

Standardized Testing: Mend It, Don't End It

**Shanker Institute Forum, “Quality Assessments for Educational
Excellence,” Washington, DC**

November 13, 2013

Steven Glazerman

MATHEMATICA
Policy Research

The Testing Debate

A blue speech bubble with a white border, containing the text "No excuses!".

**No
excuses!**

■ Reformers:

- Test all grades and subjects every year
- Track students over time
- Report results and use for accountability

A blue speech bubble with a white border, containing the text "Stop the madness!".

**Stop the
madness!**

■ Critics:

- Capture critical thinking, writing, more subjects
- End high stakes
- Don't use tests for teacher evaluation
- Boycott tests

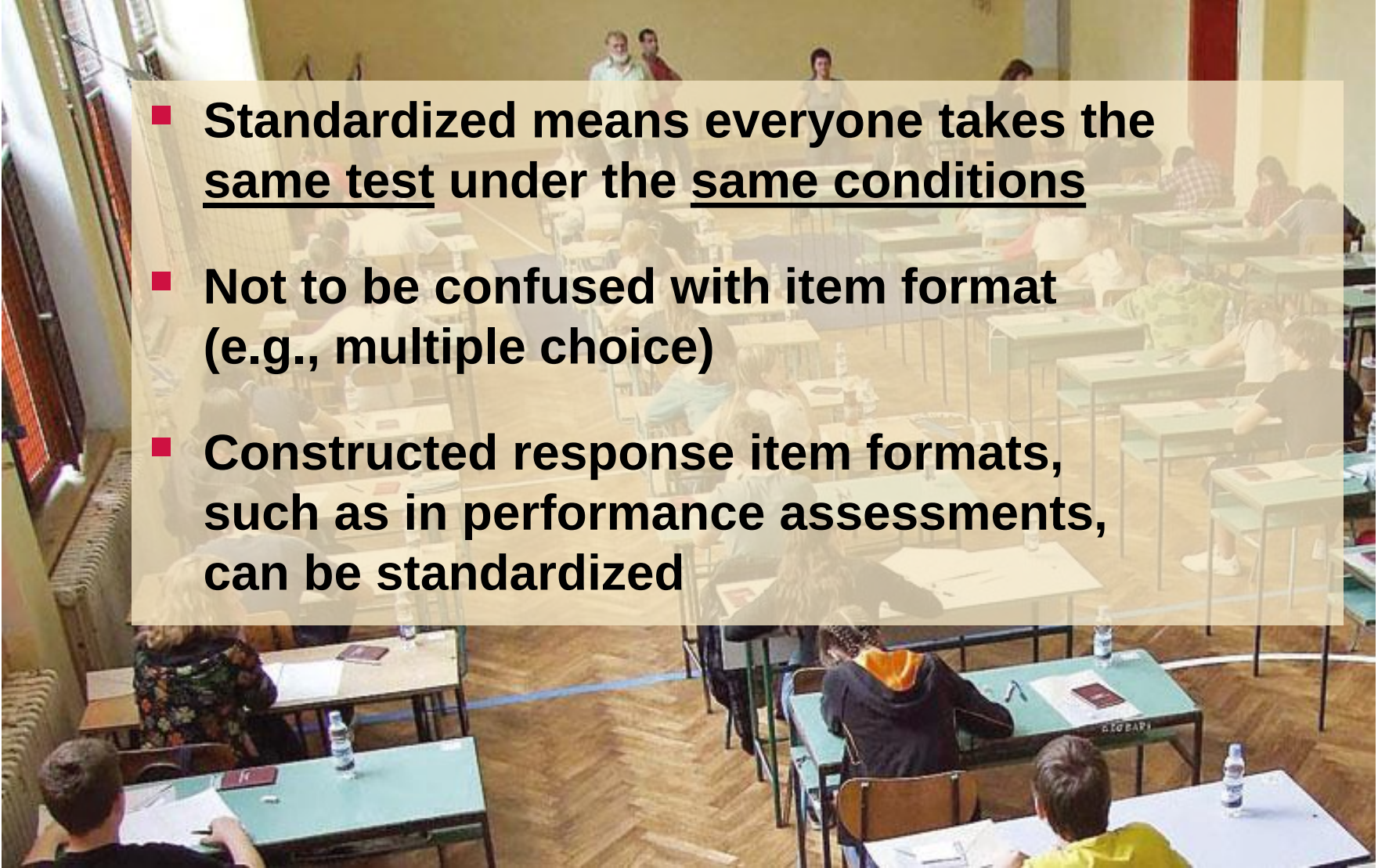
Researcher perspective:

Need outcomes to evaluate what works.

How to Make Progress?

