RILS 2021: Design Innovation in Educational Assessment Systems

The National Center for the Improvement of Educational Assessment

September 2021
Welcome to RILS 2021

Purpose of RILS 2021:

Discuss the different aspects of the design innovation process for assessment systems including:

• exploring the process for designing an innovative assessment system,
• providing an overview of current innovations in assessment systems, and
• delving into the attributes necessary for the design of an innovative assessment system.
<table>
<thead>
<tr>
<th>Session</th>
<th>Date</th>
<th>Presenter(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Session 2:</strong> Understanding the Problems for Design Innovation – <em>Root Cause Analysis</em></td>
<td>Monday, September 20, 2021 1:00-2:30</td>
<td>Juan D’Brot and Chris Brandt Guests</td>
</tr>
<tr>
<td><strong>Session 3:</strong> Leveraging Community for Design Innovation - <em>Engaging Stakeholders</em></td>
<td>Monday, September 20, 2021 3:00-4:30</td>
<td>Carla Evans Guests</td>
</tr>
<tr>
<td><strong>Session 4:</strong> Planning for Design Innovation – <em>Assessment Systems and Theory of Action</em></td>
<td>Thursday, September 23, 2021 1:00-2:30</td>
<td>Erika Landl Nathan Dadey</td>
</tr>
<tr>
<td><strong>Session 5:</strong> Exploring the Design Innovation Process – <em>Iteration in Assessment System Design</em></td>
<td>Thursday, September 23, 2021 3:00-4:30</td>
<td>Brian Gong Guests</td>
</tr>
<tr>
<td><strong>Session 6:</strong> Exploring the IADA Innovation Process – <em>Challenges and Opportunities</em></td>
<td>Friday, September 24, 2021 1:00-2:30</td>
<td>Scott Marion and Carla Evans Chris Domaleski</td>
</tr>
</tbody>
</table>
Design Innovation – What is it?

Ideology

+ 

Process

To solve “wicked” problems in a user-centric way

WICKED PROBLEM.
noun | wi-kəd l prä-blem

a problem that is difficult or impossible to solve because of incomplete, contradictory, and changing requirements that are often difficult to recognize.
Assessment Systems as a **Wicked Problem**

How to appropriately assess students?

- What does this mean?
- What is the end result
- Will it work for all students?
- Will it work in all content areas?
- Will it work in all environments?
- How does the solution of one aspect impact the next situation?
One Model for the Design Innovation Process

**UNDERSTANDING**
- Hypothesize
- Plan
- Discover
- Identify and engage a range of diverse stakeholders
- Identify problem statement(s)
- Engage in root cause analysis
- Create a Theory of Action underlying the assessment system design
- Clarify, analyze, and synthesize learning
- Repeat and review process to ensure understanding and issues of equity

**PROTOTYPING**
- Test
- Create
- Adjust
- Design assessments for user needs
- Administration with a small sample
- Collect data and review applying pre-established criteria
- Prioritize user needs
- Redesign, refine and revise assessments based on formative evaluation and user feedback
- Repeat with multiple rounds of iterations at multiple levels

**SCALING**
- Evaluate
- Pilot
- Refine
- Set up and run the assessment system with pilot group(s)
- Collect data and evaluate applying evaluation criteria
- Gather insights needed to redefine or revise
- Improve user satisfaction
- Repeat to make revisions for large-scale use
- Share the findings broadly

RILS 2021: Innovation in Educational Assessment Systems
“Understanding” Stage

Hypothesize
Discover
Plan

- Identify and engage a range of diverse stakeholders
- Identify problem statement(s)
- Engage in root cause analysis
- Create a Theory of Action underlying the assessment system design
- Clarify, analyze, and synthesize learning
- Repeat and review process to ensure understanding and issues of equity

Root Cause Analysis
Stakeholder Engagement
Identify problem statement(s)
Theory of Action
Clarify, analyze, synthesize

RILS 2021: Innovation in Educational Assessment Systems
Session 4 Focus

"Design adds value faster than it adds costs." -- Joel Spolsky, web programmer, writer, and creator of Trello

