



NextGen Assessment and Accountability:

Innovations in Policy and Practice

Michigan Educational Research Association

21 November 2016

Gregory J. Cizek

Guy B. Phillips Distinguished Professor of Educational Measurement and Evaluation
School of Education, University of North Carolina-Chapel Hill
cizek@unc.edu



Overview

- I) Introduction
- II) What We Can/Should Do (7)
- III) What We Can't/Shouldn't Do (3)
- IV) Conclusions, Questions



Broad Support...

- * NAACP
- * League of United Latin American Citizens (LULAC)
- * National Council of La Raza
- * National Disability Rights Network
- * The Leadership Conference on Civil and Human Rights
- * The American Association of University Women
- * Association of University Centers on Disabilities
- * Council of Parent Attorneys and Advocates
- * Disability Rights Education and Defense Fund
- * National Urban League
- * Southeast Asia Resource Action Center
- * TASH (The Association for Persons with Severe Handicaps)



Broad Support

“We Oppose Anti-Testing Efforts”

“Participation in assessments [is] critical for expanding educational opportunity for all students”

“anti-testing efforts ... sabotage important data and rob us of the right to know how our students are faring”

(May 5, 2015)



What We Can/Should Do



What We Can/Should Do

1) Score Reporting

STANFORD

ACHIEVEMENT TEST SERIES

TEACHER: JOHN WILLIAMS

SCHOOL: **B** LAKESIDE ELEMENTARY

DISTRICT: NEWTOWN

GRADE: 4
TEST DATE: 4/88

C
NORMS:
NATIONAL
LEVEL:
FORM:

STANFORD
GRADE 4
SPRING
INTER 1
J

OLSAT
GRADE 4
SPRING
E
1

STUDENT SKILLS ANALYSIS

A BRIAN ELLIOTT

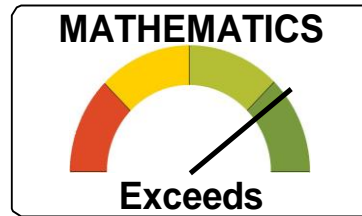
TESTS	NO. OF ITEMS	RAW SCORE	SCALED SCORE	NATL PR-S	LOC PR-S	GRADE EQUIV	AAC RANGE	NATIONAL GRADE PERCENTILE BANDS							
								1	10	30	50	70	90	99	
Total Reading	98	70	143	50-5	45-5	4.8	LOW								
Vocabulary	40	27	148	52-5	46-5	5.1	LOW								
Reading Comp.	54	43	143	50-5	44-5	4.8	LOW								
Total Math	118	58	143	44-5	45-5	4.6	LOW								
Concepts of No. Computation	34	14	137	32-4	26-4	3.8	LOW								
Applications	44	22	145	50-5	56-5	4.8	MIDDLE								
Total Language	60	47	149	61-6	56-5	5.4	MIDDLE								
Lang. Mechanics	30	21	148	60-6	54-5	5.4	MIDDLE								
Lang. Expression	30	26	150	62-6	58-5	5.4	MIDDLE								
Spelling	40	22	143	40-5	34-4	4.0	LOW								
Study Skills	30	21	142	50-5	44-5	4.8	LOW								
Science	50	31	139	44-5	38-4	4.6	LOW								
Social Science	50	42	161	82-7	76-6	6.2	MIDDLE								
Listening	45	25	130	49-5	45-5	4.7	LOW								
Using Information	70	35	152	59-5	55-5	5.2	MIDDLE								
Thinking Skills	100	45	146	52-5	53-5	4.9	LOW								
Basic Battery	387	243	143	50-5	45-5	4.8	LOW								
Complete Battery	487	315	145	52-5	47-5	4.9	LOW								
OTIS-LENNON SCHOOL ABILITY TEST		RAW SCORE	SAT	AGE PR-S	AGE NCE	SCALED SCORE	GRADE PR-S	GRADE NCE							
Total	72	53	116	84-7	70-9	290	82-7	69-3							
Verbal	36	27	117	85-7	71-8	295	83-7	70-1							
Nonverbal	36	26	115	82-7	69-3	285	80-7	67-7							

