Multiple Measures, Multiple Meanings

Charles A. DePascale
Center for Assessment (NCIEA)

2011 Reidy Interactive Lecture Series
Boston, Massachusetts
September 22, 2011

Code of Professional Responsibilities in Educational Measurement (Section 6.7), NCME (1995)

- Section 6.7
  - The interpretation, use, and communication of assessment results should promote valid inferences and minimize invalid ones. Persons who interpret, use, and communicate assessment results have a professional responsibility to use multiple sources and types of relevant information about persons or programs whenever possible in making educational decisions.

Standards for Educational and Psychological Testing
(AERA, APA, NCME, 1999)

- # Standard 13.7
  - In educational settings, a decision or characterization that will have major impact on a student should not be made on the basis of a single test score. Other relevant information should be taken into account if it will enhance the overall validity of the decision.

IASA (1994) ➡ NCLB (2001)

- Section 1111(b)3.
  - Assessments.--Each State plan shall demonstrate that the State has developed or adopted a set of high-quality, yearly student assessments, including assessments in at least mathematics and reading or language arts, that will be used as the primary means of determining the yearly performance of each local educational agency and school served under this part in enabling all children served under this part to meet the State's student performance standards. Such assessments shall–
  - ``(A) be the same assessments used to measure the performance of all children, if the State measures the performance of all children;
  - ``(B) be aligned with the State's challenging content and student performance standards and provide coherent information about student attainment of such standards; `
  - `(C) be used for purposes for which such assessments are valid and reliable, and be consistent with relevant, nationally recognized professional and technical standards for such assessments;
  - ``(D) measure the proficiency of students in the academic subjects in which a State has adopted challenging content and student performance standards and be administered at some time during–
    - ``(i) grades 3 through 5;
    - ``(ii) grades 6 through 9; and
    - ``(iii) grades 10 through 12;
  - ``(E) involve multiple up-to-date measures of student performance, including measures that assess higher order thinking skills and understanding;
  - ``(F) provide for–
    - ``(i) the participation in such assessments of all students;
    - ``(ii) the reasonable adaptations and accommodations for students with diverse learning needs, necessary to measure the achievement of such students relative to State content standards; and
    - ``(iii) the inclusion of limited English proficient students who shall be assessed, to the extent practicable, in the language and form most likely to yield accurate and reliable information on what such students know and can do, to determine such students' mastery of skills in subjects other than English;
  - ``(G) include students who have attended schools in a local educational agency for a full academic year but have not attended a single school for a full academic year, however the performance of students who have attended more than one school in the local educational agency in any academic year shall be used only in determining the progress of the local educational agency;
  - ``(H) provide individual student interpretive and descriptive reports, which shall include scores, or other information on the attainment of student performance standards; and
  - ``(I) enable results to be disaggregated within each State, local educational agency, and school by gender, by each major racial and ethnic group, by English proficiency status, by migrant status, by students with disabilities as compared to nondisabled students, and by economically disadvantaged students as compared to students who are not economically disadvantaged.

  - ``(4) Special rule.--Assessment measures that do not meet the requirements of paragraph (3)(C) may be included as one of the multiple measures, if a State includes in the State plan information regarding the State's efforts to validate such measures.

IASA (1994) → NCLB (2001)

- Section 1111(b)3.
  - Assessments.--Each State plan shall demonstrate that the State has developed or adopted a set of high-quality, yearly student assessments, including assessments in at least mathematics and reading or language arts, that will be used as the primary means of determining the yearly performance of each local educational agency and school served under this part in enabling all children served under this part to meet the State's student performance standards. Such assessments shall–
  - ``(A) be the same assessments used to measure the performance of all children, if the State measures the performance of all children;
  - ``(B) be aligned with the State's challenging content and student performance standards and provide coherent information about student attainment of such standards; `
  - `(C) be used for purposes for which such assessments are valid and reliable, and be consistent with relevant, nationally recognized professional and technical standards for such assessments;
  - ``(D) measure the proficiency of students in the academic subjects in which a State has adopted challenging content and student performance standards and be administered at some time during

