The Real Race Begins

Lessons from the First Round of Race to the Top | April 2010

Executive Summary

In Round 1 of Race to the Top, the U.S. Department of Education delivered on its promise to hold states to a high bar for reform. Only 2 states out of 16 finalists and 41 total applicants were selected for awards: Delaware and Tennessee.

These states won because they outlined bold, comprehensive visions of reform and demonstrated the ability to make them a reality. Statewide teacher effectiveness policies were the foundation for their success. They focused on putting effective teachers in every classroom and giving teachers the critical feedback and support they need to do their best work. They shifted to evaluation systems that improve their ability to recognize great teachers and respond to poor performance. Together they set a new benchmark for reform that Round 2 applicants must meet in order to win.

This analysis offers a close look at the scoring of the Round 1 finalists. It refutes some of the most common myths about Race to the Top and offers important lessons for states applying for the $3.4 billion in funding that remains available in Round 2.

At the same time, it examines scoring deficiencies that the Department of Education must address. While these issues did not result in a lowering of the bar for Round 1 winners, they could mean the difference between winning and losing for states applying in Round 2.

Less-deserving states could win at the expense of states truly committed to and capable of dramatic reform. This outcome would undermine what has been a visionary education agenda by the Obama administration to date.

Many of these problems can be easily corrected through improvements to the competition's scoring tools and in reviewer training and norming. The Department of Education has done a laudable job making adjustments on tight timelines in the past, and likely is developing solutions even now. Nonetheless, we believe corrections are imperative, not optional.

Race to the Top has already accelerated education reform by decades in some states. In fact, 16 states enacted legislation or regulatory changes in keeping with the contest's policy priorities before the first scores were announced. Round 1 raised the standard.

But Round 2 is the real race. We have a new set of best-in-class state strategies. We have long overdue dialogue and momentum. We have a review process that, though flawed in its pilot implementation, will be stronger in its second iteration. The next eight weeks offer an unprecedented window of opportunity for education reform in the United States.

Time to get to work.

Race to the Top Myths and Round 1 Results

Lessons for State Applicants

Findings and Recommendations on the Scoring Process

Appendices

Round 1 results refute some of the most common Race to the Top myths.

MYTH: **Politics will force the administration to reward states with political significance.**

FACT: **None of the three states considered political battlegrounds—Ohio, Pennsylvania and Florida—won in Round 1.** Neither did Democratic bastions such as California, New York, or the District of Columbia. President Obama and Secretary Duncan's home state, Illinois, also lost in Round 1. Three of the top six scoring states went to McCain in the 2008 presidential election, and three went to Obama. There is no evidence, in the outcomes or the scores and comments of individual reviewers, that political considerations influenced which states won.

MYTH: **Teachers unions across the country blocked meaningful reforms.**

FACT: **In many states that put forward exceptional plans—especially in the Great Teachers and Leaders section—state and local union affiliates stepped up to support strong policies.** In Illinois, Tennessee and Delaware, unions supported changes to states laws on evaluation. State unions endorsed the applications of Louisiana and Rhode Island, the top-scoring states on the Great Teachers and Leaders section. While state and local teachers unions certainly can do more to support reform, their reactions to Race to the Top were not monolithic or in lock-step with some national union rhetoric.

MYTH: States cannot win without full buy-in from local districts and unions, and that means watering down their reform agenda.

FACT: Winning states passed laws that locked in major reforms, and buy-in followed. They had a statewide vision for reform and left nothing to chance. States without full buy-in from local school districts and teachers union chapters also scored in the top five, such as Georgia, Florida and Illinois. These states chose to concentrate on a subset of LEAs that showed greater readiness for reform. This strategy tended to earn slightly fewer points, but it was by no means the only reason these states did not receive awards. Were such states to address their other deficiencies, they likely would score high enough to win in Round 2.

MYTH: Urban-oriented states have an advantage.

FACT: States with heavily rural populations performed quite well among finalists in Round 1. Tennessee, Georgia, South Carolina and Kentucky outscored Ohio, Massachusetts, New York and the District of Columbia.

MYTH: States that did not finish near the top in Round 1 should not bother re-applying.

FACT: Relatively small flaws in many applications prevented significantly higher scores. Moreover, scoring from reviewer to reviewer varied significantly. States should not necessarily assume that their policies will be scored identically in the next round. Lastly, more awards will be given in Round 2 (10-15, by some reports), giving states a better chance to end up in the winner's circle.

Race to the Top Myths and Round 1 Results

▶ Lessons for State Applicants

Findings and Recommendations on the Scoring Process

Appendices

Seven Lessons from Round 1 for States Seeking Grants in Round 2

1 **Reform must reach statewide and beyond the four-year grant period.**

Tennessee and Delaware passed legislation that locked in reforms statewide and will extend beyond the end of the grant in 2014. Writing key reforms into state law brought a range of stakeholders on board early and enabled the states to avoid enlisting local districts and unions one at a time. As a result, these two states presented bold plans that also had broad buy-in. It is important to note that **boldness and state leadership came first.**

2 **Implementation must be certain.**

Reviewers struggled to support plans that had many future contingencies that might or might not happen. They were conscious that **political support and focus can wane.** Tennessee, on the other hand, arrived at its interview with signed letters from all of its 2010 gubernatorial candidates endorsing the state's plan.