E

D

F

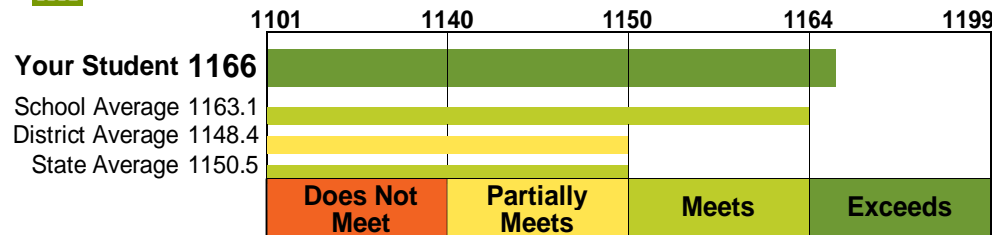
These are your student's results from Minnesota Comprehensive Assessments taken in the spring of 2016.

**STANDARDS**

Paul's score shows performance **above** grade 11 standards

STATE PERCENTILE RANK

Paul scored higher than **86%** of Minnesota grade 11 students

**MATHEMATICS: PAUL'S OVERALL MCA-III RESULTS****Performance Level Description:**

Students at the **Exceeds the Standards** level exceed the mathematics skills of the Minnesota Academic Standards.

Mathematics Performance Details

Your student's score is compared to the state expectations for mathematics and its content areas. Performance in content areas is reported as Below Expectations, At or Near Expectations, or Above Expectations.

= Below Expectations

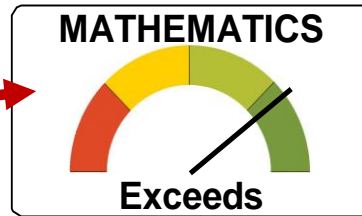
= At or Near Expectations

= Above Expectations

MATHEMATICS AREA	DESCRIPTION	PERFORMANCE
Algebra:	Identify features of functions and use them to solve real-world and mathematical problems, generate equivalent expressions, and solve equations and inequalities.	At or Near Expectations
Geometry and Measurement:	Calculate measurements, construct logical arguments to prove results, and apply properties of figures to solve problems.	Above Expectations
Data Analysis and Probability:	Display and analyze data; use various measures to draw conclusions, make predictions, and calculate probabilities.	Above Expectations

These are your student's results from Minnesota Comprehensive Assessments taken in the spring of 2016.

My
student
is OK.



STANDARDS

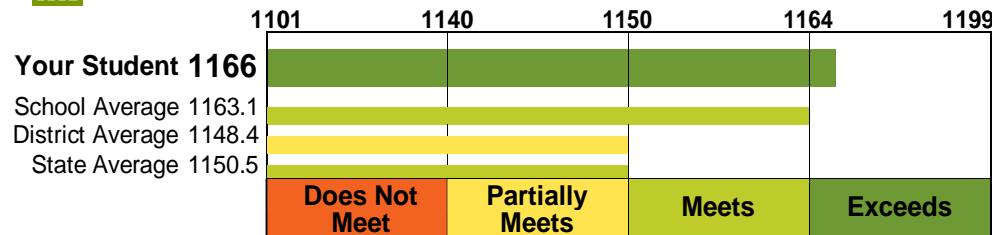
Paul's score shows performance **above** grade 11 standards

STATE PERCENTILE RANK

Paul scored higher than **86%** of Minnesota grade 11 students



MATHEMATICS: PAUL'S OVERALL MCA-III RESULTS



Performance Level Description:
Students at the **Exceeds the Standards** level exceed the mathematics skills of the Minnesota Academic Standards.

Mathematics Performance Details

Your student's score is compared to the state expectations for mathematics and its content areas. Performance in content areas is reported as Below Expectations, At or Near Expectations, or Above Expectations.