> ``(E) involve multiple up-to-date measures of student performance, including measures that assess higher order thinking skills and understanding;

State content standards; and
  - ``(iii) the inclusion of limited English proficient students who shall be assessed, to the extent practicable, in the language and form most likely to yield accurate and reliable information on what such students know and can do, to determine such students' mastery of skills in subjects other than English;
  - ``(G) include students who have attended schools in a local educational agency for a full academic year but have not attended a single school for a full academic year, however the performance of students who have attended more than one school in the local educational agency in any academic year shall be used only in determining the progress of the local educational agency;
  - ``(H) provide individual student interpretive and descriptive reports, which shall include scores, or other information on the attainment of student performance standards; and
  - ``(I) enable results to be disaggregated within each State, local educational agency, and school by gender, by each major racial and ethnic group, by English proficiency status, by migrant status, by students with disabilities as compared to nondisabled students, and by economically disadvantaged students as compared to students who are not economically disadvantaged.

> ``(4) Special rule.--Assessment measures that do not meet the requirements of paragraph (3)(C) may be included as one of the multiple measures, if a State includes in the State plan information regarding the State's efforts to validate such measures.

NCLB – Special Rule

- (4) SPECIAL RULE- Academic assessment measures in addition to those in paragraph (3) that do not meet the requirements of such paragraph may be included in the assessment under paragraph (3) as additional measures, but may not be used in lieu of the academic assessments required under paragraph (3). Such additional assessment measures may not be used to reduce the number of or change, the schools that would otherwise be subject to school improvement, corrective action, or restructuring under section 1116 if such additional indicators were not used, but may be used to identify additional schools for school improvement or in need of corrective action or restructuring except as provided in paragraph (2)(I)(i).

- National Education Association
  - http://www.nea.org/home/16874.htm
- Selected Civil Rights groups
  - http://www.edaccountability.org/MultipleMeasuresCRletter%20_5_.pdf
- National Conference of State Legislatures
  - http://www.ncsl.org/documents/educ/NCLBRecommendations.pdf
- ASCD
  - http://www.ascd.org/news-media/ASCD-Policy-Positions/ASCD-Positions.aspx

In the beginning

- In almost every situation in which psychological measurement is applied, more than one measurement is involved. The introduction of multiple measures of the same individual, or set of individuals, raises questions concerning how these several measures should be chosen, in what terms they should be expressed, and how they can be combined into a meaningful whole. The combination of the several scores for one individual may either merge the several parts or preserve their respective identities.
  – Charles I Mosier (Educational Measurement, 1951)

3 distinct purposes

1. Provide finer degrees of discrimination
2. Increase the reliability of measurement
3. Provide measures of unrelated aspects of the behavior-to-be-predicted

And a 4ᵗʰ purpose

4. As an alternative measure
   - Use of an alternative measure of the same construct with NO intent of using information from both measures.
   - Could be considered within the framework of the other purposes, but it may be best to treat separately from "multiple measures"

1. Finer degrees of discrimination

- The most simple concept…
- But can lead to the most complex designs
- Adding more measures allows for more distinctions among performances.
- More distinctions among performances can lead to
  - More precision in measurement
  - Improved interpretation

With Number Correct Scoring

1-item test
2-item test
4-item test
10-item test

→

2 categories
3 categories
5 categories
11 categories

With Pattern Scoring

1-item test
2-item test
4-item test
10-item test

→

2 categories
4 categories
16 categories
1024 categories

But????

```
                1
             1     1
          1     2     1
        1     3     3     1
      1     4     6     4     1
   1     5    10    10    5     1
```
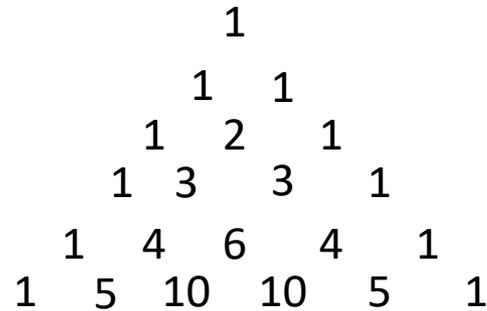
- The number of categories quickly increases as items or components are added.

