

# Considerations for the Alabama State Assessment System

Juan D'Brot, Ph.D. and Scott F. Marion, Ph.D.

*National Center for the Improvement of Educational Assessment*

Presentation to Alabama Assessment  
Policy Leaders and Stakeholders

# Overview of discussion

- A shared understanding of key assessment concepts
- Goals, purposes, and uses of assessment data
- What does it take to change assessments?
- Potential approaches for getting from here to there

# National Center for the Improvement of Educational Assessment (aka The Center for Assessment)

- A NH-based non-profit technical & policy consulting firm established in 1998 with the mission of improving student learning through improved assessment and accountability practices
- Current contracts with 30+ states, districts, and other entities
  - Almost all are long-term contracts designed to provide technical and design support for a range of assessment and accountability issues
  - Including many of your neighbors—Louisiana, Georgia, Arkansas, North Carolina
- Purposely small—12 full-time professionals
  - All with doctoral degrees and almost all have worked in the “real world” of state assessment and/or as assessment contractor staff.
- Non-partisan and independent of any governmental agency or testing company

# The challenge of assessment design

We want an assessment that:

- Provides information useful for evaluating programs and interventions
- Provides information for improving teaching and learning
- Provides high-quality data for fair accountability
- Can be administered in one class period
- Is administered during the last week of school
- Can deliver results at least a month before school gets out
- Is inexpensive

**Pick one!**

# Goals for An Assessment

- *Spend 5 minutes **working alone***, write down your highest priority purposes and uses for assessment results. In other words, what uses do you want the assessment to be able to support?
- *Spend 10 minutes in **groups of 3-4*** from those sitting around you and discuss what each of you has written down to search for commonalities. See if you can **eliminate** any purposes and uses.

# Common uses of assessments

- Student Level
  - Measure Achievement
  - Measure strengths and weakness
  - Make individual student decisions
- School Level
  - Accountability
  - Teacher evaluation
  - Program evaluation
- District & State Level
  - Accountability
  - Program Evaluation
  - Comparisons

# Common uses of assessments

- Student Level
  - Measure Achievement
  - Measure strengths and weakness
  - Make individual student decisions

---

- School Level
  - Accountability
  - Teacher evaluation
  - Program evaluation
- District & State Level
  - Accountability
  - Program Evaluation
  - Comparisons

# Common uses of assessments

- Student Level
  - Measure Achievement
  - Measure strengths and weakness
  - Make individual student decisions

---

- School Level
  - Accountability
  - Teacher evaluation
  - Program evaluation
- District & State Level
  - Accountability
  - Program Evaluation
  - Comparisons

**What do you notice about example uses above and below the dotted line?**



# Common uses of assessments

- Student Level
  - Measure Achievement
  - Measure strengths and weakness
  - Make individual student decisions

---

- School Level
  - Accountability
  - Teacher evaluation
  - Program evaluation
- District & State Level
  - Accountability
  - Program Evaluation
  - Comparisons

**What do you notice about  
example uses above and  
below the dotted line?  
Process vs. Outcome**

# A Call for Assessment Systems...

- The differing purposes and intended uses of large-scale and classroom level assessments make clear that **different assessments** are needed
  - standardized vs. dynamic/flexible
  - uniform vs. variable dates
  - independent vs. assisted (scaffolded) performance
  - delayed vs. immediate feedback
  - stringent requirements for technical accuracy vs. less stringent requirements

# Balanced Assessment Systems to Serve Multiple Purposes



- Since *Knowing What Students Know* (Pellegrino, et al., 2001), we've seen increasing calls for Balanced Assessment Systems
  - Coherent
    - Based on a common vision and theory of learning
  - Comprehensive
  - Continuous

Assessment systems designed to serve multiple purposes require thoughtful planning about which data will be privileged at each level (Chattergoon & Marion, 2016).

# Focus of today's discussion



- **Assessment systems** are one way to try to address the multiple purposes for multiple users

This may be the focus of subsequent discussions, but **today** we need to attend to the pressing issue of discussing how to design and implement Alabama's **next statewide summative assessment**

# How long does it take to change a state test?

It always depends...