= Below Expectations

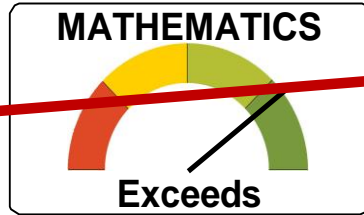
= At or Near Expectations

= Above Expectations

MATHEMATICS AREA	DESCRIPTION	PERFORMANCE
Algebra:	Identify features of functions and use them to solve real-world and mathematical problems, generate equivalent expressions, and solve equations and inequalities.	At or Near Expectations
Geometry and Measurement:	Calculate measurements, construct logical arguments to prove results, and apply properties of figures to solve problems.	Above Expectations
Data Analysis and Probability:	Display and analyze data; use various measures to draw conclusions, make predictions, and calculate probabilities.	Above Expectations

These are your student's results from Minnesota Comprehensive Assessments taken in the spring of 2016.

My student is OK.

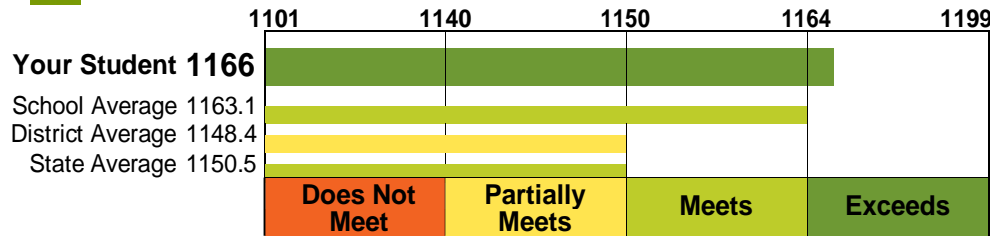


STANDARDS
Paul's score shows performance **above** grade 11 standards

STATE PERCENTILE RANK
Paul scored higher than **86%** of Minnesota grade 11 students



MATHEMATICS: PAUL'S OVERALL MCA-III RESULTS



Performance Level Description:
Students at the **Exceeds the Standards** level exceed the mathematics skills of the Minnesota Academic Standards.

Mathematics Performance Details

Your student's score is compared to the state expectations for mathematics and its content areas. Performance in content areas is reported as Below Expectations, At or Near Expectations, or Above Expectations.

= Below Expectations

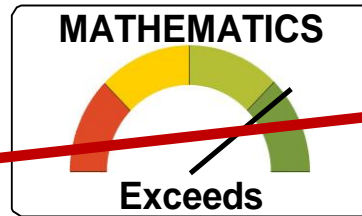
= At or Near Expectations

= Above Expectations

MATHEMATICS AREA	DESCRIPTION	PERFORMANCE
Algebra:	Identify features of functions and use them to solve real-world and mathematical problems, generate equivalent expressions, and solve equations and inequalities.	At or Near Expectations
Geometry and Measurement:	Calculate measurements, construct logical arguments to prove results, and apply properties of figures to solve problems.	Above Expectations
Data Analysis and Probability:	Display and analyze data; use various measures to draw conclusions, make predictions, and calculate probabilities.	Above Expectations

These are your student's results from Minnesota Comprehensive Assessments taken in the spring of 2016.

My
student
is
better
than...



STANDARDS

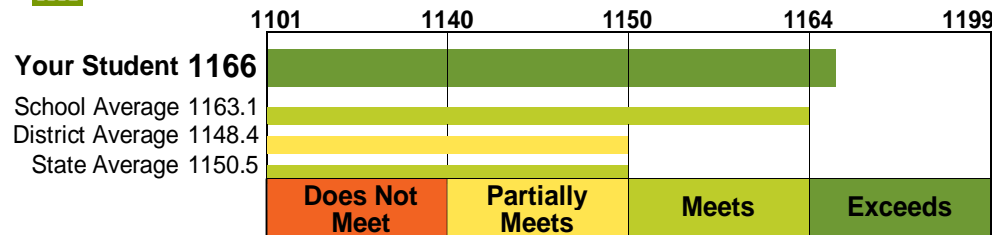
Paul's score shows
performance **above**
grade 11 standards

STATE PERCENTILE RANK

Paul scored higher
than **86%** of
Minnesota grade 11
students



MATHEMATICS: PAUL'S OVERALL MCA-III RESULTS



Performance Level Description:

Students at the **Exceeds the Standards** level exceed the mathematics skills of the Minnesota Academic Standards.

