Tests & Measurements for Teachers
(Tiegs, 1931)

"***The principal function of measurement is to contribute directly or indirectly to the effectiveness of teaching and learning.***"

The Ψ of Accountability

GOAL:

Maximize effort, innovation, self-criticism, optimal decision making

OUTCOME:

Maximize self-justification, select options that are easy to justify

FINDING:

Increased effort is not sufficient

Our Accountability System: NCLB

- Some Positives
- Lots of Negatives stemming from ill-informed assumptions

Untested Assumptions

- It's simply a matter of accounting – of holding each school accountable
- Schools are equally resourced
- It's a matter of schooling, not family or community conditions.
- Every school has staff to meet its needs – they simply need appropriate motivation.
- Teacher experience is irrelevant.

(Gary Orfield, 2014)

More Assumptions…

- Research is irrelevant – since much of it is full of excuses (why things don't work).
- Segregated schools can be equal.
- Success in education can be sufficiently measured by test scores.
- Children from linguistically isolated communities require no special programming.
- Schools created outside a typical district are inherently better.

High School Reform: Lessons Learned

- Personalized learning environment for both students and teachers
- Rigorous and relevant instruction
- Social and academic supports for students
- Engaging students through connections to their world

(Russell Rumberger, 2012)

Need specialized training to

- understand and relate effectively to students of different cultures and languages
- manage multicultural classrooms
- create local and regional collaboration
- evaluate evidence needed when adopting popular-sounding reforms or programs
- use information effectively to support teaching and learning

(Gary Orfield, 2014)

A Role for Measurement & Assessment

National Academies (2009)

A test score is an **estimate** rather than an exact measure of what a person knows and can do. The items on any test are a sample from some larger universe of knowledge and skills, and scores for individual students are affected by the particular questions included.

National Academies (2009)

A student may have done better or worse on a different sample of questions. In addition, guessing, motivation, momentary distractions, and other factors introduce uncertainty into individual scores.

Measurement Error

Sampling Error

2013-2014 Technical Manual for Minnesota's Title I and Title III Assessments

Understanding Measurement Error

When interpreting test scores, it is important to remember that test scores contain some amount of measurement error. That is to say, test scores are not infallible measures of student characteristics… **measurement error must always be considered** when making score interpretations. (p. 80)

2013-2014 Technical Manual

Using Objective/Strand-Level Information

Strand or substrand level information can be useful as a preliminary survey to help identify skill areas in which further diagnosis is warranted. The standard error of measurement associated with these generally brief scales makes drawing inferences from them at the individual level very suspect; more confidence in inferences is gained when analyzing group averages. (p. 81)

2013-2014 Technical Manual

When considering data at the strand or substrand level, the error of measurement increases because the number of possible items is small. In order to provide comprehensive diagnostic data for each strand or substrand, the test would have to be prohibitively lengthened. (p. 81)

MCA for Individual Interpretation

2013-2014 Yearbook Tables for Minnesota's Title I and Title III Assessments

Example: Grade 3 Reading Score Distributions, p. 90

Scale Score	SEM	Achievement Level
338	5.0	D
339	5.0	D
340	5.0	P
341	5.0	P
…		
348	5.0	P
349	5.0	P
350	5.0	M
351	5.0	M

Seeking More Information

- Looking to Subscales for more info
- Knowing subscales are shorter – less precision
- Consider the role of measurement error in correlations
- Randomness doesn’t correlate with anything
- Measurement error (random noise) limits correlations

2014 MCA-III Summary Statistics Grade 3 Reading, p. 133

	Items	Reliability
Total Scale	48	.88
Literature	21-27	**.81**
Information	21-27	**.80**

2014 MCA-III Subscale Correlations Grade 3 Reading, p. 160

	Total Scale	Literature
Literature	.94	
Information	.93	**.80**

3rd Grade Mathematics Subscale Corrected Correlations

	Number & Operation	Algebra	Geometry & Measurement
Algebra	.98		
Geometry & Measurement	1.00	.99	
Data Analysis	.98	.97	.99

4th Grade Mathematics Subscale Corrected Correlations

	Number & Operation	Algebra	Geometry & Measurement
Algebra	.99		
Geometry & Measurement	.97	.98	
Data Analysis	.97	.99	.98

5th Grade Mathematics Subscale Corrected Correlations

	Number & Operation	Algebra	Geometry & Measurement
Algebra	1.00		
Geometry & Measurement	.96	.98	
Data Analysis	.97	.97	.95

6th Grade Mathematics Subscale Corrected Correlations

	Number & Operation	Algebra	Geometry & Measurement
Algebra	.99		
Geometry & Measurement	.99	.99	
Data Analysis	1.00	.97	.97

7th Grade Mathematics Subscale Corrected Correlations

	Number & Operation	Algebra	Geometry & Measurement
Algebra	.99		
Geometry & Measurement	.99	1.00	
Data Analysis	1.00	.99	.99