3 **Plans must be clear.**

Reviewer comments frequently lament the amount of jargon, padding, and exposition in applications. **Reviewers want concise substance.**

4 **Local advantages are key.**

Winning applications built on their **states' inherent strengths**. For example, Delaware is a small state with relatively few districts, which makes centralized reform more plausible. The state's unique character was very well described as an advantage. Tennessee had a best-in-class data system that could fuel better evaluations and teacher development. The application put the data system front-and-center.

Seven Lessons from Round 1 for States Seeking Grants in Round 2

⑤ Points can be won and lost in unexpected places.

Several states lost big because they had **made insufficient progress on data systems.** Others were penalized for **vague strategies** to equalize distribution of teachers. Two finalists failed to win the 15 "all-or-nothing" **STEM points.** These trends could be make-or-break in Round 2, when more awards will be given and more states will be competitive.

⑥ On "Teachers and Leaders," bold policies are rewarded.

Section D: Great Teachers and Leaders carries the most weight in terms of points. Both winning states have comprehensive statewide approaches—not pilots. However, it has generally gone unnoticed that **Louisiana and Rhode Island** actually earned the top scores in the contest's most critical section. Their applications go even further than the winners in incorporating student achievement into teacher evaluations and clearly linking evaluations to major decisions such as how teachers are hired, developed and retained—**and the scorers approved.**

⑦ Borrow concepts, but do not cut-and-paste.

Reviewers frequently commented on whether a state's application held together thematically. It will be tempting for Round 2 states to simply import the highest-scoring provisions from Round 1 applications. This is likely a mistake. For policies to have meaning, they **must be rooted in a state's local context** and tailored to a larger vision. Further, no state's application was perfect. As we will show later, even the winners' proposals can be improved upon.

Several finalists offer model policy solutions to the most vexing challenges states face in addressing the Great Teachers and Leaders criteria.

Honor Roll
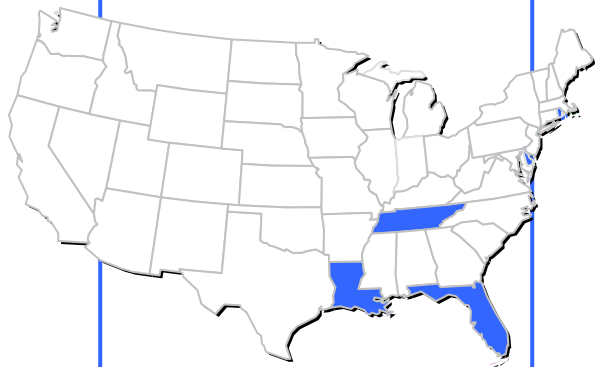
Delaware

Florida

Louisiana

Tennessee

Rhode Island

Challenge	Sample Solutions
Assessing teacher and principal effectiveness based on student academic outcomes	**Tennessee's** "First to the Top" legislation requires that LEAs implement an evaluation system that is based 50% on student achievement data (35%will be based on value-added data, where applicable). **Florida, Louisiana** and **Rhode Island** also submitted plans basing at least 50% of evaluations on student outcomes.
Linking teacher certification and tenure to classroom effectiveness	**Rhode Island** demands evidence of effectiveness for advanced teacher certification and renewal of certification.
Using effectiveness data to inform retention and layoff decisions	**Florida** ensures that when layoffs of teachers and leaders are necessary in participating LEAs, summative evaluations carry more weight than seniority.

Note: Policies described were part of Round 1 applications. Since not all states won, not all policies will necessarily be implemented as proposed.

Model Policy Solutions (continued)

Challenge	Sample Solutions
Improving teacher preparation programs to maximize impact on students	**Louisiana** already measures and publicly ranks the effectiveness of its teacher preparation providers using value-add data. **Rhode Island** will deny program renewal to programs that produce relatively high percentages of ineffective leaders, if they fail to show significant improvement. Rhode Island has *already* closed programs in recent years.
Providing effective professional development (PD)	**Delaware** requires LEAs to adopt a comprehensive PD approach that includes job-embedded, evaluation-aligned PD; common planning time; and participation in a statewide approach to certify and measure PD, including tracking participant and student outcomes.
Ensuring equitable distribution of effective teachers	**Rhode Island** has made significant progress toward mutual consent hiring in all schools by prohibiting teacher assignment on the sole basis of seniority. Rhode Island also forbids any student from having two ineffective teachers in a row. **Louisiana** holds participating LEAs accountable for equitably distributing effective teachers and withholds grant money if goals are not met.