Mathematics Performance Details

Your student's score is compared to the state expectations for mathematics and its content areas. Performance in content areas is reported as Below Expectations, At or Near Expectations, or Above Expectations.

= Below Expectations

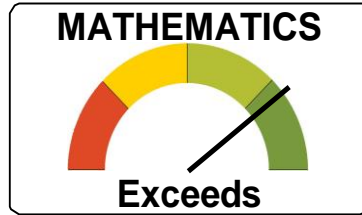
= At or Near Expectations

= Above Expectations

MATHEMATICS AREA	DESCRIPTION	PERFORMANCE
Algebra:	Identify features of functions and use them to solve real-world and mathematical problems, generate equivalent expressions, and solve equations and inequalities.	At or Near Expectations
Geometry and Measurement:	Calculate measurements, construct logical arguments to prove results, and apply properties of figures to solve problems.	Above Expectations
Data Analysis and Probability:	Display and analyze data; use various measures to draw conclusions, make predictions, and calculate probabilities.	Above Expectations

These are your student's results from Minnesota Comprehensive Assessments taken in the spring of 2016.

How my student compares to school, district, state students



STANDARDS

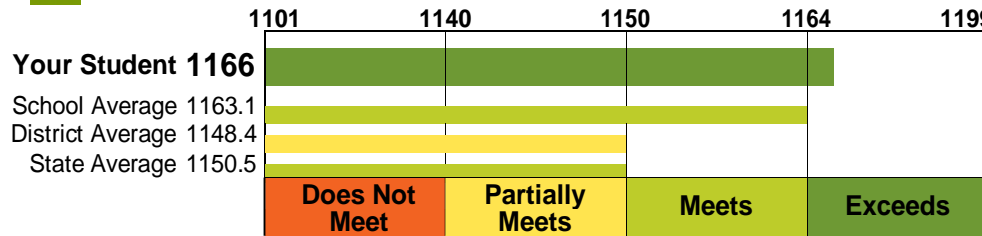
Paul's score shows performance **above** grade 11 standards

STATE PERCENTILE RANK

Paul scored higher than **86%** of Minnesota grade 11 students



MATHEMATICS: PAUL'S OVERALL MCA-III RESULTS



Performance Level Description:
Students at the **Exceeds the Standards** level exceed the mathematics skills of the Minnesota Academic Standards.

Mathematics Performance Details

Your student's score is compared to the state expectations for mathematics and its content areas. Performance in content areas is reported as Below Expectations, At or Near Expectations, or Above Expectations.

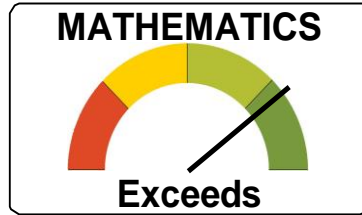
= Below Expectations

= At or Near Expectations

= Above Expectations

MATHEMATICS AREA	DESCRIPTION	PERFORMANCE
Algebra:	Identify features of functions and use them to solve real-world and mathematical problems, generate equivalent expressions, and solve equations and inequalities.	At or Near Expectations
Geometry and Measurement:	Calculate measurements, construct logical arguments to prove results, and apply properties of figures to solve problems.	Above Expectations
Data Analysis and Probability:	Display and analyze data; use various measures to draw conclusions, make predictions, and calculate probabilities.	Above Expectations

These are your student's results from Minnesota Comprehensive Assessments taken in the spring of 2016.



STANDARDS

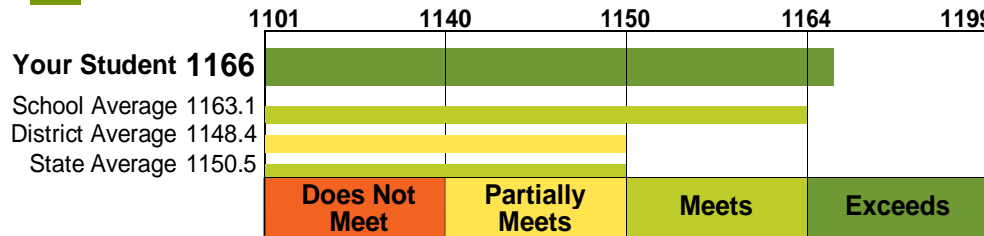
Paul's score shows performance **above** grade 11 standards

STATE PERCENTILE RANK

Paul scored higher than **86%** of Minnesota grade 11 students



MATHEMATICS: PAUL'S OVERALL MCA-III RESULTS



Performance Level Description:
Students at the **Exceeds the Standards** level exceed the mathematics skills of the Minnesota Academic Standards.