- **Clarify terms to get on the same page**
- **Set criteria for good assessments**
- **Monitor the test vendors and demand the best**

1. Standardized Testing

- 
- **Standardized means everyone takes the same test under the same conditions**
 - **Not to be confused with item format (e.g., multiple choice)**
 - **Constructed response item formats, such as in performance assessments, can be standardized**

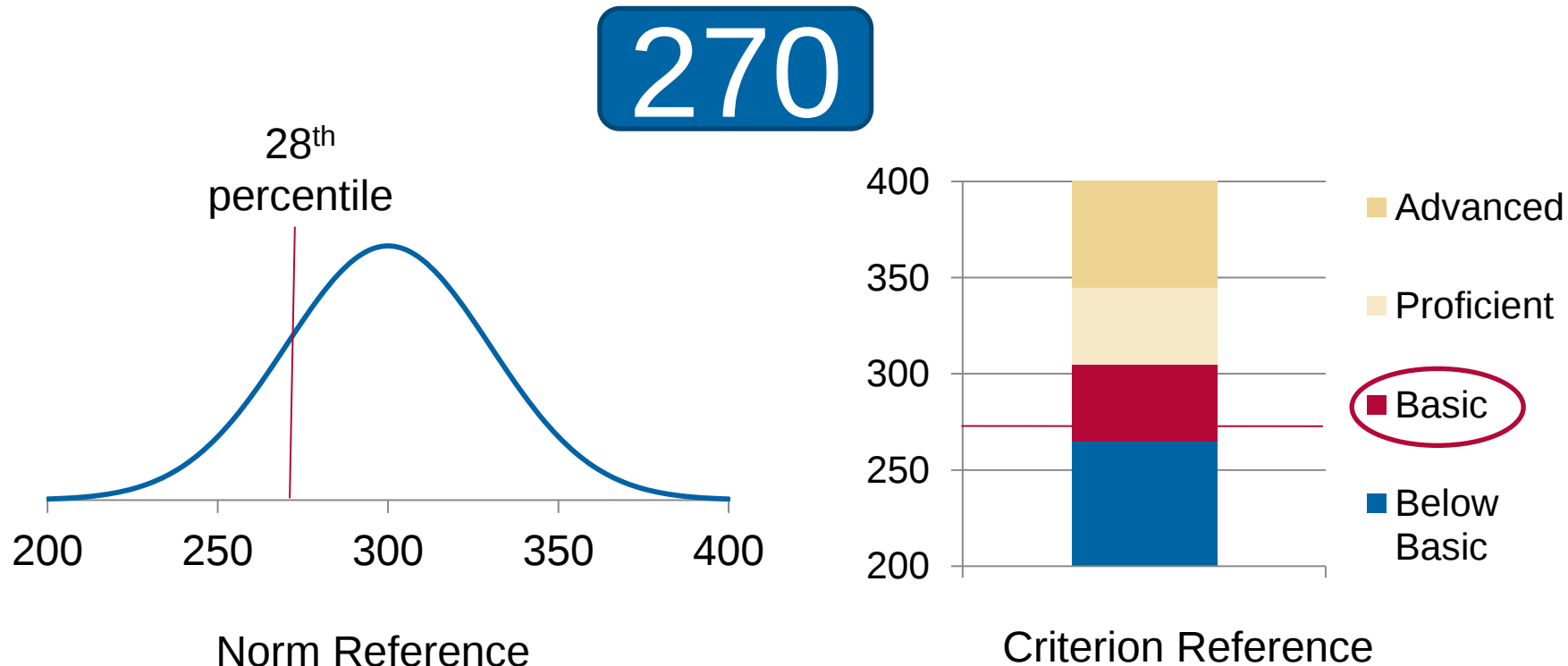
2. Teaching to the Test

- Teaching *narrowly* to the test
 - Emphasize test-taking strategy over content
- Teaching to the domain of the test
 - Test items should be a random sample of items, testing all possible elements of course content
 - Teachers should cover all content from which the test items are sampled



3. Norm-Referenced Tests

- “Norm-referenced” and “criterion-referenced” are not properties of tests but of test *interpretation*



4. High-Stakes Tests

- **High-stakes *decisions* are being made**
 - Teacher hiring, promotion, assignment, compensation, remediation
 - Education interventions, sometimes costly
 - What information are we using?
 - Imperfect information versus no information

- **What about the test produces the anxiety?**
 - Difficulty mismatch
 - Burdensome format
 - Time pressure

Properties of Good Assessments

- **Valid**

- Measures what we care about
- No bias (no “DIF”)

- **Reliable**

- Multiple scorers would give the same score
- Discriminating over the range of proficiency

- **Noncorruptible**

- Cheating can be detected, prevented
- Must elicit students’ best effort

- **Meaningful, with minimal burden**

- Students can learn from the test

- **Interpretable**

- Supports decision making to improve teaching and learning

Conclusions

- **Standardized testing doesn't have to be multiple choice, but standardization is not the problem**
- **If a test is aligned to the curriculum, teaching to it (broadly) would be a good thing**
- **It can be useful to compare test takers to each other as well as to absolute standards**
- **The cost of making tests better is less than the cost of making hard decisions with no evidence**