- How many meaningful distinctions can be made?
  - What are the relationships among components?
  - What combinations can be expected?

2. Increase reliability of measurement

- Each observation has some elements of unreliability
- The reliability of the composite is increased by increasing the number of observations*
- The measures should be highly intercorrelated, since all are intended to be measures of the same fundamental characteristic.
- * Increasing the number of observations or components is not a guarantee of increased reliability.

2. Increase reliability (Variation 1)

- Multiple tests of a common factor
  - Minimize chance error
  - Minimize systematic effects of a large number of uncorrelated factors as well
- Each test measures a common factor, but some other factor as well
  - "other" factor not shared across tests
- Aggregate minimizes the impact of the "other" factors

2. Increase Reliability (Variation 2)

- Multiple opportunities to pass the same test
  - In theory, multiple opportunities would provide a more reliable estimate of a student's true score than a single observed score.
  - In practice, only a single observed score is used as the indicator of student achievement
    - Graduation/exit exams
    - Admissions tests

3. Unrelated Aspects

- Unrelated = uncorrelated

- Purpose is to measure *different* aspects of the criterion, each related to the criterion, but not duplicating each other.

- Substantial correlation with the criterion and "absence" or "lack" of correlation with each other.

Comparing Assessment Designs

2. Increased Reliability (Related Aspects)

$$R = F_1(X)$$

$$A = F_2(X, U_A)$$

$$B = F_3(X, U_B)$$

$$C = F_4(X, U_C)$$

3. Unrelated aspects

$$R = F_1(X,Y,Z)$$

$$A = F_2(X, U_A)$$

$$B = F_3(Y, U_B)$$

$$C = F_4(Z, U_C)$$

"Many a slip between scoring model and score interpretation" – Messick (1989)

- A scoring or measurement model describes the way in which item responses are combined to form test scores, along with any control procedures taking account of conditions of testing that influence score interpretation.

- Model compounding – two or more scoring or measurement models are combined, or one is overlaid on the other, so that one aspect of test interpretation derives from one model and another aspect from the overlay.
  - Leads to confusion as to what construct theory to reference.
  - Leads to confusion about the forms of evidence needed for the construct validation of the compound interpretation.
  - Compound score interpretations require compounded sources of evidence…

- Model slippage – scores derived by one measurement model are interpreted in terms of a different measurement model.
  - Scores designed to distinguish among groups (class model) are interpreted to make inferences about the amount of some characteristic or trait (cumulative model)

- "One person's inadvertent model slippage may be another's deliberate creative leap"

Alternative Models for scoring and score interpretation – Messick (1989)

- Cumulative Quantitative
  - Higher scores associated with more of the trait
- Quality or intensity of a response
  - Non-compensatory; highest rating attained
- Class models
  - Higher score indicates greater probability of class membership
  - Criterion-referenced or domain-referenced interpretations
- Dynamic Models
  - Two or more manifestations of the same trait my be mutually exclusive or negatively correlated
- Ipsative Models
  - Comparisons within individuals (relative strengths and weaknesses)
- Normative
  - Comparisons across individuals
- Criterion-referenced or domain-referenced

Composite Score v. Score Profile

- Either can support a summative decision
- Choice requires a policy decision which may be based on several factors
  - Policy/politics, measurement, cost, ease
- Choice should be
  - related to the reason for having multiple measures
  - consistent with the assessment system design
  - consistent with the proposed uses of the system

Composite Scores

- Tend to reflect a compensatory model
- Require decisions about weighting of components
  - There are no "non-decisions" about weighting
  - Weightings are not as simple and straightforward as they may appear.
  - Certain weightings may improve reliability or validity, but not both

Profile Scores

- Tend to reflect a non-compensatory model
  - Conjunctive, Disjunctive or Complementary
- Can focus on patterns of performance as well as individual components
- Usually require some type of processing of individual component scores to support appropriate comparisons
- May defer complex considerations to the judgment of the interpreter – beware of "seeming simplicity" of profiles

Variations on Composites and Profiles

- Tiers, Screening, Successive Hurdles,
    - Components are administered consecutively and only those who "meet" or "fail to meet" a specified criteria on one component are administered the following component.
    - Under a conjunctive system, each component measures critical, necessary knowledge/skills
    - Efficiency is a prime consideration in determining order of administration.
- Dashboards

Time

- Introduction of time impacts measurement and the interpretation of scores.