Details
(if you want them)

Mathematics Performance Details

Your student's score is compared to the state expectations for mathematics and its content areas. Performance in content areas is reported as Below Expectations, At or Near Expectations, or Above Expectations.

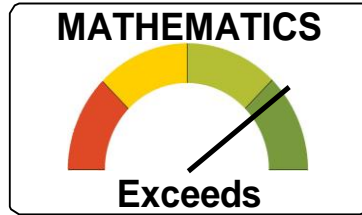
= Below Expectations

= At or Near Expectations

= Above Expectations

MATHEMATICS AREA	DESCRIPTION	PERFORMANCE
Algebra:	Identify features of functions and use them to solve real-world and mathematical problems, generate equivalent expressions, and solve equations and inequalities.	At or Near Expectations
Geometry and Measurement:	Calculate measurements, construct logical arguments to prove results, and apply properties of figures to solve problems.	Above Expectations
Data Analysis and Probability:	Display and analyze data; use various measures to draw conclusions, make predictions, and calculate probabilities.	Above Expectations

These are your student's results from Minnesota Comprehensive Assessments taken in the spring of 2016.



STANDARDS

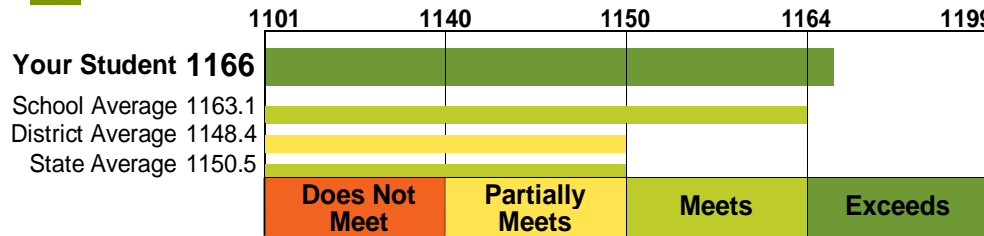
Paul's score shows performance **above** grade 11 standards

STATE PERCENTILE RANK

Paul scored higher than **86%** of Minnesota grade 11 students



MATHEMATICS: PAUL'S OVERALL MCA-III RESULTS



Performance Level Description:
Students at the **Exceeds the Standards** level exceed the mathematics skills of the Minnesota Academic Standards.

Ideas for following up

Mathematics Performance Details

Your student's score is compared to the state expectations for mathematics and its content areas. Performance in content areas is reported as Below Expectations, At or Near Expectations, or Above Expectations.

= Below Expectations

= At or Near Expectations

= Above Expectations

MATHEMATICS AREA	DESCRIPTION	PERFORMANCE
Algebra:	Identify features of functions and use them to solve real-world and mathematical problems, generate equivalent expressions, and solve equations and inequalities.	At or Near Expectations
Geometry and Measurement:	Calculate measurements, construct logical arguments to prove results, and apply properties of figures to solve problems.	Above Expectations
Data Analysis and Probability:	Display and analyze data; use various measures to draw conclusions, make predictions, and calculate probabilities.	Above Expectations



What We Can/Should Do

- 1) Score Reporting
- 2) Educator Involvement
 - * contracted work
 - * professional development



What We Can/Should Do

- 1) Score Reporting
- 2) Educator Involvement
- 3) Other Indicators**



What We Can/Should Do

- 1) Score Reporting
- 2) Educator Involvement
- 3) Other Indicators
- 4) Testing Time**



Federal/State Mandated Summative Testing Time in hours (grade)