8th Grade Mathematics Subscale Corrected Correlations

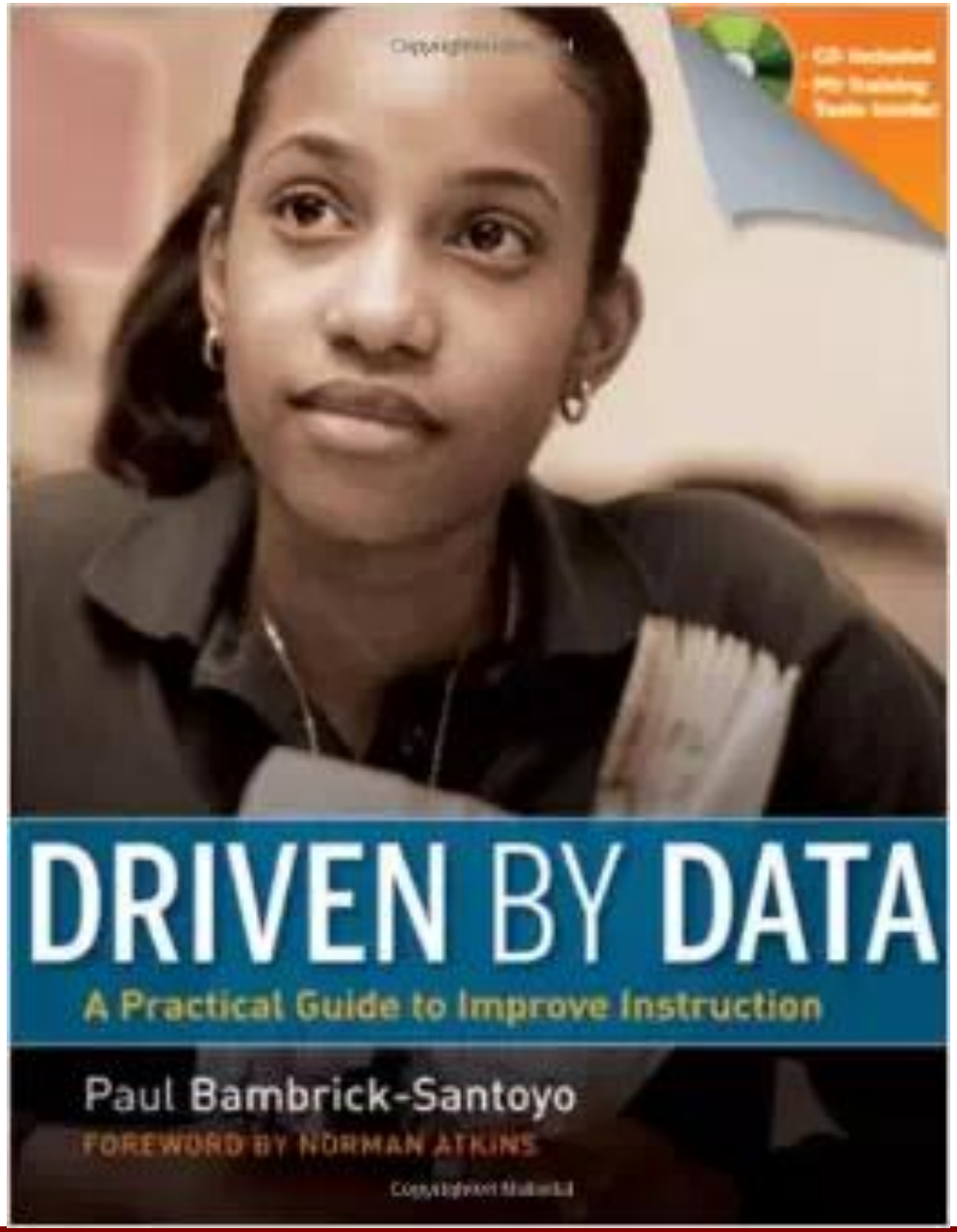
	Number & Operation	Algebra	Geometry & Measurement
Algebra	.99		
Geometry & Measurement	.96	.99	
Data Analysis	.96	1.00	.97

11th Grade Mathematics Subscale Corrected Correlations

	Algebra	Geometry & Measurement
Geometry & Measurement	1.00	
Data Analysis	.99	.99

Grade	Corr.
3	.99
4	.98
5	.97
6	.97
7	.96
8	.93
10	.97

MCA Reading Subscale Corrected Correlations: ***Literature & Information***


DRIVEN BY DATA
A Practical Guide to Improve Instruction
Paul Bambrick-Santoyo
FOREWORD BY NORMAN ATKINS

p. 8

CORE IDEA

- Assessments are not the end of the teaching and learning process; they're the starting point.

… we should not teach and then write an assessment to match; instead, we should create a rigorous and demanding test and then teach to meet its standards.

p. 13

CORE IDEAS: Interim Assessments

- Start from the end-goal exam.
- Align the interim assessments to the end-goal test.

…

p. 28

Analyze the Interim Assessment or End-Goal Test

Acquire the closest version that you can find of your state test, interim assessment, or other year-end assessment by which your students' learning will be measured.

…

National Academies 2009

The choice of appropriate assessments for use in instructional improvement systems is critical. Because of the extensive focus on large-scale, high-stakes, summative tests, policy makers and educators sometimes mistakenly believe that such tests are appropriate to use to provide rapid feedback to guide instruction. This is not the case.

National Academies 2009

Tests that mimic the structure of large-scale, high-stakes, summative tests, which lightly sample broad domains of content taught over an extended period of time, are unlikely to provide the kind of fine-grained, diagnostic information that teachers need to guide their day-to-day instructional decisions.

National Academies 2009

…BOTA urges the Department to clarify that assessments that simply reproduce the formats of large-scale, highstakes, summative tests are not sufficient for instructional improvement systems.