- Across short time periods

- Across longer time periods "growth" becomes a factor. Are we interested in
  - Growth from Time A to Time B
  - Achievement at Time B
  - "Typical" performance across Time A to Time B

Examples

- Grades
- High school graduation
- College Admission
  - SAT and ACT
  - College admissions process
- Teacher Evaluation
  - DC Impact
  - Rhode Island
- Race to the Top

Grades
Rocky River High School

HIGH SCHOOL GRADING POLICY

Grades are quantitative statements reflecting instructor assessment of student performance. Grades represent the degree of mastery of the prescribed curriculum of a given content area or course at a given point in time in a student's educational development. Measurement of a student's performance provides a means for educators to communicate with students and parents. *Grades indicate academic achievement of a student, not necessarily one's ability.* Assessment is to be a daily function so that feedback, both written and verbal, is constant. Grades reflect actual student's academic performance. Parameters include authentic assessments, tests, class work, homework, and class participation. Teachers will follow additional guidelines in the courses of study when determining a grade. Grades are interpreted as follows:

GRADES IN GPA

A+	=	Superior
A, A-, B+	=	Excellent
B, B-, C+	=	Good
C, C-, D+	=	Average
D, D-	=	Below Average
F	=	Failing
WF	=	Withdrew Failing

GRADES NOT IN GPA

I	=	Incomplete
p	=	Passed
EX	=	Excused
WP	=	Withdrew Passing

Grades
Calcasieu Parish School Board

GRADE COMPUTATION

Reports related to student progress will be made to parents of regular and special education students at the end of each marking period throughout the session. Teachers are expected to evaluate at least ten (10) different times during the period to arrive at a fair grade (tests, class participation, homework, reports, experiments, themes, etc.) Teachers shall not assign multiple grades to individual tasks in order to meet the required number of evaluations. Special projects like term papers, which include multiple steps that are graded separately, do not violate the spirit of this policy. The grade for the marking period will be an average of these grades. All grades shall be recorded numerically.

High School Graduation
Darling-Hammond et al. (2005)

Multiple measures approaches to graduation provide diverse opportunities for students to demonstrate what they have learned, including research papers, projects, exhibitions, and other performance assessments that evaluate a wide range of skills. The concept of multiple measures is routinely used by policymakers to make critical decisions about such matters as employment and economic forecasting (for example, the Dow Jones Index or the GNP), as well as admission to universities, where grades, essays, activities, and accomplishments are considered along with test scores. Successful businesses use a "dashboard" set of indicators to evaluate their health and progress, aware that no single indicator is sufficient to understand their operations. This concept was embraced in successive revisions of the Elementary and Secondary Education Act, including the No Child Left Behind Act in 2002, which calls for multiple measures of student performance, extending beyond test scores.

ACT and SAT

ACT

- Composite
  - English
  - Mathematics
  - Reading
  - Science
- Composite score is an average of test scaled scores ("equal" weights)
- Separate scales were designed to have consistent relationship between scaled scores and standard error

SAT

- Profile
  - Critical Reading
  - Mathematics
  - Writing

- Separate scales were designed to have the same mean and standard deviation

ACT and SAT College Readiness

ACT

- Profile
- Compute benchmarks for individual tests
- Related to probability of success in particular credit-bearing courses

SAT

- Composite
- Compute a college readiness score based on an aggregate of Critical Reading, Mathematics, and Writing scores
- Related to first-year college GPA

APPLYING ONLINE: Both the Common Application and the Universal College Application (**www.commonapp.org** and **www.universalcollegeapp.com**) allow students to apply online. The Duke Student Supplement is also available at these Web sites and can be submitted online.