Unit	Elementary	Middle	H.S.
NC	9.0 (3)	13.0 (8)	15.0 (10)
Ohio	17.2 (3)	19.3 (8)	21.7 (11)
*CGCS	20.6 (3)	25.3 (8)	22.5 (11)
US	16.6 (3)	17.1 (7)	---

* Includes district mandated tests



Federal/State Mandated Summative Testing Time

Total Testing Time/Instructional Time

* 1.7% (Teoh, et al, 2014)

* 1.6% (Guindon, et al, 2014)

$$180 \times 6 = 1080 \times .0165 = 18.7$$



Reality vs. Perceptions of Testing Time

n > 300 K, 3, 7 teacher estimates of testing time

Kindergarten: x 3.04

Third Grade: x 1.95

Seventh Grade: x 1.10



Reality vs. Perceptions of Testing Time

n > 300 K, 3, 7 teacher estimates of testing time

Kindergarten: x 3.04

Third Grade: x 1.95

Seventh Grade: x 1.10

...but likely underestimates



Reality vs. Perceptions of Testing Time

Additional Factors:

- * district, other mandated tests
- * non-mandated tests
- * test preparation activities (tests, prep)
- * test administration, scoring, reporting activities
- * test administration "windows"
- * "non-interference" actions
- * facilities scheduling disruptions



Five Conclusions

- 1) Little state/federal mandated testing
- 2) Modest total testing
- 3) Substantial perceived amount of testing
- 4) Many perceived negative consequences
- 5) Few (negligible?) perceived benefit(s)



Patterns of Non-participation

- * Increase across grades
- * Greater for Math than ELA
- * Lower for larger districts
- * Lower for more disadvantaged districts
- * Greater for districts with lower scores
- * Greater for accountability tests
- * Association with educator evaluation



Definitely Ineffective... (3/3)

- 1) Repeat information on actual amount of testing
- 2) Appeals to civic duty
- 3) Statistical adjustments without accurate information on ability distribution



Reasonable Responses...

(5/5)

- * Do nothing
- * More complete data collection
- * Reassure that all scores, categories are still valid
- * Incentivize participation by providing something of value or consequence
- * Actually reduce testing time



Seven Strategies for Reducing Testing Time

- 1) Embedded field tests (not stand alone)
- 2) Operational field testing
- 3) Narrow testing windows
- 4) Eliminate untimed tests
- 5) Shorter tests
- 6) Reduce/eliminate other state/district mandated and optional testing
- 7) Out-of-school testing



What We Can/Should Do

- 1) Score Reporting
- 2) Educator Involvement
- 3) Other Indicators
- 4) Testing Time
- 5) Validity**
- 6) Test Security**



Test Security and Validity: Three New Ideas



1) Test Security = Validity

Test security is a validity concern; we most appropriately stay within the boundaries of our expertise when we weigh in with authority on the validity of test data.

2) The Weight of Quantitative Evidence



“It would seem prudent that the weight of statistical evidence only be brought to bear when some other circumstances—that is, a trigger—provide a reason for flagging cases for subsequent statistical analysis.”

(1999, p. 142)

2) The Weight of Quantitative Evidence

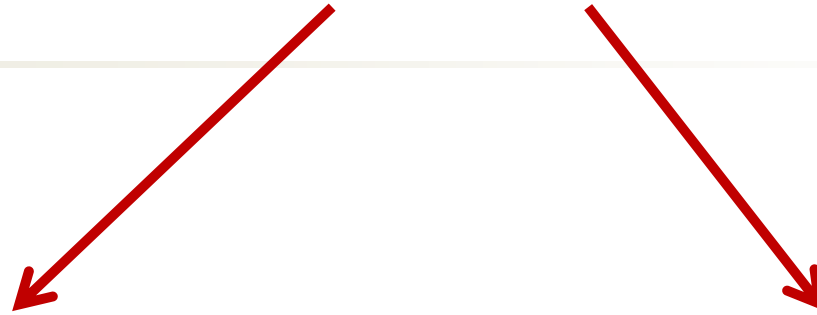


“It is entirely defensible not only to use statistical analysis itself as a trigger to prompt additional investigation but, in some cases, the application of appropriate statistical methods yielding highly improbable findings can alone be sufficient to conclude that a test score should not be considered a valid representation of an examinee’s knowledge or ability.”