Model Policy Solutions (continued)

Challenge	Sample Solutions
Removing ineffective teachers	**Rhode Island** requires participating LEAs to remove teachers who are rated ineffective for two consecutive years. **Florida** requires all participating LEAs to report annually on the number of ineffective teachers and leaders dismissed for poor performance—as well as the number of highly effective teachers lost by the district.
Using effectiveness to inform compensation	**Florida** requires participating LEAs to base the most significant gains in compensation on effectiveness, as measured by the official evaluation system.
Supporting strong school staffing practices	**Louisiana** requires mutual consent hiring in all participating LEAs and will provide direct assistance to high-poverty schools in staffing practices such as vacancy projections, recruitment, school-based screening and on-boarding.

Even the top-scoring states can continue to improve their strategies.

Recommendations for <u>Delaware</u> on Great Teachers and Leaders

- **Raise the bar for teacher effectiveness:** To be eligible for a Highly Effective rating, a teacher must show student growth that is "greater than one year." This is a low cutoff for the top category—Tennessee requires 1.5 years, or 50% more. Delaware's plan suggests that teachers will be able achieve Effective ratings with less than a year of growth for students. It may be advisable to revisit these policies as Delaware works to define satisfactory levels of student growth by July 2011. Tennessee's benchmarks are a good place to start.

- **Refuse to be satisfied with "basic" performance:** For the five individual components of the Delaware teacher evaluation system, teachers are considered to have achieved a satisfactory result if they are rated Distinguished, Proficient or Basic. Only Unsatisfactory is considered a negative result. However, a Basic rating is not generally considered an acceptable ongoing performance level for an experienced teacher. It would be more rigorous to demand Proficient or Distinguished work from seasoned teachers.

- **Institute annual evaluations:** Delaware does not propose full annual evaluations for all experienced teachers. Those rated Highly Effective or Effective will receive a single announced observation annually and a full evaluation every two years. While student growth is to be monitored annually, infrequent evaluations for veteran teachers send the message that continuous feedback and improvement are not top priorities. We recommend comprehensive annual evaluations for all teachers.

- **Do not tolerate consistently poor performance:** Delaware law directly addresses removal for poor performance and says that teachers with two consecutive Ineffective ratings MAY be removed—but doesn't guarantee it. We recommend that any teacher who is demonstrably ineffective for two years not be permitted to continue in the classroom. Rhode Island's application has a strong policy to this effect.

Even the top-scoring states can continue to improve their strategies (continued).

Recommendations for <u>Tennessee</u> on Great Teachers and Leaders

- **Stay focused:** Tennessee passed groundbreaking "First to the Top" legislation just prior to submitting its application, but much of the substance behind the act is yet to be defined. A key state-level committee will design a new evaluation system and give guidance on critical issues by July, 2011. Over a 15 month period, the committee must maintain the same focus and momentum that allowed Tennessee to submit a top-notch application. It must be insulated from political pressure to water down its plan now that it has won.

- **Build out the compensation plan:** Tennessee's legislation allows LEAs to develop modified salary schedules and requires evaluation to be used as a component of compensation. However, it does not say much more. We advise adoption of a compensation strategy that is as well grounded as Tennessee's educator evaluation plan. Florida's application may be a starting point—the state would require LEAs to base the "most significant" gains in teacher salary on effectiveness, as measured by the official evaluation system.

- **Take action to guarantee equal access to effective teachers:** Tennessee's plans to improve the distribution of effective teachers to high-need schools are less robust than most of the state's other plans. In fact, Tennessee received just 10 out of 15 possible points for that section, tied for lowest among the 16 finalists. Tennessee should adopt two policies included in Rhode Island's plan, which earned the highest score on equitable distribution. First, permit schools to select their teachers, ending forced-placement practices that often funnel the least effective teachers to schools serving poor children. Second, demand that no child in a participating LEA be taught by two ineffective teachers in a row, ensuring that equitable distribution reaches all the way to the student level.

- **Set clear consequences for ineffective teaching:** Tennessee will soon have a much-improved evaluation system, but it is not necessarily the case that a better evaluation system will lead to better approaches for the removal of ineffective teachers. Tennessee could have done more to spell out the specific changes to dismissal procedures that will overcome a historic tendency not to confront poor performance. We recommend that Tennessee consider adopting Rhode Island's policy of requiring LEAs to release from employment any teacher who is rated ineffective for two consecutive years.

Race to the Top Myths and Round 1 Results

Lessons for State Applicants

▶ **Findings and Recommendations on the Scoring Process**

Appendices

The Department of Education should be commended for the overall transparency, rigor and integrity of the competition.

Structural Strengths

- **Credibility:** The Department resisted what was likely considerable pressure to make decisions based on politics, and in the process proved wrong many commentators who assumed that a multi-billion dollar grant program could not be run with integrity.

- **Clarity and vision:** The contest guidelines were exceptionally clear and forward-looking, representing a coherent push for educational improvement, not a bloated laundry list of pet projects. Program administrators did an outstanding job of resisting calls to water down program requirements prior to issuing final guidance.