Planning for Design

- Defining Theory of Action
- Characteristics of a Balanced Assessment System
- Questions within a Theory of Action for a Balanced Assessment System
- Using the Template
- Q&A / Closure
Planning for Design –
Theory of Action for a
Balanced Assessment
System

Erika Landl

Nathan Dadey
Leveraging Theories of Action to Support Design Innovation in Assessment Systems

Nathan Dadey & Erika Landl, with contributions from Brian Gong

The National Center for the Improvement of Educational Assessment
1 Theory of Action
A brief review of theory of action within the context of assessment systems

2 Dimensions of Innovation
Defining kinds of innovation within a balanced assessment system

3 Example Theories of Action
Worked, hypothetical examples of assessment systems with innovative components

4 Q&A
Question and answer session facilitated by Scott Marion
1. Theory of Action

Would program logic by any other name smell as sweet?

-Shakespeare, probably
Design innovation is complex.

Managing this complexity is critical.

A theory of action is *the* approach to help manage this complexity.
A Theory of Action is

A logical argument that connects the goals of a system to its component parts

By describing the actions and conditions that lead to the goals

as well as the rationales, assumptions and evidence that support and justify the connections within the system

For a deep dive in the context of general programs, see Patton (2008), particularly chapter 10. For consideration of theory of action in the context of assessment, see the Center’s Theory of Action Template.
The Importance of a Theory of Action

A theory of action makes **design of a system explicit** and in doing so:

- Acts as a roadmap for design and a touchstone for iterative design
- Provides shared language & understanding
- Supports the investigation of problem areas
A Simple Logic Model Framing

- There are a number of ways to visualize and structure a theory of action, including Logic Models (e.g., Frechtling, 2007, W.K. Kellog Foundation, 1998) or Driver Diagrams (e.g., Bennett & Provost, 2015).
- We present theories of action graphically. Others use formats like tables (e.g., SCILLSS, 2017, p. 5). There is no one correct format and each application is tailored by the developer.
# A Simplified, Hypothetical State Accountability Plan

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Action Mechanisms</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>School Identification</td>
<td>A Turn Around Plan is Developed and Implemented</td>
<td>Students are Provided with Individualized</td>
</tr>
<tr>
<td>Triggers Support</td>
<td></td>
<td>Supports</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Student Achievement Improves</td>
</tr>
</tbody>
</table>
Implications for Innovation Design?

So what does this mean for design innovation?
Implications

- There are **multiple levels of complexity**
- Innovation design requires us to both
  - **Drill down** into specific parts of the theory of action, and
  - **Attend to the whole**.
- While also being clear about **what** is being innovated on:
  - Assessments, whether in part of whole, and
  - The use of that assessment information.

---

¹This focus on linking assessments explicitly to the program they function within draws from the work that ties score use to interpretation (e.g., Cronbach’s work throughout the 70’s and 80’s; Kane, 2006, p. 53; and more recently, Bennett, Kane & Bridgeman, 2011).
This theory of action is just one of many that are operating in an educational context.
**Levels**

**State.** Statewide Accountability Assessment

**District.** District-Wide Middle of Year Interim (e.g., Benchmark)

**Classroom.** End of Unit & Mid-Unit Check in Assessments, Weekly Exit Tickets, Daily Conversations

---

**Uses**

- ESSA School Identification & Support
- District Resource Allocation
- Formative Assessment Cycle for Tailored Instruction

---

### Instructional Unit

<table>
<thead>
<tr>
<th>Quarter 1</th>
<th>Quarter 2</th>
<th>Quarter 3</th>
<th>Quarter 4</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Levels

State. Statewide Accountability Assessment

District. District-Wide Middle of Year Interim (e.g., Benchmark)

Classroom. End of Unit & Mid-Unit Check in Assessments, Weekly Exit Tickets, Daily Conversations

What are we trying to innovate?

