Contributing to the Development of *NGSS*-Based Large-Scale Assessments

Tom Regan

AAPT Summer 2014 Session AA03

NGSS Lead States. 2013. *Next Generation Science Standards: For States, By States*. Washington, DC: The National Academies Press.

My Perspective

- Science Assessment Developer
  - AAAS Project 2061—research related to *BSL*, *SFAA*, *NSES*
  - American Institutes for Research—large-scale state science assessments
  - Currently at The College Board
- HS Physics teacher
- Participated in AAPT review of *NGSS* 2nd Draft
- The views expressed here are my own.

Theme and Purpose

- Theme: Large-scale science assessment* would benefit from more physics community involvement

- Purpose
  - To suggest specific ways for the physics community to get involved in large-scale assessments
  - Issues of particular relevance to *NGSS* will be highlighted.

- *need a more precise term

Assessment Goals

- The assessment is fair.
  - Transparency (know what will be assessed, with sufficient time to prepare)
  - Matches what you teach (content and emphasis; content and student tasks)
- The assessment presents correct physics.
- The assessment provides useful information.
  - Information about important things
  - Sufficient granularity
  - Timeliness
  - Sufficient information
- The assessment is inexpensive
  - $
  - Time (preparation and testing)

Overview of Large-Scale Assessment Design Process

- State issues RFI (Request for Information) and RFP (Request for Proposal), evaluates proposals, and chooses a vendor
- SOW (Scope of Work) is written
- Item Specifications document is written.
- Test Specifications/Blueprint is written.

- Item Development Begins

RFI/RFP and Assessment Goals

RFI: Request for Information: opportunity to learn about vendor capabilities
RFP: Request for Proposal: opportunity to shape the contract

- Usefulness—ensure the vendor can assess what you think is important, and can get the most information from the student responses
- Correctness—ask vendor to demonstrate subject-matter competence
- Inexpensiveness—watch for construct-irrelevant bells & whistles

RFI/RFP and Usefulness

- Sample Questions:
  - "How will you assess each of the science practices?"
  - "How will you assess the student's facility with common scientific representations (e.g. free-body diagrams, ray tracing)?"
  - "How will you extract diagnostic information from student responses, including constructed-response items?"
  - "Please describe a test blueprint that is suited to assessing the *NGSS*."

- You do not have to be familiar with large-scale assessment—simply ask the vendor to tell you what they can do.

RFI/RFP and Correctness

- First decide what "subject-matter competence" means to you:
  - Expertise: Physics or Physical Science or High-School Science or Science
  - Expertise: Science Content or Science Practices
  - Experience: Teaching or Assessment or Research

- Then ask the vendor to demonstrate those characteristics.

- Large-Scale Assessment (LSA) Context Note: ELA and Math assessment are not strongly discipline-specific.

RFI/RFP and Inexpensiveness

- Expensive elements of the assessment system should be worth the expense.
- Vendor should articulate rationale for such elements.
- Example: simulations/animations are expensive in terms of development time and cost, possibly bandwidth, possibly test-taking time.
- “What information will be provided by simulations/animations that cannot be provided by cheaper item types?”

Scope of Work and Assessment Goals

Scope of Work: Detailed description of what vendor and client will do over the life of the contract.

- Correctness—ensure that subject-matter competence is applied at appropriate points in the development process
- Usefulness—ensure that useful information is reported.

Scope of Work and Correctness

- Vendor internal item development
  - Specify the roles of subject-matter-experts versus non-experts.
- State internal item review
  - Volunteer to assist.
- Content Committee Review
  - Specify subject-matter-expert representation
  - Discipline-based committees (“Physics” or “Physical Science”) versus Grade-level-based committees (“High School Science”)
- LSA Context note: Writing/negotiating the SOW is laborious, so it is sometimes assumed that what works for ELA and/or Math will work for Science.

Scope of Work and Usefulness

- Ensure that diagnostic information is built into the system:

"Misconceptions will be built into the item bank and multiple-choice item distractors will be linked to the misconceptions."

Item Specifications and Assessment Goals

The Item Specifications document describes the desired items.

- Learning goals to target
- Item types
- DOK/CD

Often it is an augmented version of the standards (the *NGSS*)

- Fairness (transparency)—ensure timely availability of document
- Correctness—fix any errors in the standards
- Usefulness—if you want information about it, there's got to be a standard.

Item Specifications and Fairness

- Typical standards documents (and therefore item specs) are insufficiently detailed, so it's not clear what will be assessed.

- Stuff the Item Specifications with details!
  - List all assessable forces.
  - Describe the assessable physical situations
  - Add necessary definitions like "model" "explanation" "evidence"—*SCBSCS*[*] is a good source.

- LSA Context Note: "e.g." is very common in standards documents. Replace it with "i.e." and complete the list!
- LSA Context Note: There are differing opinions about making Item Specifications public.