- **Focus:** Significant points were allotted to some of the most difficult priorities. In particular, Great Teachers and Leaders was worth 138 of 500 total points. For decades, states and districts have opted to deal with reforms that are less controversial and difficult than improving the quality of educators. Race to the Top sent a message that those days are over.

- **Directness:** Application reviewers had the opportunity to question state leaders in person, leading to unprecedented displays of political will from the states.

- **Transparency:** All scores and comments were released publicly—an impressive degree of transparency, especially given the difficult position in which low-scoring states could be placed. Making detailed scores available opens the Department to scrutiny—including from the authors of this analysis—but it was unquestionably the right thing to do.

- **Selectivity:** Just two awards were given in the first round, which keeps the bar high and will result in greater overall reform.

However, TNTP's analysis of the scoring of the Round 1 finalist applications reveals significant concerns about the scoring process.

Areas of Concern

- **Lack of Differentiation:** Overall, the scoring of finalists' applications produced relatively minimal differentiation. Reviewers' scores often did not reflect significant variations between states in the ambition, rigor and feasibility of their proposed reforms. Given the variation in quality across applications, the lack of score differentiation is problematic. States with particularly ambitious proposals were disadvantaged to a degree because peer states earned nearly as many points for weaker plans.

- **Inflated Scores:** Reviewers were given guidance on "high," "medium," and "low" score ranges for most criteria. However, when rating finalist applications, they almost always assigned "high" points and almost never assigned "low" points—even when the application appeared to have meaningful weaknesses. The result was 16 applications scoring over 400 points out of a possible 500. This appears to have been far too many.

- **Deviation from Scoring Guidance:** The challenges above were exacerbated by uneven scoring guidance. Yet even when reviewers were given clear guidance, they sometimes deviated from it. Some of these deviations appear to have gone unaddressed.

- **Excessive Influence of Outliers:** Despite the general lack of differentiation in scoring, there were instances of wide variance in reviewer agreement across applications. Review panels for some states assigned fairly similar scores; other states' review panels included significant outliers. These outliers were enough to change the final rank of a state's application—meaningfully, in a few cases. A scoring process that discarded the highest and lowest scores awarded to each state would have significantly changed scoring outcomes for some states.

Finalist states consistently received "high" scores in each selection criterion and almost never received "low" scores, resulting in minimal differentiation and significant inflation.

(A)(2)(i) – "Ensuring the Capacity to Implement"

(B)(3) – "Supporting the Transition to Enhanced Standards and Assessments"

(D)(2)(ii) – "Developing Evaluation Systems"

(E)(2)(ii) – "Turning Around the Lowest -Achieving Schools"

All 16 finalists received "high" scores in the sections identified above.

Criterion	DE	TN	GA	FL	IL	SC	PA	RI	KY	OH	LA	NC	MA	CO	NY	DC
(A)(1)(i)	5	5	5	5	4	4	5	5	5	5	5	5	5	4	5	4
(A)(1)(ii)	45	44	39	36	39	40	44	35	45	39	40	44	36	37	37	34
(A)(1)(iii)	15	14	9	9	10	12	11	11	14	11	10	14	14	11	12	12
(A)(2)(i)	18	18	17	17	15	16	18	17	19	19	16	16	18	18	17	16
(A)(2)(ii)	10	10	6	7	9	7	9	8	10	10	9	9	9	9	9	6
(A)(3)(i)	5	5	4	5	4	5	4	5	4	5	4	4	5	5	5	5
(A)(3)(ii)	22	16	22	22	13	17	16	19	17	13	18	16	24	12	20	17
(B)(1)(i)	20	20	20	20	20	20	18	19	20	20	19	20	20	20	20	20
(B)(1)(ii)	20	20	18	20	20	20	20	19	20	20	19	20	5	20	19	20
(B)(2)	10	10	10	10	10	10	10	10	10	10	9	10	10	7	9	10
(B)(3)	19	19	17	19	19	19	18	18	19	20	18	17	19	16	17	18
(C)(2)	5	5	5	4	4	5	5	4	5	4	4	3	5	5	5	4
(D)(1)	17	15	15	16	19	16	12	18	20	16	19	17	19	14	20	18
(D)(2)(i)	4	5	2	4	4	4	3	4	4	5	5	5	4	4	4	5
(D)(2)(ii)	13	14	12	11	13	13	13	14	12	13	14	12	12	12	12	14
(D)(2)(iii)	9	10	10	8	7	9	10	10	6	9	9	8	6	9	9	8
(D)(2)(iv)	24	24	25	24	23	26	23	26	21	23	25	24	17	22	17	20
(D)(3)(i)	12	10	14	11	11	10	13	14	11	10	13	13	10	11	13	12
(D)(3)(ii)	9	8	8	7	7	8	6	6	7	4	9	7	7	7	8	8
(D)(4)	11	13	10	11	10	11	9	13	11	9	12	10	9	12	12	12
(D)(5)	19	15	15	16	16	16	16	17	18	14	17	17	17	14	17	16
(E)(2)(i)	4	5	5	5	5	5	5	5	5	4	5	5	5	5	4	5
(E)(2)(ii)	28	33	32	29	34	29	31	30	31	29	35	30	32	31	29	35
(F)(1)	8	9	9	10	9	8	9	4	10	10	10	8	10	5	8	10
(F)(2)	31	30	37	39	35	29	31	31	8	34	34	23	29	40	27	40
(F)(3)	4	5	4	5	5	4	5	5	5	5	4	4	5	4	4	5

