

# On Being Specific: Identifying Uses and Developing Systems to Match

Nathan Dadey & Joseph Martineau\*

*Center for Assessment*

\*authorship is alphabetical

Reidy Interactive Lecture Series

Portsmouth, NH

September 27, 2018



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

# Context

Literature on systems of assessments\* emphasizes both the need for a **common theory of learning** and supporting a variety of **intended purposes and uses.**



This presentation is aimed at the later, exploring use and how it connects to the *design* of assessments.

\*i.e., balanced, comprehensive, or next generation assessment systems. See Pellegrino, Chudowsky & Glaser, (2001); NRC, (2014); Perie, Marion, & Gong; (2009); Darling-Hammond & Pecheone; (2010); Herman, (2010); Herman, (2017).

# Some Big Questions

This exploration runs into a number of big questions along the way, including:

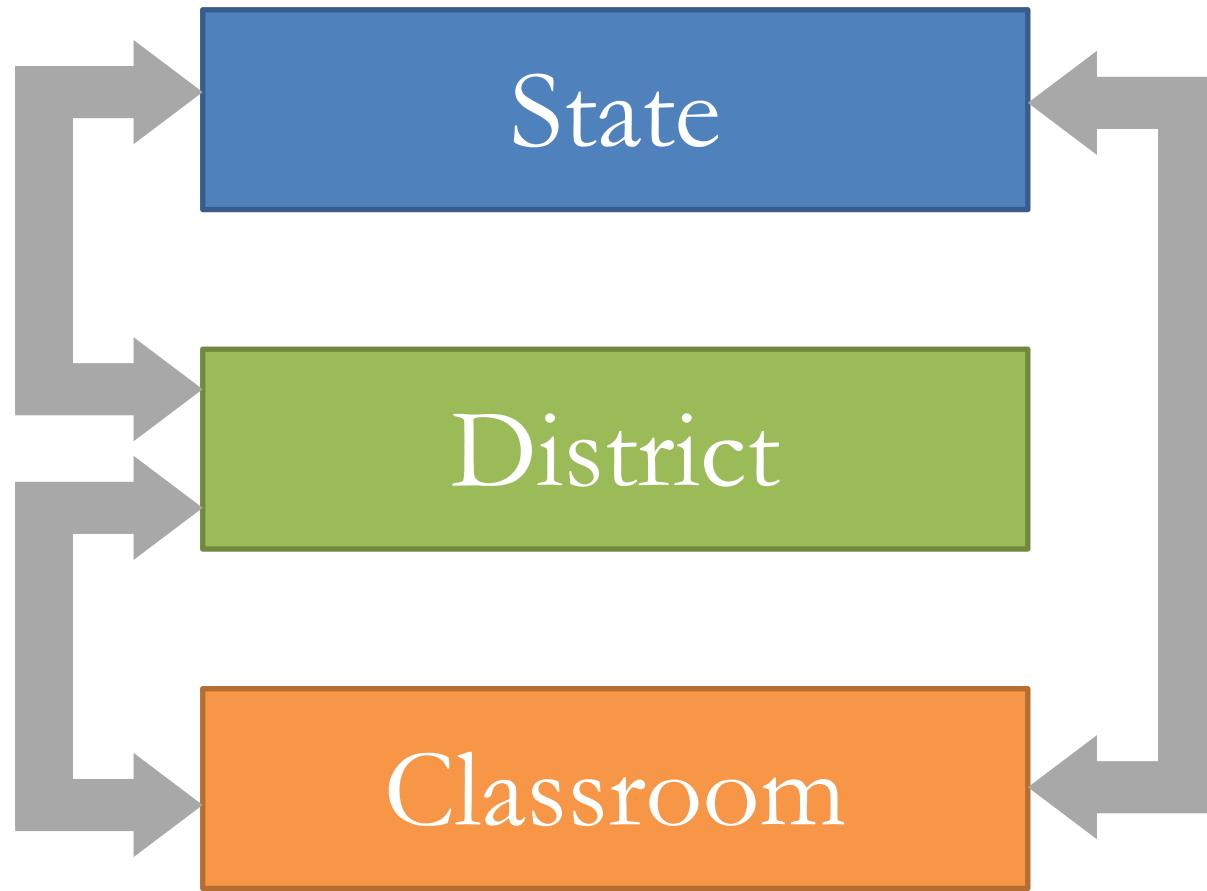
- There are many, many possible uses to be examined. Do we need to validate them all?
- How can we clearly delimit what is and isn't a system?
- How can we effectively firewall classroom and monitoring uses from one another?
- How can good systems designed be scaled?