Instructional Unit

Quarter 1 | Quarter 2 | Quarter 3 | Quarter 4

Uses

ESSA School Identification & Support

District Resource Allocation

Formative Assessment Cycle for Tailored Instruction
Levels

State. Statewide Accountability Assessment

District. District Wide Middle of Year Interim (e.g., Benchmark)

Is it, for example, an aspect of the state test a small grain-size?

Classroom. End of Unit & Mid-Unit Check in Assessments, Weekly Exit Tickets, Daily Conversations

Uses

ESSA School Identification & Support

District Resource Allocation Formative Assessment Cycle for Tailored Instruction

Quarter 1 | Quarter 2 | Quarter 3 | Quarter 4
Or is it some much larger aspect of the state assessment that has implications both for the state level use, but also implications for other levels.
### Levels

**State.** Statewide Accountability Assessment

**District.** District Wide Middle of Year Interim (e.g., Benchmark)

**Classroom.** End of Unit & Mid-Unit Check in Assessments, Weekly Exit Tickets, Daily Conversations

<table>
<thead>
<tr>
<th>Instructional Unit</th>
<th>Instructional Unit</th>
</tr>
</thead>
<tbody>
<tr>
<td>Quarter 1</td>
<td>Quarter 2</td>
</tr>
<tr>
<td>Quarter 3</td>
<td>Quarter 4</td>
</tr>
</tbody>
</table>

---

### Uses

- **ESSA School Identification & Support**
- **District Resource Allocation**
- **Formative Assessment Cycle for Tailored Instruction**

---

**Or is the innovation crossing all levels?**
State. Statewide Accountability Assessment

District. District-Wide Middle of Year Interim (e.g., Benchmark)

Classroom. End of Unit & Mid-Unit Check in Assessments, Weekly Exit Tickets, Daily Conversations

Instructional Unit

Quarter 1 | Quarter 2 | Quarter 3 | Quarter 4

Levels

Uses

ESSA School Identification & Support

District Resource Allocation Formative Assessment Cycle for Tailored Instruction

Ultimately, our theory of action can and should acknowledge other levels.
Theory of Action in terms of Balanced Assessment Systems

• This framing is rooted in a balanced assessment system approach.

• This framing helps shift our focus to include the information provided to the assessments and the use of that information within the theory of action (see the Appendix for more detail).

Knowing What Students Know (2001), crystallized the appeal for balanced systems of assessment (see p. 253 – 257). Recent work has shown that building these systems faces a number of barriers (e.g., Marion et al., 2019a; Marion et al., 2019b; Conley, 2018). These challenges are also considered in a 2018 Special Issue of EM:IP.
Multiplicative Roles of TOA in Design Innovation

Theory of Action

- Root Cause Analysis
- Identify problem statement(s)
- Stakeholder Engagement

Serves as a tool by which to explicate the assumptions, inputs and actions necessary to ensure the innovation has the desired impact.

Serves to highlight the differences between typical programs and more innovative programs which supports communication.
Implications of Layered Design of TOA on Innovation

• When it comes to design innovation in a balanced assessment system understanding how/where the smaller of theories of action support or potentially constrain the innovation is key.

• Must understand the interplay among the components, as they currently exist, and also understand where and how a proposed innovation may impact what is happening within and across different levels of the system.

• The impact could be significant and broad or minor and isolated depending on the dimension of change necessitated.
System vs Component Level Innovation

How much are we innovating?

Full System

Innovation

Innovation

Innovation

www.nchiea.org
Dimensions of Innovation

- The dimension of innovation represents the key element(s) that you are looking to modify in order to address an existing problem or improve upon a current existing state.
  - Can be broad or granular.
  - Can be simple or complex
  - Can impact the entire assessment system, (across levels), a level within the system, or a specific process/test within a level.

