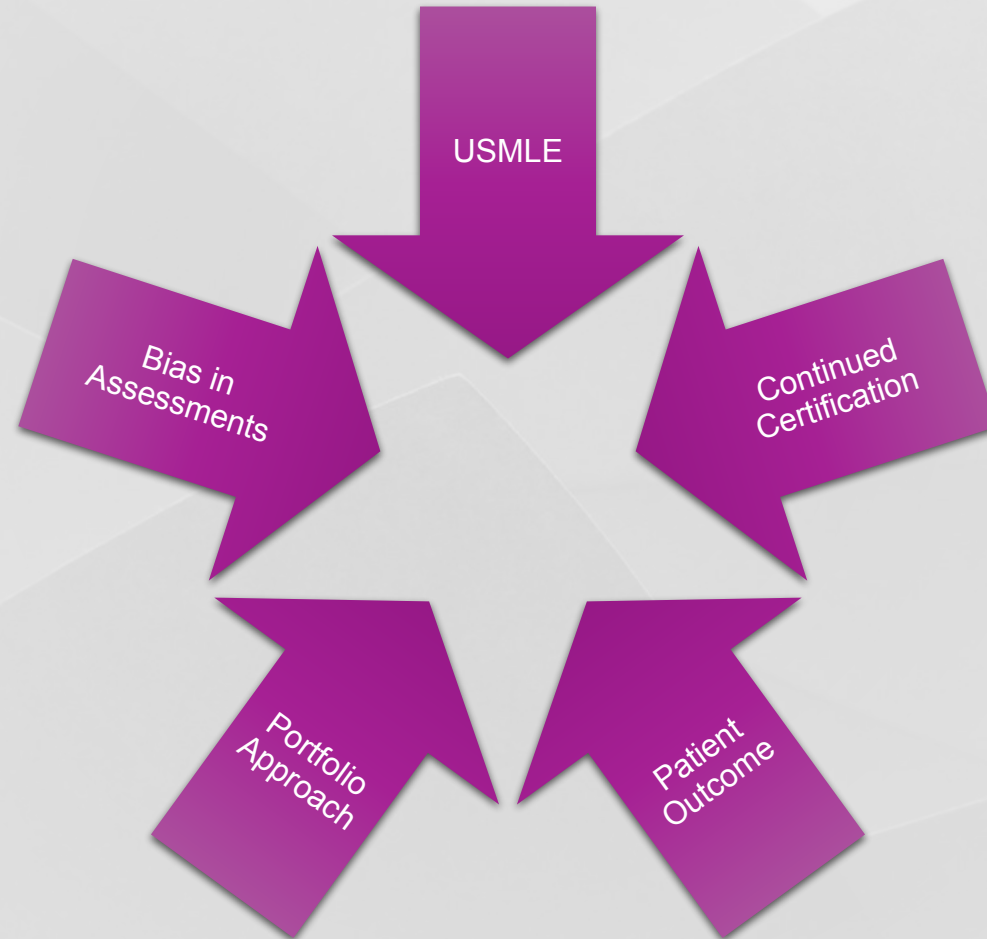


Milestone



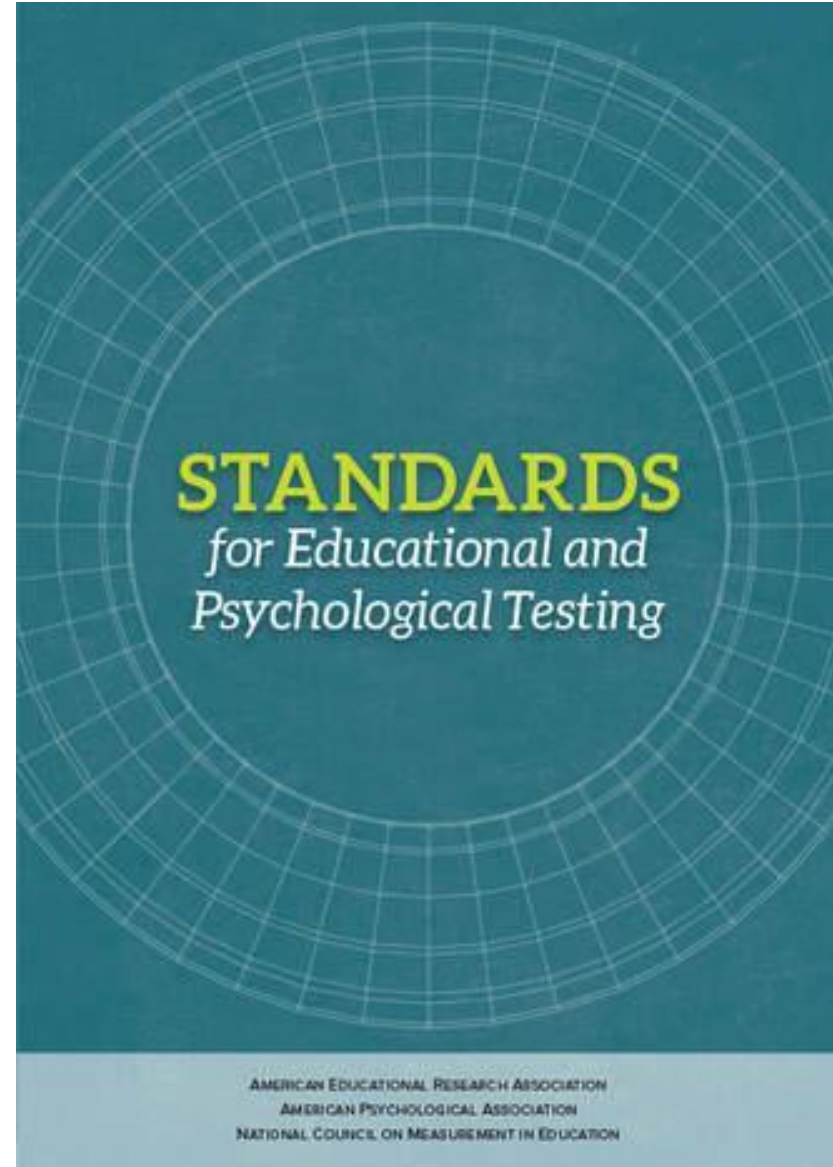
Continued Certification

Framework to Address Consequences



The Standards

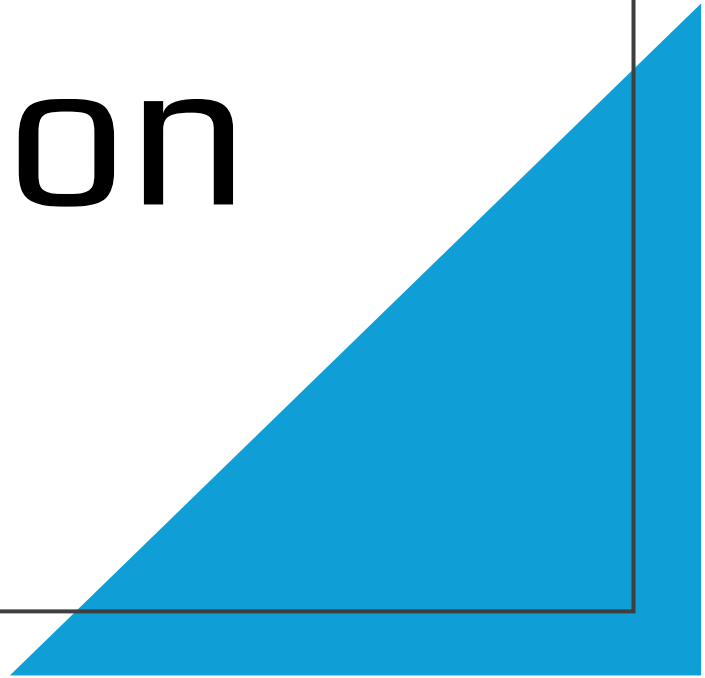
- Revision
- 16-person Committee
- Major themes to address:
 - Fairness and Equity
 - Use of Technology
 - Score misuse (consequence)
- Please provide input



Assessments in Medical Education

Ye Tong, Ph.D.