1. What use do you want to prioritize?
2. How big do you want your test to be?
  - a. An assessment system
  - b. A shorter end-of-year assessment
3. What do you want to test?

# How long does it take to change a state test?

It always depends--Do you want:

- a. A test created specifically for Alabama?
  - b. A “shelf test” (e.g., ACT Aspire or Measured Progress Empower) that is not designed specifically for your standards?
  - c. A consortium test such as PARCC or Smarter Balanced
- 
- If you want an Alabama-standards-based test...

# An Example of a Test Development Schedule

2017-2018

- Conceptualize the new assessment system
- Draft and release RFP
- Hire contractor

2018-2019

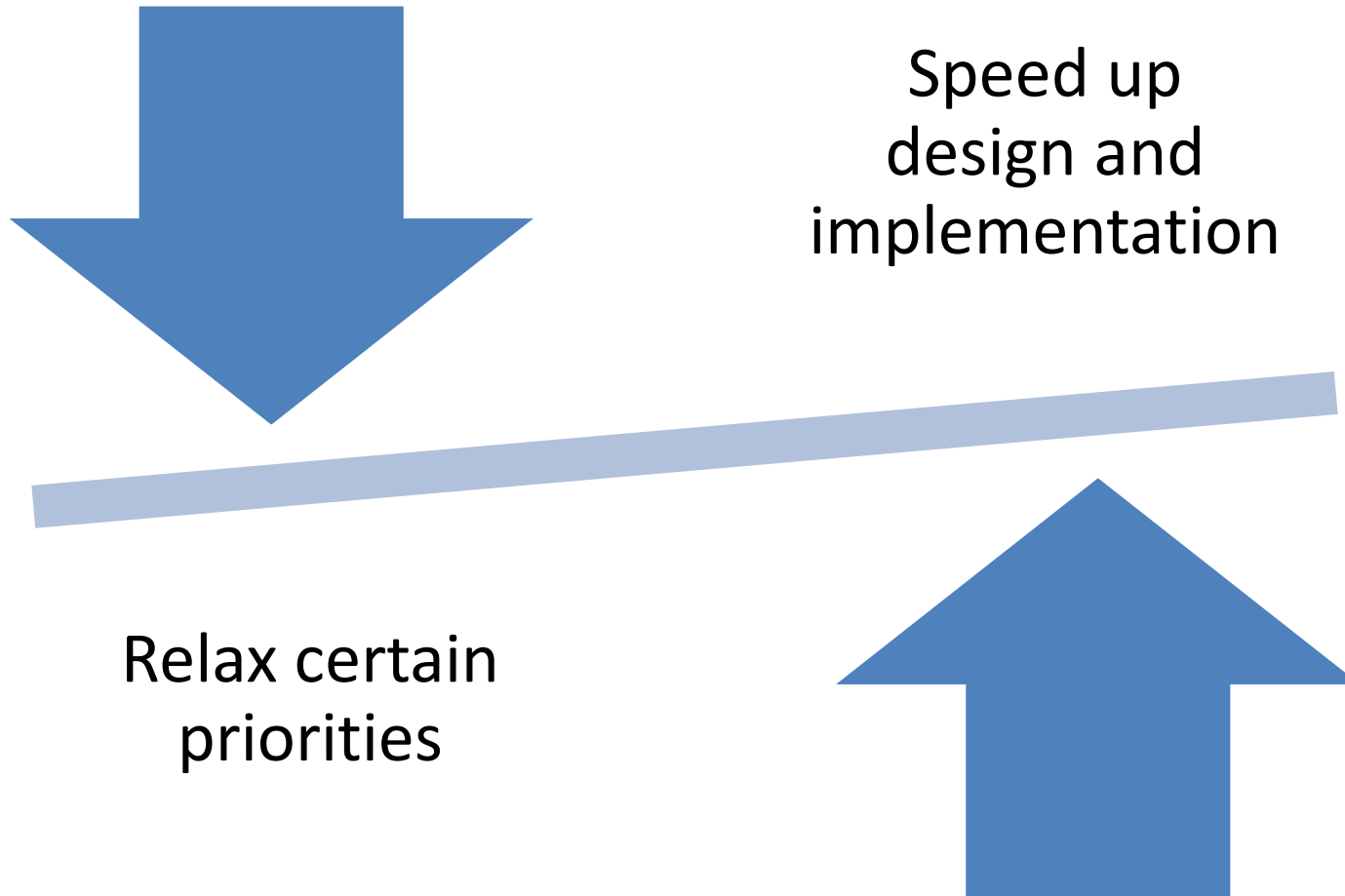
- Engage in item development, repurposing existing items and writing new items
- Engage in item tryouts and cognitive laboratories and field test items on the 2019 legacy test if possible