- In some cases an innovation results from implementing an existing solution at a different level of a system. (e.g., using a locally scored performance assessments as an element of the state summative assessment).
Identifying the Focus of Innovation in BAS

The *what* of innovation may stem from:

- root cause analysis
- advances in technology
- an “event” by
  - establishing new requirements or flexibilities (federal or state laws)
  - shining a light on deficiencies in existing systems
  - reflecting a shift in values or priorities
<table>
<thead>
<tr>
<th>Where/What is the innovation?</th>
<th>Example</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Test Score</strong></td>
<td>Type (NCE, Scaled Score, Growth Score), Interpretation (Criterion/Norm Reference), Estimation (Multidimensional IRT), Use (e.g., Instruction &amp; Accountability), Reporting</td>
</tr>
<tr>
<td><strong>Test Items or Tasks</strong></td>
<td>Type (TEI, OE) ; Scoring (Machine, AI, Human)</td>
</tr>
<tr>
<td><strong>Test Form</strong></td>
<td>Design (Length, Content, Representation) ; Administration (standardized/individual); Mode of Delivery (paper-pencil/computer/CAT)</td>
</tr>
<tr>
<td><strong>Test Development/Evaluation</strong></td>
<td>Who is involved (community, representation); How quality is evaluated/endorsed (e.g., peer review);</td>
</tr>
<tr>
<td><strong>Construct Definition</strong></td>
<td>Traditional, Anti-racist</td>
</tr>
<tr>
<td><strong>Theory of Learning</strong></td>
<td>Learning progressions; Learning expectations/ appropriate demonstrations of learning</td>
</tr>
<tr>
<td><strong>Infrastructure</strong></td>
<td>Design of Data and Learning Management systems</td>
</tr>
<tr>
<td><strong>Role of State/District</strong></td>
<td>Partner and Resource Provider, Auditor</td>
</tr>
</tbody>
</table>
Why useful to highlight the key dimension(s) underlying a proposed innovation?

• Serves to highlight/focus articulation of the theory of action (e.g., inputs, interactions, assumptions) on those things that not well understood or have not yet been played out within the system.

• And, in doing so, clarifies one’s understanding of how/where the innovation will have an impact on the assessment system.
<table>
<thead>
<tr>
<th>Problem</th>
<th>Source</th>
<th>Hypothesized Solution (Based on RCA)</th>
<th>What is the innovation?</th>
<th>Level of Impact</th>
</tr>
</thead>
<tbody>
<tr>
<td>Persistent gaps in achievement outcomes across student groups</td>
<td>State</td>
<td>Engage stakeholders to review and modify the content standards to ensure they outline expectations are appropriate/fair for all students</td>
<td>Construct or Domain Specification (process and content)</td>
<td>All</td>
</tr>
<tr>
<td>Persistent gaps in achievement outcomes across student groups</td>
<td>District</td>
<td>Provide educators in schools serving predominantly low income and historically disadvantaged students with access to high quality educational resources including curriculum embedded assessments that provide ongoing feedback and support progress monitoring.</td>
<td>District role Curriculum and assessment design (high quality, aligned, relevant, accessible)</td>
<td>District, classroom</td>
</tr>
</tbody>
</table>
3. Examples & Tools

*Unfortunately, no one can be told what a Theory Of Action is. You'll have to see it for yourself.*

-Morpheus, probably
Example 1: District has not been effective in supporting the success of it’s lowest performing students.

Example 2: The summative assessment does not provide educators with information in time for within year instructional shifts.
Example 1

District assessment system is incoherent and not effective in supporting the success of all students.
The Challenge for Example 1

• District has not been effective in supporting the achievement and growth of its lowest performing students.

Vision for Teaching and Learning

• What conditions need to hold to support instruction and help students learn?

Role of Assessment

• What role does assessments play within that vision?
• What information do different stakeholders need to be successful?

Role of District

• What role should the district play to support this vision?
• What components and information should it provide versus inform?
### Part 3: Assessment Vision: The assessment information we value, irrespective of what we have

After thinking about what students, teachers, and principals should do that promote student learning, it is important to think about the kinds of assessment information that support these activities. What type of information would each group of people find most useful in their work? What, specifically, would help them make better decisions about their next steps, instruction, planning, or guidance?