Point Ranges for Scores

- 🟩 **High**
- 🟨 **Medium**
- 🟥 **Low**

Point ranges provided by USDE as part of scoring guidance for reviewers. **See _Appendix 1_ for details**.

Note: Data presented on this slide do not include criteria (C)(1), (C)(3), or (E)(1), which were not scored on the "high"/"medium"/ "low" continuum. Scores are rounded to the nearest whole number.

When reviewers were given highly specific scoring guidance for particular criteria, the states' scores showed much healthier differentiation.

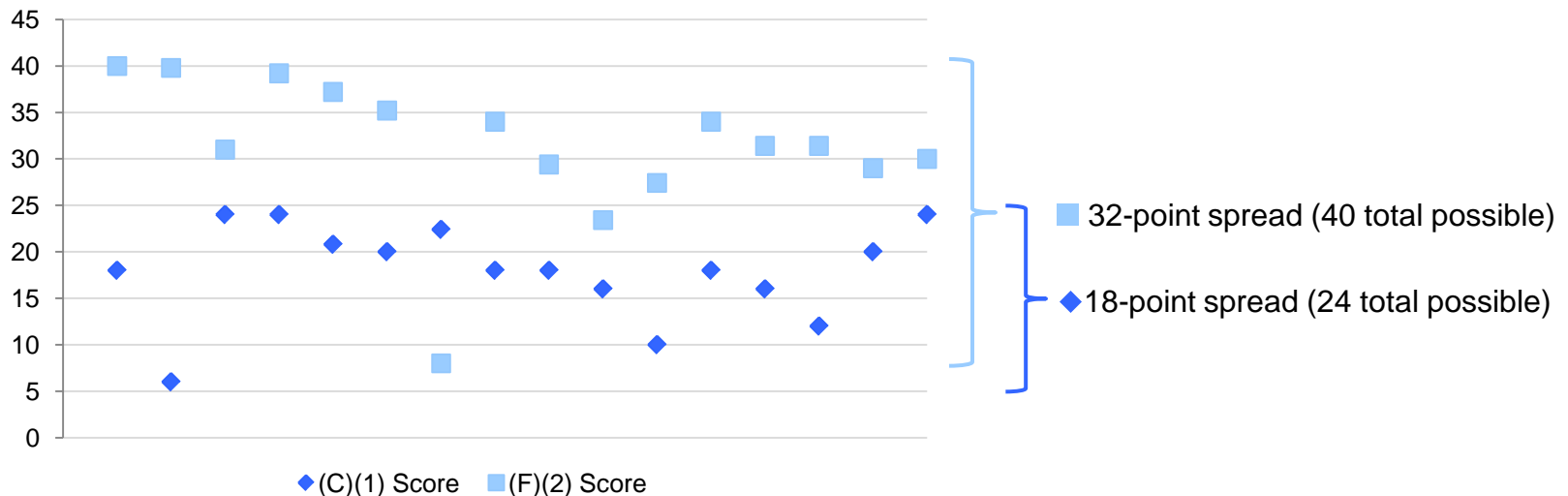## Sample Criteria with Highly Specific Scoring Guidance

◆ **(C)(1)—Fully implementing a statewide longitudinal data system (*24 points max*)**

Two points for each of the 12 elements of America COMPETES in place in the state.

■ **(F)(2)—Ensuring successful conditions for high-performing charter schools and other innovative schools (*40 points max*)**

"High"/"medium"/"low" points based on specific numeric guidelines related to the number of schools allowed under a state's charter cap and the per-pupil funding available to charter schools.

Finalist State Scores on (C)(1) and (F)(2)

■ 32-point spread (40 total possible)

◆ 18-point spread (24 total possible)

◆ (C)(1) Score   ■ (F)(2) Score

In contrast, there was minimal differentiation in scores when the scoring guidance was ambiguous.

Section D: Great Teachers and Leaders was an area in which the degree of difficulty was expected to sharply distinguish states with the strongest plans.

Yet in all the Section D sub-criteria, the only scoring guidance provided was the general direction to award "high"/"medium"/"low" points based on the quality of the state's plan and the extent to which it addressed both teachers and principals. **This was a lost opportunity,** and had the effect of reducing the importance of what appeared to be a crucial priority for the contest.