ESSAY AND SHORT ANSWER QUESTIONS: Your essay is an important part of your application. It is your primary chance to speak to us in your own voice and a way for us to learn about you as a person, so write from the heart and be yourself. We encourage you to take time with your essay; don't simply write it online and submit it. Print it out, have others read it, and submit it only when you are completely satisfied that it represents you and your best effort. Applicants to Trinity College may also choose to answer the short answer question on the Duke Student Supplement; applicants to the Pratt School of Engineering are required to answer this question. In either case, please be sure to give your responses thoughtful consideration.

TEACHER RECOMMENDATIONS: Duke requires two teacher recommendations. Complete the student portion of the Teacher Evaluations and give the forms to teachers who have taught you within the last two years of secondary school in major academic courses (English, mathematics, social studies, sciences, foreign language). If you are applying to the Pratt School of Engineering, one recommendation should be from a math or science teacher.

TESTING: All candidates for admission must complete one of the following and arrange to have official test score reports sent to Duke:

- **ACT, including the writing exam**

  OR

- **SAT plus two SAT Subject Tests**
  Applicants to the Trinity College of Arts & Sciences who take the SAT may take any two SAT Subject Tests. Applicants to the Pratt School of Engineering who take the SAT must take one SAT Subject Test in Mathematics.

SCHOOL REPORTS: Complete the student portions of the School Report and the Midyear Report, and give these forms to your guidance counselor or advisor. Be sure to request that an official transcript be sent to Duke. If you are applying Early Decision, you should also give your counselor or advisor Duke's First Quarter Grade Report (available at **www.admissions.duke.edu**). If you apply via the Common Application, your counselor and teachers will have the option to complete their forms online through the Common Application Online School Forms system.

OPTIONAL ARTISTIC MATERIAL: If you have exceptional talent in art, dance, music, theater, photography, or film/video/digital media, you may submit an artistic material form along with supplementary material to be evaluated by an appropriate faculty member.

Both the Common Application and Universal College Application provide material forms. For specific guidelines about what to submit, please visit **www.admissions.duke.edu** and click on "How to Apply," then "Optional Artistic Material."

DCPS Impact

SAMPLE SCORE CHART
TEACHING AND LEARNING FRAMEWORK (TLF)

TEACHING AND LEARNING FRAMEWORK (TLF)	ADMIN CYCLE ENDS 12/1	ADMIN CYCLE ENDS 3/1	ADMIN CYCLE ENDS 6/15	ME CYCLE ENDS 2/1	ME CYCLE ENDS 6/15	OVERALL ANNUAL COMPONENT SCORE (Average of Cycles)
TLF SCORE (Average of Teach 1 to Teach 9)	3.7	3.8	3.7	3.7	3.8	3.7
Teach 1: Lead Well-Organized, Objective-Driven Lessons	3.0	4.0	4.0	4.0	4.0	
Teach 2: Explain Content Clearly	4.0	3.0	4.0	4.0	4.0	
Teach 3: Engage Students at All Learning Levels in Rigorous Work	4.0	4.0	4.0	4.0	4.0	
Teach 4: Provide Students Multiple Ways to Engage with Content	3.0	4.0	3.0	3.0	4.0	
Teach 5: Check for Student Understanding	4.0	4.0	4.0	3.0	3.0	
Teach 6: Respond to Student Misunderstandings	4.0	3.0	4.0	4.0	4.0	
Teach 7: Develop Higher-Level Understanding through Effective Questioning	4.0	4.0	4.0	4.0	4.0	
Teach 8: Maximize Instructional Time	3.0	4.0	3.0	3.0	4.0	
Teach 9: Build a Supportive, Learning-Focused Classroom Community	4.0	4.0	3.0	4.0	3.0	