**Activity for Part 1:** In each row of the table, briefly respond to each question about what information an assessment could help benefit them. The responses to the questions should be related to the response from Part 1. Please note that assessment information does not have to be the results of a test, but can include a process of gathering evidence.

| Student | Generally, the known outcomes of the learning in the classroom, student comments, grades, etc.
| Teacher | The current learning goals, the curriculum, the teacher's beliefs, etc.
| Principal | The needs of the school, the school's mission and vision, etc.
| Local Educational Leaders | The needs of the district, the district's mission and vision, etc.
| Others | Please add as many rows as necessary.

<table>
<thead>
<tr>
<th><strong>What assessment information would be most helpful for each question?</strong></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Student</strong></td>
</tr>
<tr>
<td>Which kind of assessment information do students need given your vision for teaching and learning and why? How should students use assessment information given your vision for teaching and learning?</td>
</tr>
<tr>
<td><strong>Teacher</strong></td>
</tr>
<tr>
<td>What type(s) of assessment information do teachers need to facilitate student learning and why? How should teachers use each type of assessment information?</td>
</tr>
<tr>
<td><strong>Principal</strong></td>
</tr>
<tr>
<td>What assessment information do principals need and why (e.g., support teachers)? How should principals use each type of assessment information?</td>
</tr>
<tr>
<td><strong>Local Educational Leaders</strong></td>
</tr>
<tr>
<td>What assessment information do district leaders need and why? How should district leaders use each type of assessment information?</td>
</tr>
<tr>
<td><strong>Others</strong></td>
</tr>
<tr>
<td>Others' needs and uses (please add as many rows as necessary)</td>
</tr>
</tbody>
</table>
Identifying Sources of Information

State. Statewide Accountability Assessment

District. District-Wide Middle of Year Interim (e.g., Benchmark)

Classroom. End of Unit & Mid-Unit Check in Assessments, Weekly Exit Tickets, Daily Conversations

- What sources/tools/processes will provide the needed information and support the intended use?

- What already exists? How is it working?

- What are the gaps?
## Defining the Parameters of Innovation

<table>
<thead>
<tr>
<th>The Why (RCA)</th>
<th>Hypothesized Solution</th>
<th>Nature of innovation</th>
</tr>
</thead>
</table>
| • Constant changes to assessment and curriculum  
• District expectations for student performance are unclear and defined differently across schools.  
• High numbers of transient students  
• Lack of assessment literacy  
• Educator use/modification of pre-existing tools that do not align to the curriculum | Providing schools with resources that serve to clarify and demonstrate the expectations of the standards (e.g., rubrics/exemplars) and common district-developed interim assessments that allow for monitoring of student progress relative to those expectations will foster collaboration and consistency within and across educators and schools. | Role of the district in defining and measuring expectations for student performance.  
Establishing resources that provide for coherence across schools in the absence of a common curriculum or pacing. |
The Center’s Theory of Action Template is available online. Please use and modify as you see fit.
Addressing Innovation in the TOA

How/where will this innovation be represented in the District’s theory of action?

• Specification of the characteristics of system components
• Identification of assumptions underlying the revised system working as intended/specification of potential solutions.
  • Quality and utility of district-provided resources
  • Assessment literacy
• Conditions and inputs necessary to ensure the system works as intended.
Example 2

District assessment system is incoherent and not effective in supporting the success of all students.
Hypothetical Example High Level Theory Of Action for a Through-Year Assessment Program (Math)

<table>
<thead>
<tr>
<th>If the state provides:</th>
<th>Then:</th>
<th>So that:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Assessment scores in early fall, winter and spring based through course administration</td>
<td>• Educators will adjust instruction to meet the needs of students</td>
<td>• Students achievement improves</td>
</tr>
<tr>
<td>• Assessment literacy training on use of assessment data</td>
<td>• Administrators will examine trends to allocate additional support</td>
<td></td>
</tr>
</tbody>
</table>

This topic is the subject of an upcoming online convening hosted by the Center: [Claims and Evidence for Through Year Assessments: What We Know and What We Need to Know](https://example.com/).
Framing Assumptions

• Educators will value and engage with state provided (a) professional development and (b) assessment information

• The provided assessment evidence can be used by educators to adjust instruction