Finalist State Scores on Section D

■ 15-point spread (58 points possible)

◆ 8-point spread (21 points possible)
▲ 9-point spread (25 points possible)
● 4-point spread (14 points possible)
■ 5-point spread (20 points possible)

◆ (D)(1) Score   ■ (D)(2) Score   ▲ (D)(3) Score   ● (D)(4) Score   ■ (D)(5) Score
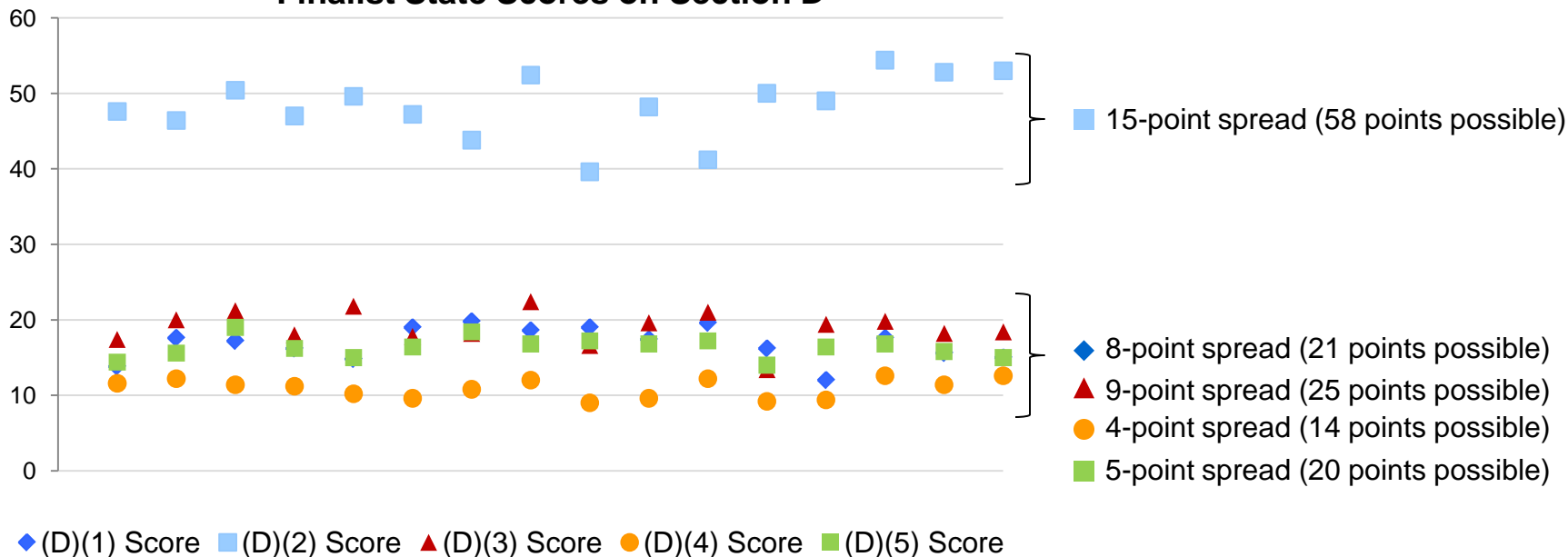
For instance, the scoring of (E)(2)(ii): Turning Around the Lowest-Achieving Schools shows how reviewers sometimes struggled to meaningfully differentiate the relative strengths of finalists' proposals.

Is it really *this easy* to turn around schools?

The significant challenge posed by turning around low-achieving schools is well known, and is evidenced by the weighting of (E)(2)(ii) at 35 points, one of the highest possible point totals of any single RTTT criterion. One would expect that this would be an area in which a small number of the most aggressive states could distinguish themselves.

However, the District of Columbia and Louisiana each received perfect scores in (E)(2)(ii), and four more states received 90% or more of the available total. Moreover, 31 individual reviewers awarded perfect scores in (E)(2)(ii) to 15 different finalist states.

