Evaluation Considerations: Use and Utility

Juan D'Brot

Erika Landl

The National Center for the Improvement of Educational Assessment

September 27th, 2019
Reidy Interactive Lecture Series
Portsmouth, NH

Coupled with Program Evaluation Methods

- 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.
  - Can we generalize our findings?
  - Are we meant to generalize our findings?
  - What can we learn about use of an assessment and whether it is serving our intended purposes?

Reference: Shackman, Gene. *"What Is program evaluation: A beginner's guide"*. The Global Social Change Research Project.

Program Evaluation to obtain evidence of our TOAs

- 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

- Formative Evaluation vs. Summative Evaluation
  - What is the distinction?

- It's all about information!
  - **Formative Evaluation**: Evaluation to improve the design, development, or implementation of a program or effort (D'Brot)
  - **Summative Evaluation**: Evaluation intended to make a retrospective judgment about a program or effort (D'Brot)

Evaluation is Context Dependent

- Assumption: Implementing an assessment system or assessment is a akin to a program

- Consider the Program Evaluation Standards (JCSEE, 2014)
  - Utility
  - Feasibility
  - Propriety
  - Accuracy

- Many (if not all) are stakeholder specific

Evaluation is Context Dependent (PES)

- Utility
  - Credibility
  - Stakeholders
  - Purposes
  - Values
  - Relevance
  - Consequences

- Feasibility
  - Management
  - Procedures

- Propriety
  - Inclusivity and Equity
  - Roles and Responsibilities
  - Human Rights and Respect
  - Fairness
  - Transparency
  - Conflicts of Interest

- Accuracy
  - Justified Conclusions
  - Valid Information
  - Reliable Information
  - Explicit Descriptions
  - Info Management

- Documentation

Evaluation is Context Dependent (PES)

- **Utility Standards:** Intended to increase the extent to which people find evaluation processes and products valuable in meeting their needs.

- **Feasibility Standards:** Intended to increase effectiveness and efficiency.

- **Propriety Standards:** Support what is proper, fair, legal, right and just.

- **Accuracy Standards:** Intended to increase the dependability and truthfulness associated with interpretations about quality.

- **Evaluation Accountability Standards:** Encourage adequate documentation and a metaevaluative perpsective

Considering Stakeholders

- For whom are we evaluating alignment to a Theory of Action?

- As always, it depends
  - Parents
  - Students
  - Educators
  - Administrators
  - Local Leaders (e.g., district, LEA, regional)
  - State Education Agency
  - Community and the Public

Evaluation Depends on Purposes and Uses

- What are some purposes and uses for assessment information?
  - Diagnosis
  - Progress monitoring
  - Evaluation
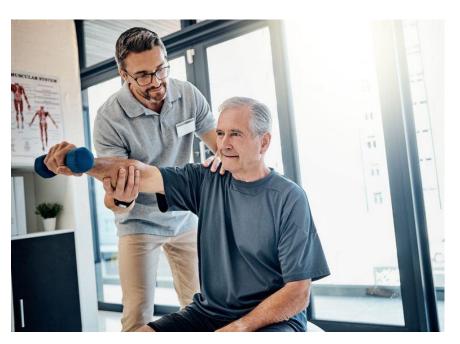 - Curricular decisions

Evaluation Depends on Purposes and Uses

- What are some purposes and uses for assessment information?
  - Diagnosis
  - Progress monitoring
  - Evaluation
  - Curricular decisions
- Evaluation, physical therapy, and sports
  - Bad shoulder ← initial diagnosis
  - Physical therapy ← progress monitoring
  - Can I participate fully? ← final evaluation
  - Should I change my training? ← changing the curriculum

Assessments and Medical Treatment

Diagnosis (Assessment)	Prescription (What should we do)	Treatment (Carrying out with fidelity)	Intended Outcome
?			?
?			?
?			?

Assessments and Medical Treatment

Diagnosis (Assessment)	Prescription (What should we do)	Treatment (Carrying out with fidelity)	Intended Outcome
X	X	X	X

Assessments and Medical Treatment

Diagnosis (Assessment)	Prescription (What should we do)	Treatment (Carrying out with fidelity)	Intended Outcome
X			
X	X		
X	X	X	X

Another Example: Early Warning Systems

Important to delineate between the input, output, environment, and outcome

Consider the Assessment Outcomes

- Assessment X was delivered with alignment to administration constraints

- Students showed little to no evidence of improvement over administrations. Why might that be?

- How do we disentangle those reasons from the results (or lack thereof) we see on the assessment to understand its impact?

What do we want to *actually* evaluate?

Conditions for Implementation Fidelity	Considerations for Appropriate Assessment Use	Assumptions Related to Utility of Assessment Outcomes
• Aligned curriculum • Well-delivered instruction • Well-selected assessment	• Behaviors of educators • Knowledge of users • Mismatch of need and assessment	• Improvements in student learning • Diagnosing progress • Supporting prescription of next steps

What do we want to *actually* evaluate?

1. In small groups (table or larger groups), you will address a few questions

2. Appoint a presenter and a note-taker (not a Center staffer)

3. Discuss the best starting point (or propose your own)

4. Take notes on the flip chart

What do we want to *actually* evaluate?

1. If you were designing Phase 3, where would you start and why (you can even make your own bucket!)?

2. What additional considerations should be included in each bucket?

3. What is the process the district would/should use to answer questions related to these?

4. What are the outputs that would be used to determine that you've met or did not meet these considerations?

5. How would you make these results of this examination relevant and useable?

Conditions for Implementation Fidelity	Considerations for Appropriate Assessment Use	Assumptions Related to Utility of Assessment Outcomes
• Aligned curriculum • Well-delivered instruction • Well-selected assessment	• Behaviors of educators • Knowledge of users • Mismatch of need and assessment	• Improvements in student learning • Diagnosing progress • Supporting prescription of next steps