• Adjusting instruction is sufficient to improve student learning and resulting student achievement

Which of these assumptions are both critical and risky? Which of these do we need to investigate?
Hypothetical Example *High Level Theory Of Action* for a Through-Year Assessment Program (Math)

<table>
<thead>
<tr>
<th>If the state provides:</th>
<th>Then:</th>
<th>So that:</th>
</tr>
</thead>
<tbody>
<tr>
<td>• Assessment scores in <em>early fall, winter and spring</em> based through course administration</td>
<td>• Educators will adjust instruction to meet the needs of students</td>
<td>• Student learning improves</td>
</tr>
<tr>
<td>• Assessment literacy training on use of assessment data</td>
<td>• Administrators will examine trends to allocate additional support</td>
<td></td>
</tr>
</tbody>
</table>

We will drill down into these areas.
Example *Slightly More Detailed* Logic Model

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Action Mechanisms</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Teachers learn to better use assessment data</td>
<td>Student Learning Improves</td>
</tr>
<tr>
<td></td>
<td>Assessment Score(s) indicating Student Mastery</td>
<td>Teachers make instructional adjustments</td>
</tr>
<tr>
<td></td>
<td>Interpretive Guide</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Assessment Literacy Professional Development</td>
<td></td>
</tr>
</tbody>
</table>

|------|------|------|------|------|

RILS 2021: Innovation in Educational Assessment Systems
Teachers make instructional adjustments

Teachers access reports in a timely fashion

Teachers understand the score reports

Teachers triangulate assessment results with classroom assessments

Teachers determine what kinds of instructional supports are needed

Teachers implement supports at the individual and small group level

Student misconceptions on subdomain concepts are addressed
Specific Assumptions

- Teachers have access to, and understand, a variety of instructional supports connected to the math subdomains.
- Teachers can use assessment evidence to select amongst the various kinds of support.
- The assessment evidence is needed to select amongst these kinds of support.

Teachers determine what kinds of instructional supports are needed.
Example *Slightly More Detailed* Logic Model

<table>
<thead>
<tr>
<th>Inputs</th>
<th>Action Mechanisms</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interpretive Guide</td>
<td>Teachers make instructional adjustments</td>
<td>Student Learning Improves</td>
</tr>
<tr>
<td>Assessment Score(s) indicating Student Mastery</td>
<td>Teachers learn to better use assessment data</td>
<td></td>
</tr>
<tr>
<td>Assessment Literacy Professional Development</td>
<td>Assessment Score(s) indicating Student Mastery</td>
<td></td>
</tr>
</tbody>
</table>

**Inputs:***

- Interpretive Guide
- Assessment Literacy Professional Development

**Action Mechanisms:**

- Teachers make instructional adjustments
- Teachers learn to better use assessment data

**Effects:**

- Student Learning Improves

**Timeline:**

|------|-------|------|------|------|

RILS 2021: Innovation in Educational Assessment Systems
4. Q&A

Wisdom from Scott
Facilitator: Scott Marion
All materials shared here, including the **Theory of Action Template**, can be found on a shared Google Drive folder & will be posted to [www.nciea.org](http://www.nciea.org) shortly.
Appendix: Additional References
A Balanced Assessment System is:

Multiple assessments with potentially different designs, sponsored by different people, who are at different levels of control.

Coordinated by a common theory of learning.

Working together to meet a specific use or uses.
In Other Words...

• to have the desired impact:
  • “to provide information and data that informs policy, programs, and individual teachers and learners in a coherent and coordinated manner….and improve student learning and school capacity” (Gong, 2010).

• a balanced assessment system cannot be a bunch of assessments that are designed and implemented independently of one another.
Conceptualizing the Components of a Balanced Assessment System

Levels (e.g., Dadey, 2018; Shepard & Penuel, 2018)

Type or Tier (e.g., Perie, Marion & Gong, 2009; Sigman & Mancuso, 2017)

Purpose (e.g., NRC, 2014)

Regardless of how the system is conceptualized, the “overall” theory of action can and must connect all of the parts together for it to be truly balanced.