Percent of Total Points Received, (E)(2)(ii): Turning Around the Lowest-Achieving Schools
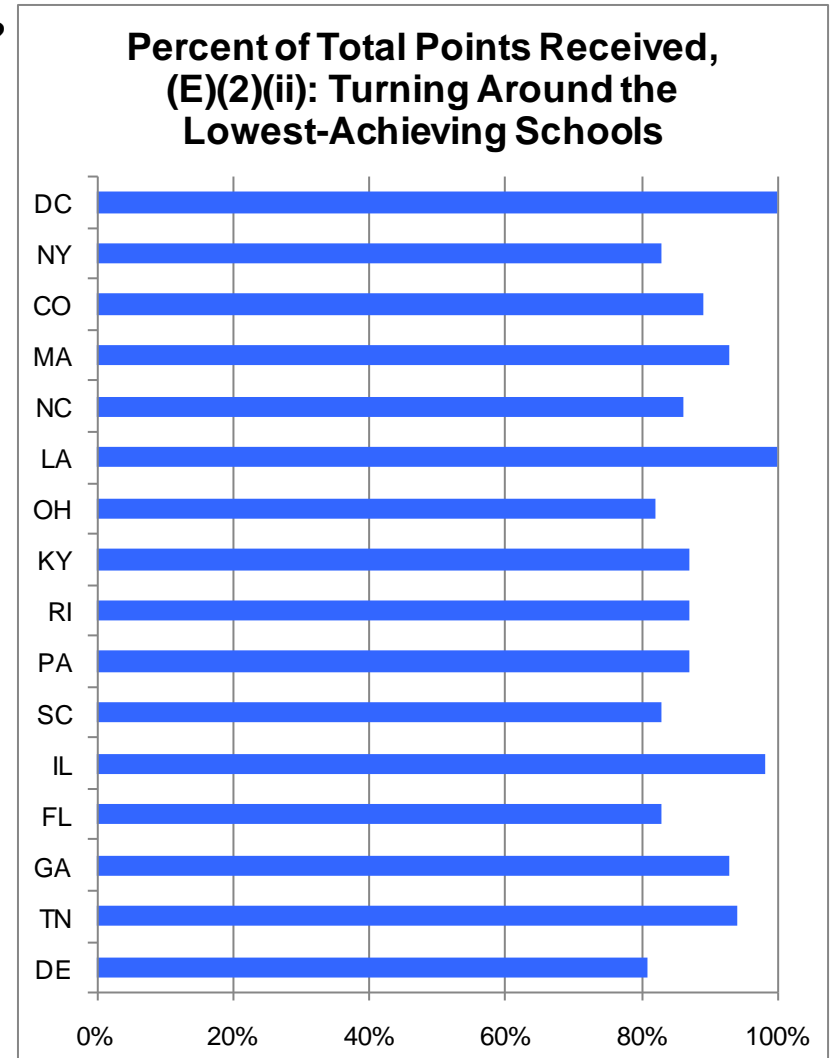
Likewise, reviewers tended to be very generous in their assessment of how states have improved student outcomes—in contrast with external assessments that are far more cautious.

Top (A)(3)(ii) Scores		Independent Analysis of Actual State Performance		
	State Score (out of 25) (*individual reviewer scores*)	**Closing Achievement Gaps** (Source: Ed Trust analysis of 2003-2007 NAEP scores in "Gauging the Gaps")	**Raising Achievement** (Source: Ed Trust analysis of 2003-2007 NAEP scores in "Gauging the Gaps")	**Increasing High School Graduation Rates** (Source: EdWeek analysis of 2006 graduation rates vs. 1996 figures in "Diplomas Count" 2009)
Delaware	**22.2** (*25, 20, 23, 20, 23*)	🟡	🟡	🟡
Florida	**22.4** (*22, 20, 22, 23, 25*)	🟡	🟡	🔴
Georgia	**21.8** (*22, 18, 19, 25, 25*)	🟡	🟡	🔴 ←
Massachusetts	**24.2** (*25, 23, 25, 23, 25*)	🔴	🟡	🔴
New York	**20.4** (*18, 19, 23, 17, 25*)	🟡	🟡	🟢

Georgia's 2006 graduation rate was **55.9%,** an increase of only 0.7 percentage points since 1996.

Source: Editorial Projects in Education Research Center and *Education Week.* Diplomas Count 2009.
http://www.edweek.org/media/ew/dc/2009/33sos_gains.pdf
The Education Trust (2010). Gauging the Gaps.
Washington, D.C.
http://www.edtrust.org/sites/edtrust.org/files/publications/files/NAEP%20Gap_0.pdf

🟢 Statistically significant progress across subgroups and subjects

🟡 Some statistically significant progress among specific subgroups/subjects

🔴 No statistically significant progress

🟢 Statistically significant progress across subgroups and subjects

🟡 Some statistically significant progress among specific subgroups/subjects

🔴 No statistically significant progress

🟢 Rate of increase above U.S. average
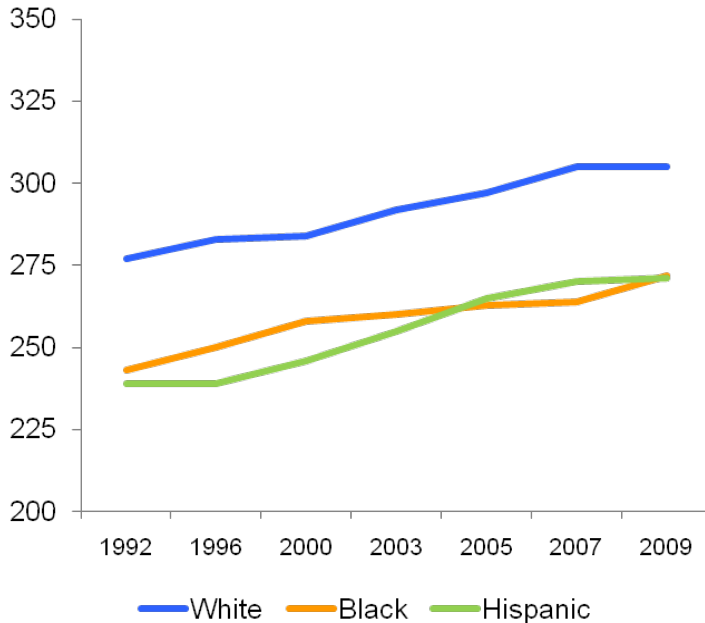
🟡 Rate of increase near U.S. average