# Who is using these results?

District

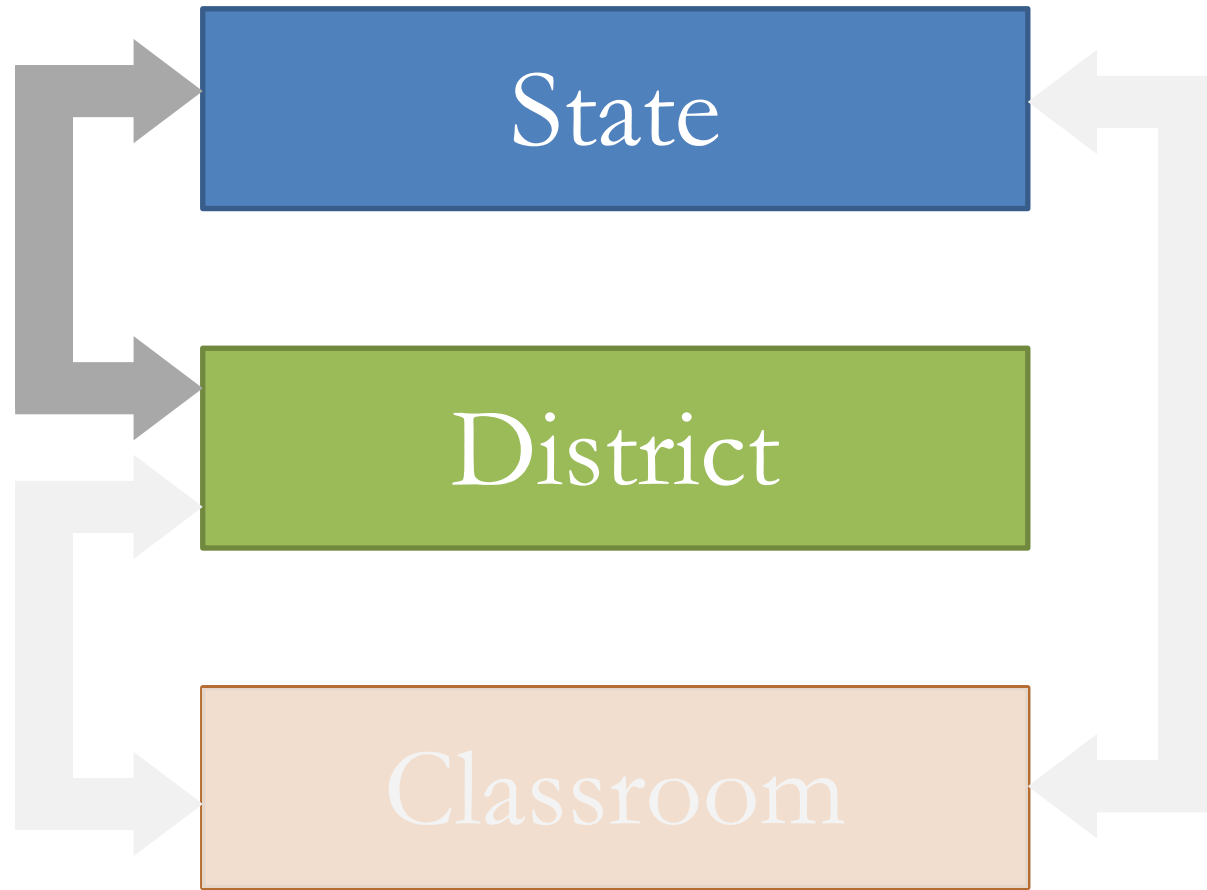
To start considering use, we first examine the district layer, then use this framing to open up discussion on how we can consider coupling along a number of dimensions (e.g., modularity, coverage).

# Coupling Across Layers



By doing so we consider *coupling* more broadly, with a focus on coupling between district and state levels.

# Coupling Across Layers



By doing so we consider *coupling* more broadly, with a focus on coupling between district and state levels.

# Starting within the District Level

- Considering the design of district-level, i.e., interim, assessments from the perspective of use.
  - I.e., focusing on a single level
- This was the basis for the *District Assessment System Design Toolkit*, which attempted to structure the large number of choices involved in designing or selecting district assessments.
  - Meant to relate the purposes of the assessment with a number of design choices