[*]The College Board, 2009. *Science College Board Standards for College Success*.

Item Specifications and Correctness

- Use the Item Specifications to shore up content weaknesses in the standards.
- Strategy: Add “Testing Points” to every Performance Expectation
- The Testing Points (1) state content correctly and (2) add specificity (improving Fairness)
- State that assessment items will be based on the Testing Points.

Item Specifications and Testing Points (1)

NGSS Performance Expectation HS-PS2-5

Plan and conduct an investigation to provide evidence that an electric current can produce a magnetic field and that a changing magnetic field can produce an electric current.

Assessment Boundary: Assessment is limited to designing and conducting investigations with provided materials and tools.

- "Can" is not appropriate.
- "limited to designing and conducting"—why not also analyze data?
- What "materials and tools" are acceptable?

Item Specifications and Testing Points (2)

NGSS

Sample Testing Points for PE HS-PS2-5

- Assessable student activities
  - Planning investigations
  - Conducting investigations
  - Analyzing the data from investigations.
- Assessable content
  - An electric current produces a magnetic field.
  - A changing magnetic field produces an electric current.
- Assessable materials and tools:
  - Batteries and power supplies
  - Wires, wire loops, and solenoids
  - Permanent magnets and electromagnets
  - Ammeters, voltmeters, gaussmeters, magnetic compasses

Item Specifications and Usefulness

- Any new Performance Expectations (see next slides) that are written will go in the Item Specifications
- If it's not in the Item Specifications, it won't be assessed!

Test Specifications/Blueprint and Assessment Goals

The Test Specifications/Blueprint describes the test.

- Total # of items
- # of items per Learning Goal or LG Group
- # of items per other categories

It expresses priorities.

- Usefulness—reporting categories, granularity…
- We'll discuss this with reference to a sample blueprint.

NGSS-Based Test Blueprint

- A general strategy for *NGSS*-Based Blueprints
  - Define the *test* in terms of the *Framework*[*]
  - Define the *items* in terms of the *NGSS*

- Say your priorities are
  - Science content
  - Science practices
  - *not* CCC, DOK…

- Pieces of sample "High School Physical Science" Blueprint
  - *Framework* Core and Component Ideas in the Physical Sciences
  - *Framework* Practices for K-12 Science Classrooms
  - *NGSS* High School Physical Sciences Performance Expectations (secondary connections not included)

National Research Council. 2011. *A Framework for K-12 Science Education: Practices, Crosscutting Concepts, and Core Ideas.* Washington, DC: The National Academies Press.

Sample *NGSS*-Based Test Blueprint

NGSS

Disciplinary Core Ideas (total 4) ↓

Component Ideas (total 13) ↓

Science and Engineering Practices (total 8) →

Disciplinary Core Ideas (total 4)	Component Ideas (total 13)	1, Questions	2, Models	3, Investigations
PS1, Matter and Its Interactions	PS1.A, Structure and Properties of Matter		HS-PS1-1 HS-PS1-4	HS-PS1-3
	PS1.B, Chemical Reactions		HS-PS1-4	
	PS1.C, Nuclear Processes		HS-PS1-8	
PS2, Motion and Stability…	PS2.A, Forces and Motion			
	PS2.B, Types of Interactions			HS-PS2-5
	PS2.C, Stability and Instability in Physical Systems			

- Upper left corner of blueprint only
- Note "holes" where *NGSS* did not write PEs

hole

Performance Expectation (total 24)

Possible Implementations of Blueprint Structure

- Report a subscore for every NGSS-populated cell
  - 28 populated cells @ 10 points per reporting category → 280 points @1min/point → Almost 5 hours. Expensive!
  - Blueprint has many holes
    - Usefulness/Fairness: Is it an adequate representation of the curriculum?
    - Subscore validity: Some Practices are assessed with only one of the thirteen Component Ideas.

- Report a subscore for every cell
  - 8 practices × 13 component ideas = 104 cells → 1040 points → 17 hours. Extremely expensive!
  - Requires 78 new Performance Expectations to fill the empty cells.

Suggested Implementation of Blueprint Structure

NGSS

- Report on every column (Science Practice) and every row (Component Idea)
  - Need 10 points per Science Practice
  - Need 10 points per Component Idea

- Write additional PEs to bolster subscore validity. 24 new PEs can ensure that
  - Each Component Idea subscore includes at least one item from at least four of the eight Science Practices
  - Each Science Practice subscore includes at least one item from each of the four Core Ideas.
  - Position new PEs such that items 'count double'

- 130 items → 2+ hour exam. Long but perhaps reasonable for year-end.

Conclusion

- I hope this presentation has given you some starting points for contributing to large-scale assessments based on the *NGSS*.
- Decide what you want and ask for it.
- Nothing ventured, nothing gained.