2019-2020

- Continue item development, start building operational “forms”
- First administration of new operational test, standard setting and score reporting

# Can you speed this up?

- Sure, but what are you willing to give up?

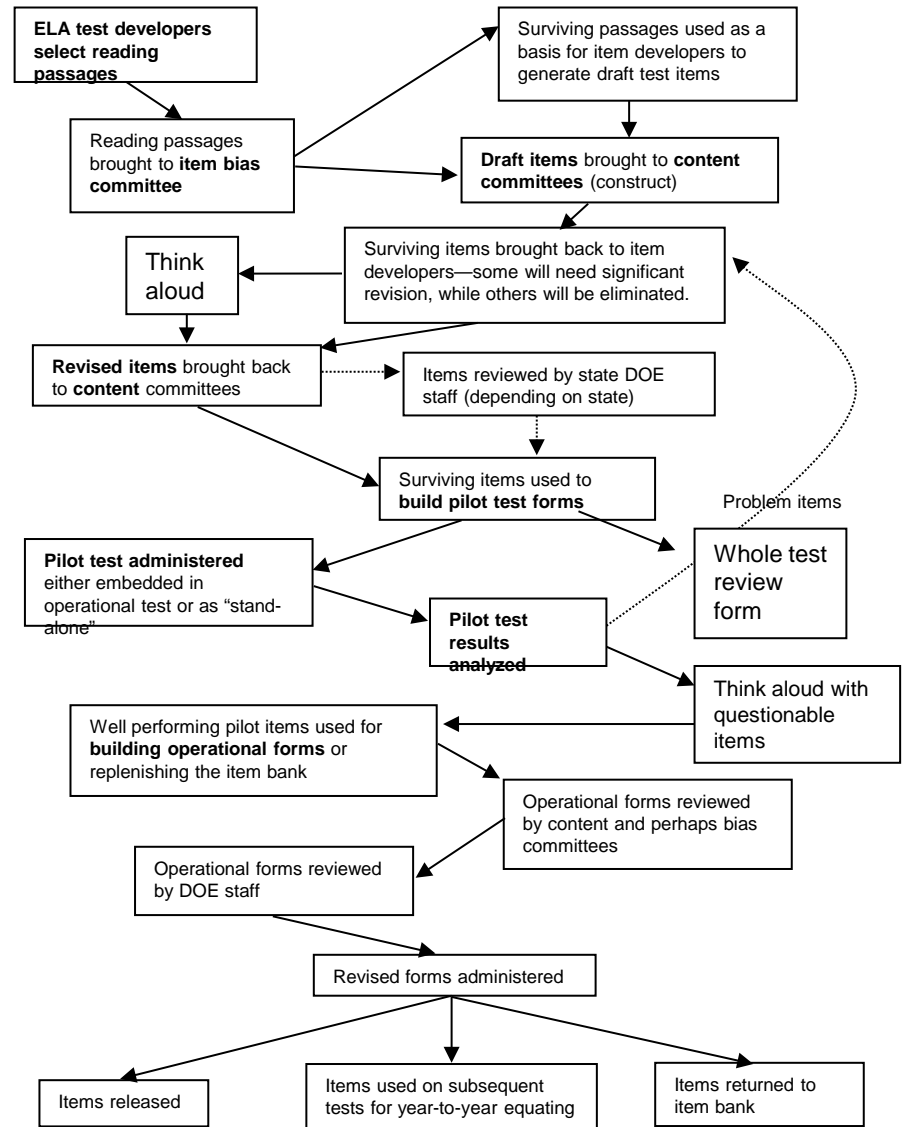




# The Life Cycle of an ELA Test item

Why it takes so long!