SAMPLE SCORE

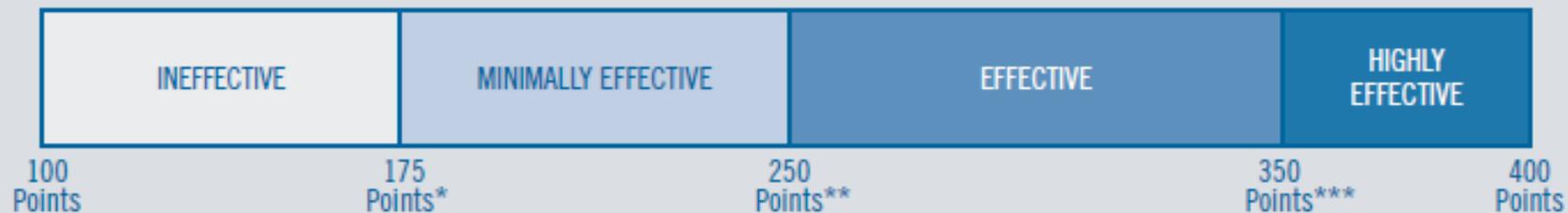| COMPONENT	COMPONENT SCORE	PIE CHART PERCENTAGE		WEIGHTED SCORE	
Individual Value-Added Student Achievement Data (IVA)	3.5	x	50	=	175
Teaching and Learning Framework (TLF)	3.7	x	35	=	130
Commitment to the School Community (CSC)	3.5	x	10	=	35
School Value-Added Student Achievement Data (SVA)	3.3	x	5	=	17
TOTAL					357

Step 4

We then adjust your total score based on your rating for Core Professionalism. If your rating for this component is Meets Standard for both cycles, then your total score remains unchanged. If not, then 10 points are subtracted from your total score for each cycle in which your rating is Slightly Below Standard, and 20 points are subtracted for each cycle in which your rating is Significantly Below Standard. In the example above, the individual's rating for all cycles is Meets Standard, so no points have been subtracted.

SAMPLE SCORE CHART
CORE PROFESSIONALISM (CP)

CORE PROFESSIONALISM (CP)	CYCLE ENDS 12/1	CYCLE ENDS 6/15	OVERALL
CP SCORE *(Lowest of CP 1 to CP 4)*	MEETS STANDARD	MEETS STANDARD	MEETS STANDARD
CP 1: Attendance	MEETS STANDARD	MEETS STANDARD	
CP 2: On-Time Arrival	MEETS STANDARD	MEETS STANDARD	
CP 3: Policies and Procedures	MEETS STANDARD	MEETS STANDARD	
CP 4: Respect	MEETS STANDARD	MEETS STANDARD	

OVERALL IMPACT SCALE

INEFFECTIVE	MINIMALLY EFFECTIVE	EFFECTIVE	HIGHLY EFFECTIVE

100 Points   175 Points*   250 Points**   350 Points***   400 Points

*A score of exactly 175 would be classified as Minimally Effective.

**A score of exactly 250 would be classified as Effective.

***A score of exactly 350 would be classified as Highly Effective.

Rhode Island Model

Professional Practice

Professional Responsibilities		Exemplary	Proficient	Emerging	Unsatisfactory
	Exceeds Expectations	4	3	2	2
	Meets Expectations	4	3	(2)	1
	Does Not Meet Expectations	3	2	1	1

Student Learning Objectives

Growth Model		Exceptional Attainment	Full Attainment	Considerable Attainment	Partial Attainment	Minimal/No Attainment
	High Growth 3	5	5	4	3	2
	Typical Growth 2	5	(4)	3	2	1
	Low Growth 1					

The Rhode Island Growth Model will not be used for ratings in school year 2011-2012

STUDENT LEARNING

PP x PR		5	(4)	3	2	1
	4	HE	HE	E	E*	D*
	3	HE	E	E	D	I*
	(2)	E*	E	D	D	I
	1	D*	D*	D	I	I

Race to the Top

II. Points Overview

The chart below shows the maximum number of points that may be assigned to each criterion.