September 27, 2024



Susan Lyons

**A Framework for Enacting Equity Aims in Assessment Use:
A Justice-Oriented Approach**

Susan Lyons, Lyons Assessment Consulting

Maria Elena Oliveri, Buros Center for Testing, University of Nebraska-Lincoln

Mya Poe, Northeastern University

```
graph LR; A((Identifying Framework Completion Team)) --> B((Interpreting Test Scores)); B --> C((Specifying Uses)); C --> D((Gathering Evidence)); D --> E((Determination));
```

Identifying
Framework
Completion
Team

Interpreting
Test Scores

Specifying
Uses

Gathering
Evidence

Determination

Framework for Justice-Oriented Assessment Use

Identifying the Framework Completion Team	<i>Who comprises the team of people who are completing and discussing this framework?</i>	
Interpreting Individual and Group Differences in Scores	<i>What is the intended interpretation of the test scores?</i>	
	<i>In what ways might the test scores reflect systemic oppression of marginalized examinees?</i>	
	<i>How might we correctly interpret group differences in a multicultural society?</i>	
Specifying Use	<i>In what ways are the test scores intended to be used?</i>	
	<i>What is the range of possibilities by which test scores might be used for additional, unintended purposes?</i>	
Gathering Evidence	<i>How might the intended and/or probable unintended uses of the test scores result in the further entrenchment of existing inequities that harm minoritized people and communities? What evidence supports this theory?</i>	<u>Theory:</u> <u>Supporting Evidence:</u>
	<i>How might the intended and/or probable unintended uses of the test scores interrupt and reshape systemic factors to advance social justice? What evidence supports this theory?</i>	<u>Theory:</u> <u>Supporting Evidence:</u>
Determination	<i>Given the interaction between the interpretation, use, and supporting evidence, is the intended use of the test scores justifiable from a social-justice perspective?</i>	<u>Decision:</u> Yes/No

Framework for Justice-Oriented Assessment Use

Interpreting Individual and Group Differences in Scores	<i>What is the intended interpretation of the test scores?</i>	
	<i>In what ways might the test scores reflect systemic oppression of marginalized examinees?</i>	
	<i>How might we correctly interpret group differences in a multicultural society?</i>	

Race and ethnicity were falsely interpreted as the cause of group score differences, rather than interpreting the group score differences as resulting from the intentional racialization of society that privileges Whiteness (Zuberi, 2000).

Race and ethnicity were falsely interpreted as the cause of group score differences, rather than interpreting the group score differences as resulting from the intentional racialization of society that privileges Whiteness (Zuberi, 2000).

1. Cultural and social processes play a central role in the activity systems that underlie human learning.

Race and ethnicity were falsely interpreted as the cause of group score differences, rather than interpreting the group score differences as resulting from the intentional racialization of society that privileges Whiteness (Zuberi, 2000).

1. Cultural and social processes play a central role in the activity systems that underlie human learning.
2. Cultural activity systems occur within sociopolitical contexts that mediate power by race and class.

(Nasir & Hand,
2006)

Race and ethnicity were falsely interpreted as the cause of group score differences, rather than interpreting the group score differences as resulting from the intentional racialization of society that privileges Whiteness (Zuberi, 2000).

1. Cultural and social processes play a central role in the activity systems that underlie human learning.
2. Cultural activity systems occur within sociopolitical contexts that mediate power by race and class.

(Nasir & Hand,
2006)

Race and ethnicity were falsely interpreted as the cause of group score differences, rather than interpreting the group score differences as resulting from the intentional racialization of society that privileges Whiteness (Zuberi, 2000).

1. Cultural and social processes play a central role in the activity systems that underlie human learning.
2. Cultural activity systems occur within sociopolitical contexts that mediate power by race and class.

(Nasir & Hand,
2006)

Interpreting Group Score Differences

1. The degree to which our society privileges the dominant class and culture in the content, language, format, and scoring of the assessments themselves;

AND

2. Reflections of the systemic limitations and barriers our society places on opportunity and access for marginalized students

Systemic Oppression

1. The degree to which our society privileges the dominant class and culture in the content, language, format, and scoring of the assessments themselves;

AND

2. Reflections of the systemic limitations and barriers our society places on opportunity and access for marginalized students

Revisiting Classical Theory

Observed Score = **Individual factors** + **Systemic factors** +
random error

Moderated Panel



Session 8

Trailblazing Transformations: Putting Ideas into Concrete Action

Reidy Interactive Learning Series (RILS) Conference
Portsmouth, NH, September 26-27, 2024
AC Marriott Hotel