# The Core of the Toolkit

| Category                         | Complete Purpose/Use   | Abbreviated Purpose/Use  |
|----------------------------------|--|--|
| Signaling                        | Maintain a feedback loop between student and teacher to signal next steps<br>Indicate valued knowledge and skills to motivate instruction and student work   | Maintain a feedback loop<br>Indicate valued knowledge and skills   |
| Corroborate                      | Corroborate formative assessment insights to improve decisions and refine practice<br>Corroborate unit grades/test results to improve decisions and refine content/scoring<br>Corroborate marking period grades/test results to improve decisions and refine content/scoring   | Formative insights<br>Unit grades/test results<br>Marking-period grades/test results   |
| Inform instruction               | Monitor instructional effectiveness for in-the-moment adaptation and rapid course correction<br>Monitor student/group needs to differentiate and/or tailor next-lesson planning & instruction<br>Monitor student/group needs to differentiate and/or tailor next-unit planning & instruction   | In-the-moment adaptation<br>Next-lesson planning & instruction<br>Next-unit planning & instruction   |
| Inform instructional programming | Evaluate achievement to guide mid-marking-period instructional grouping (including remediation)<br>Evaluate achievement to guide instructional program placement (e.g., grade, course, track)  | Instructional grouping<br>Instructional program placement  |
| Grading                          | Evaluate achievement to support traditional grading<br>Evaluate achievement to support standards-based grading   | Support traditional grading<br>Support standards-based grading   |
| Eligibility                      | Evaluate achievement to determine eligibility for course credit (w/out taking the course)<br>Evaluate achievement to determine eligibility for program entrance (e.g., EL, SWD)<br>Evaluate achievement to determine eligibility for program exit (e.g., EL, SWD)<br>Evaluate achievement to determine eligibility for graduation/diploma annotation<br>Evaluate achievement to determine eligibility for formal honors/awards | Eligibility for course credit<br>Eligibility for program entrance/services<br>Eligibility for program exit<br>Eligibility for graduation/diploma annotation<br>Eligibility for honors/awards |
| Readiness                        | Evaluate achievement to determine readiness for the next grade or course<br>Evaluate achievement to determine academic readiness to begin college coursework<br>Evaluate achievement to determine academic readiness to begin career training coursework<br>Evaluate off/on/above track status for an outcome 2+ years out for planning/intervention   | For next grade/course<br>For college coursework<br>For career training coursework<br>On track for an outcome 2+ years out  |
| Programs & policies              | Identify needs to develop policies and/or programs<br>Track progress to refine and/or evaluate policies and/or programs  | Develop programs/policies<br>Refine/evaluate programs/policies   |
| Growth                           | Measure growth during a single marking period for accountability or policy/program evaluation<br>Measure growth across multiple marking periods for accountability or policy/program evaluation<br>Growth: Isolate school effects on student growth for educator evaluation<br>Growth: Isolate educator effects on student growth for educator evaluation  | During a single making period<br>Across multiple marking periods<br>Isolate school effects on student growth<br>Isolate educator effects on student growth                                   |

Note: this is way too much text to read! If interested, download the [toolkit](#).



# Revisiting the Dimensions

- Degree of modularity (a few minutes  $\Leftrightarrow$  all of K-12)
- Depth of coverage (deep on little  $\Leftrightarrow$  sample of much)
- Item/task types (selected response  $\Leftrightarrow$  extended projects)
- Timing (before  $\Leftrightarrow$  during  $\Leftrightarrow$  after a unit)
- Control over timing (teacher  $\Leftrightarrow$  state/vendor)
- Control over content (teacher  $\Leftrightarrow$  state/vendor)
- Security (regulated  $\Leftrightarrow$  open)

# Tying TOL and Purpose/Use to Location on each Dimension

- We think this is possible to do, for example...
- Dimension locations for informing daily instruction
  - Degree of modularity (a few minutes  $\Leftrightarrow$  all of K-12)
  - Depth of coverage (deep on little  $\Leftrightarrow$  sample of much)
  - Item/task types (depends on TOL and content covered)
  - Timing (before  $\Leftrightarrow$  during  $\Leftrightarrow$  after a unit)
  - Control over timing (teacher  $\Leftrightarrow$  state/vendor)
  - Control over content (teacher  $\Leftrightarrow$  state/vendor)

# Lessons Learned

- It is important to clearly define what constitutes a purpose and what constitutes a use
- It is helpful to define a use as an action taken using assessment data, where the action is paired with an object for the action.
- It is helpful to define a purpose as the reason for taking the action.

# Looming Questions

- How do we make this work tractable, particularly if we try to address the aforementioned dimensions at both the state and district levels?
  - Is it a matter of working around the state level?
- How can we provide recommendations or guidance for practice around coupling?

# Supplemental Slides

# A Thought Experiment

What would a “system” that is coherent, comprehensive and continuous (and also efficient and utilitarian) across state and district levels look like?

- How would a theory of learning unify the two levels? What does this theory need to define?
- Do some uses, and their designs, jeopardize this theory of learning?

# Designing Across State & District?

|                      | State | District |
|----------------------|-------|----------|
| Degree of modularity |       |          |
| Depth of coverage    |       |          |
| Item/task types      |       |          |
| Timing               |       |          |
| Control over timing  |       |          |
| Control over content |       |          |

# Some Example Designs

- A tightly coupled design: Delaware's NGSS Assessment System
  - Interim: Block design aligned to unit content
  - State Summative: A fixed design aligned to a subset of the domain, but with a focus on transfer
- A more loosely coupled design: Kentucky's NGSS Assessment System
  - Interim: Modular design (task bank) with tasks aligned to individual standards, likely at a deep level of complexity, with reporting requirements
  - State Summative: A fixed design broadly aligned to the standards