3) The *Standards* Must Change

A Comprehensive Framework for Defensible Testing



Validation of
Score Meaning

≠

Justification
of Score Use

“What is the evidence that the score can be interpreted as intended?”

“What is the evidence that the score *should* be used as proposed?”

The *Standards for Educational and Psychological Testing*



Evidence based on...



The *Standards for Educational and Psychological Testing*

Evidence based on...

- 1) Test Content**
- 2) Response Process**
- 3) Internal Structure**
- 4) Relations to Other Variables**
- 5) Consequences of Testing**

The *Standards for Educational and Psychological Testing*

Evidence based on...

1) Test Content

2) Response Process

3) Internal Structure

4) Relations to Other Variables

~~**5) Consequences of Testing**~~

A Comprehensive Framework for Defensible Testing



*Standards for
Educational and
Psychological Test
Use*

Justification
of Score Use

“What is the evidence that
the score *should* be used
as proposed?”



The *Standards for Educational and Psychological Testing*

Evidence based on...

- 1) Test Content**
- 2) Response Process**
- 3) Internal Structure**
- 4) Relations to Other Variables**

The *Standards for Educational and Psychological Testing*

Evidence based on...

1) Test Content

2) Response Process

3) Internal Structure

4) Relations to Other Variables

The *Standards for Educational and Psychological Testing*



Evidence based on...

1) Test Content

2) Response Process

**3) Hypothesized Relationships
among Variables**



The *Standards for Educational and Psychological Testing*

Evidence based on...

- 1) Test Content**
- 2) Response Process**
- 3) Hypothesized Relationships among Variables**
- 4) Test Development and Administration Procedures**



Formative Assessment

A process for helping students to identify and internalize their learning goals, take responsibility for their learning, reflect on their own understanding and evaluate the quality of their work in relation to their own or objective goals, and identify strategies to improve their work and understanding

10 Principles of Formative Assessment

(Cizek, 2010)

1. Requires students to take responsibility for their own learning
2. Communicates clear, specific learning goals
3. Focuses on valuable educational outcomes with applicability beyond the learning context
4. Identifies students' current knowledge/skills and the necessary steps for reaching the desired goals

10 Principles of Formative Assessment

(Cizek, 2010)

5. Requires development of plans for attaining the desired goals
6. Encourages students to self-monitor progress toward the learning goals
7. Provides examples of learning goals including, when relevant, the specific grading criteria or rubrics that will be used to evaluate the student's work

10 Principles of Formative Assessment

(Cizek, 2010)

8. Includes frequent assessment, including peer and student self-assessment and assessment embedded within learning activities
9. Provides feedback that is non-evaluative, specific, timely, related to the learning goals, and provides opportunities for the student to revise and improve work products
10. Teaches/promotes student metacognition, reflection, and self-regulation



What We Can't/Shouldn't Do

1) College and Career Readiness

“70% chance of getting a C or better in an introductory, credit-bearing....”

“75% chance of getting a score of x or better some another test...”



What We Can't/Shouldn't Do

1) College and Career Readiness

“Intelligence as a measurable capacity must at the start be defined as the capacity to do well in an intelligence test. Intelligence is what the tests test.”

(Boring, 1923, p. 35)



What We Can't/Shouldn't Do

1) College and Career Readiness

2) Test Purposes

- * S-R vs predictive

- * C-R vs N-R standard setting



What We Can't/Shouldn't Do

1) College and Career Readiness

2) Test Purposes

- * S-R vs predictive

- * C-R vs N-R standard setting

3) Over-responsiveness



The demands....

- * diagnostic
- * tailored/individualized
- * less time/more efficient
- * more/actionable subscores
- * less standardized conditions
- * wider windows
- * fast, user-friendly reports
- * “highly reliable and valid”



Responsible Responsiveness

- * advantages/disadvantages
- * "risk assessment"



Conclusions, Questions?

- Thank you...