This schematic illustrates the many steps involved in developing a test item for an operational test form. Believe it or not, this is actually a bit of an oversimplification.



# Constraints and Requirements

- Assessment design is always a case of **optimization under constraints**
- The Alabama assessment system **must** support both state and federal **accountability** requirements
- Any given assessment can serve **only a very limited number of purposes** well and they must be **validated** for **each purpose**
  - For example, just because an assessment may be valid for predicting college readiness, it does not hold true that it is valid for informing instruction or even measuring achievement against the state's content standards.

# What's important to you?

What are your highest priorities for AL's summative assessment?

1. Subscores in addition to total scores
2. High levels of technical quality
3. Useful for measuring student growth
4. As short (least amount of time) as possible
5. Full coverage of Alabama's learning standards on state summative test
6. Use of performance or other open-ended tasks
7. Adapts to different levels of student achievement (i.e., computer-adaptive test)
8. Is coherently designed as part of an assessment system (e.g., interim assessments are also included)
9. As inexpensive as possible

**Take a few minutes to identify your top three priorities..**

# The importance of stability

- State assessments serve to monitor achievement over time (i.e., trend Lines) and to serve as a foundation of most accountability systems
- Any change to the assessment can potentially impact the ability to maintain valid achievement trend lines and to the accountability system
  - Administration policies
  - Content standards
  - Test length
  - Test composition

If you want to  
measure change...  
don't change  
the measure

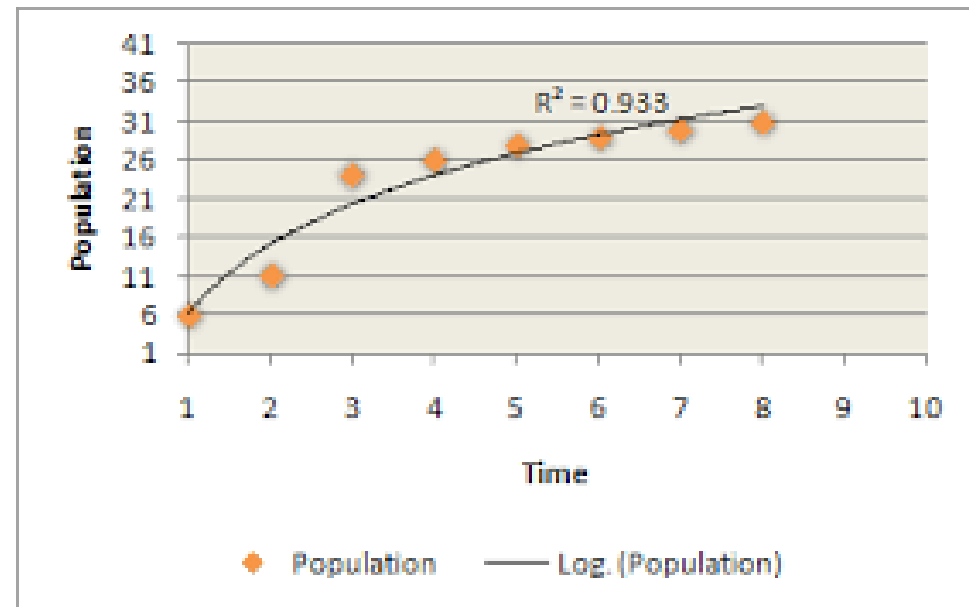
If you want to  
measure change...  
don't change  
the measure