Selection Criteria	Points	Percent
A. State Success Factors	**125**	**25%**
(A)(1) Articulating State's education reform agenda and LEAs' participation in it	65	
(i) Articulating comprehensive, coherent reform agenda	5	
(ii) Securing LEA commitment	45	
(iii) Translating LEA participation into statewide impact	15	
(A)(2) Building strong statewide capacity to implement, scale up, and sustain proposed plans	30	
(i) Ensuring the capacity to implement	20	
(ii) Using broad stakeholder support	10	
(A)(3) Demonstrating significant progress in raising achievement and closing gaps	30	
(i) Making progress in each reform area	5	
(ii) Improving student outcomes	25	
B. Standards and Assessments	**70**	**14%**
(B)(1) Developing and adopting common standards	40	
(i) Participating in consortium developing high-quality standards	20	
(ii) Adopting standards	20	
(B)(2) Developing and implementing common, high-quality assessments	10	
(B)(3) Supporting the transition to enhanced standards and high-quality assessments	20	
C. Data Systems to Support Instruction	**47**	**9%**
(C)(1) Fully implementing a statewide longitudinal data system	24	
(C)(2) Accessing and using State data	5	
(C)(3) Using data to improve instruction	18	
D. Great Teachers and Leaders	**138**	**28%**
Eligibility Requirement (b)	eligibility	
(D)(1) Providing high-quality pathways for aspiring teachers and principals	21	
(D)(2) Improving teacher and principal effectiveness based on performance	58	
(i) Measuring student growth	5	
(ii) Developing evaluation systems	15	
(iii) Conducting annual evaluations	10	
(iv) Using evaluations to inform key decisions	28	
(D)(3) Ensuring equitable distribution of effective teachers and principals	25	
(i) Ensuring equitable distribution in high-poverty or high-minority schools	15	
(ii) Ensuring equitable distribution in hard-to-staff subjects and specialty areas	10	
(D)(4) Improving the effectiveness of teacher and principal preparation programs	14	
(D)(5) Providing effective support to teachers and principals	20	
E. Turning Around the Lowest-Achieving Schools	**50**	**10%**
(E)(1) Intervening in the lowest-achieving schools and LEAs	10	
(E)(2) Turning around the lowest-achieving schools	40	
(i) Identifying the persistently lowest-achieving schools	5	
(ii) Turning around the persistently lowest-achieving schools	35	
F. General	**55**	**11%**
Eligibility Requirement (a)	eligibility	
(F)(1) Making education funding a priority	10	
(F)(2) Ensuring successful conditions for high-performing charter schools and other innovative s	40	
(F)(3) Demonstrating other significant reform conditions	5	
Competitive Preference Priority 2: Emphasis on STEM	15	5%
TOTAL	**500**	**100%**

Subtotal: Accomplishments	260	52%
Subtotal: Plans	240	48%

Selection Criteria	Points	Percent
A. State Success Factors	125	25%
(A)(1) Articulating State's education reform agenda and LEAs' participation in it	65	
(i) Articulating comprehensive, coherent reform agenda	*5*	
(ii) Securing LEA commitment	*45*	
(iii) Translating LEA participation into statewide impact	*15*	
(A)(2) Building strong statewide capacity to implement, scale up, and sustain proposed plans	30	
(i) Ensuring the capacity to implement	*20*	
(ii) Using broad stakeholder support	*10*	
(A)(3) Demonstrating significant progress in raising achievement and closing gaps	30	
(i) Making progress in each reform area	*5*	
(ii) Improving student outcomes	*25*	
B. Standards and Assessments	70	14%
(B)(1) Developing and adopting common standards	40	
(i) Participating in consortium developing high-quality standards	*20*	
(ii) Adopting standards	*20*	
(B)(2) Developing and implementing common, high-quality assessments	10	
(B)(3) Supporting the transition to enhanced standards and high-quality assessments	20	
C. Data Systems to Support Instruction	47	9%
(C)(1) Fully implementing a statewide longitudinal data system	24	
(C)(2) Accessing and using State data	5	
(C)(3) Using data to improve instruction	18	