**But what if you do?**

# Algebra is our friend (to a point!)

- First, it all depends how similar the two assessments are to one another
- Fortunately, we have a lot of experience helping states transition accountability systems across changes in assessment systems

We are also able to provide a concordance to help compare achievement results across the transition, but this still is a challenge for examining long-term trends



# How much stability can we expect?

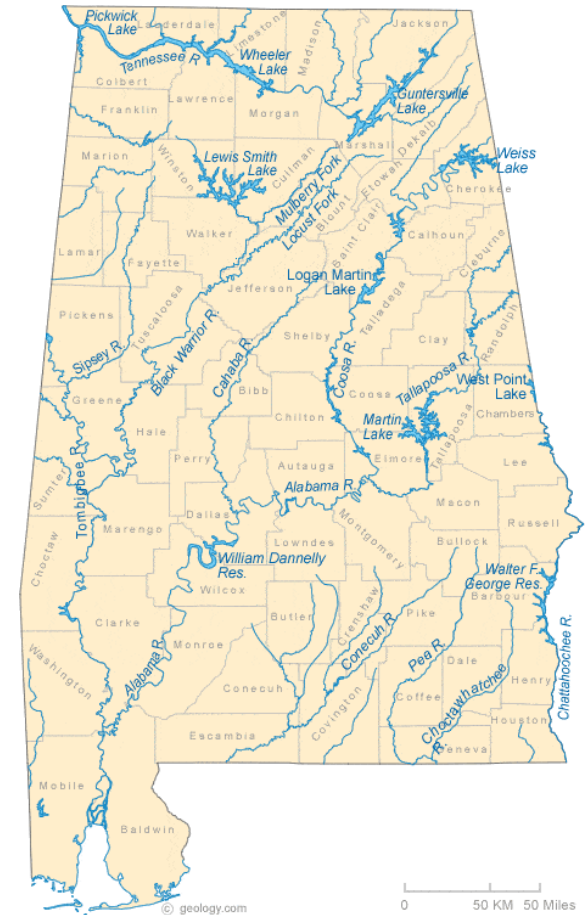


- Many states are changing their tests far too frequently
- In spite of our algebraic prowess, stability is our goal!
- Massachusetts gets credit for its rapid improvements and high national and international rankings, but we contend that the real “Massachusetts Miracle” was keeping the same standards and assessments for over 17 years
- If it takes between 5-10 years for policy initiatives to take hold in the classroom what happens if you change your test every 3-5 years?



# Getting from here to there...

- As we have tried to make clear, there are many decisions and tradeoffs associated with determining Alabama's next state assessment system
- There is not a single way to move from the current position to a new operational state assessment



# Facilitating change

- Requests for Proposals (RFP) operationalize the change from assessments to assessments (or vendors to vendors)
- RFPs communicate state values, priorities, and expectations for a new assessment system
- Transparency has led to greater scrutiny of assessment RFPs
- Diverse perspectives in an RFP can be beneficial but may be difficult to manage
- The best RFPs clearly specify the state's goals and intended outcomes to allow bidders to propose the most efficient solutions for the state

# Three basic frameworks for development

- DOE drafts and releases RFP for new assessment system (perhaps with State Board input)
- DOE convenes a small advisory group to provide input on the draft RFP
- State Board appoints an “assessment task force” comprised of a broad-based group of constituents to make recommendation to the DOE for constructing an RFP
- Many hybrids of these three...

# What do you think?

- In groups of 3-4, list the pros and cons of each of the three approaches on the previous slide (or a hybrid).
- Consider some or all of the following as either pros or cons:
  - Credibility
  - Defensibility
  - Quality
  - Efficiency
  - Cost
  - Time

# Questions, comments, and discussion

- We now have time for questions and responses for items that we either have not discussed or that we have but you'd like to circle back

# More Information:

Center for Assessment

[www.nciea.org](http://www.nciea.org)



Juan D'Brot and Scott Marion

[jdbrot@nciea.org](mailto:jdbrot@nciea.org) and [smarion@nciea.org](mailto:smarion@nciea.org)