🔴 Flat or minimal rate of increase

One finalist, Massachusetts, earned 24.2 of 25 points for criterion (A)(3)(ii): Improving Student Outcomes. Yet while scores in the state have improved overall, it has made little progress in closing its achievement gap.

NAEP Average Scale Scores: Massachusetts Grade 8 Mathematics

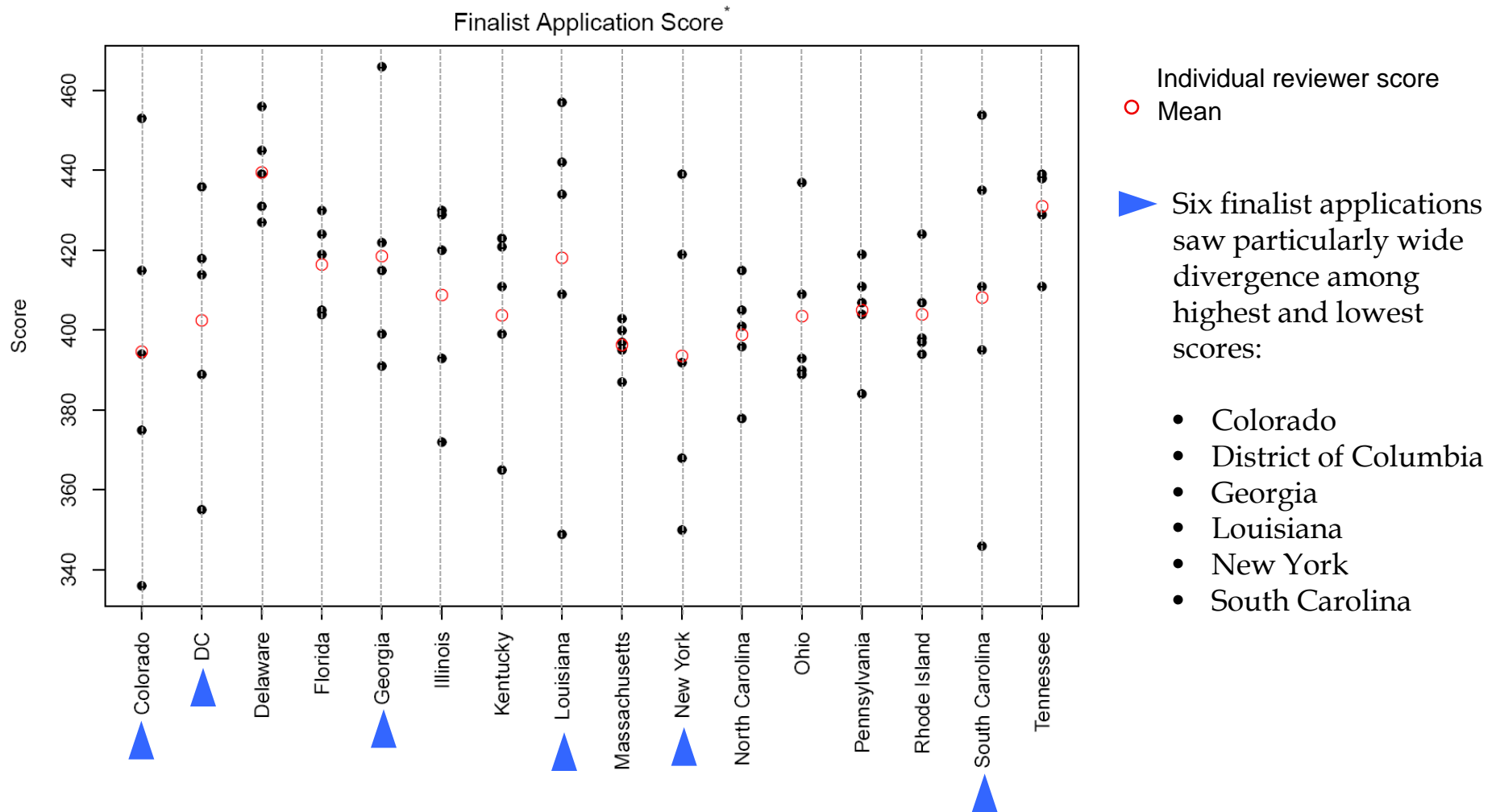**NAEP Average Scale Scores: Massachusetts Grade 8 Reading**
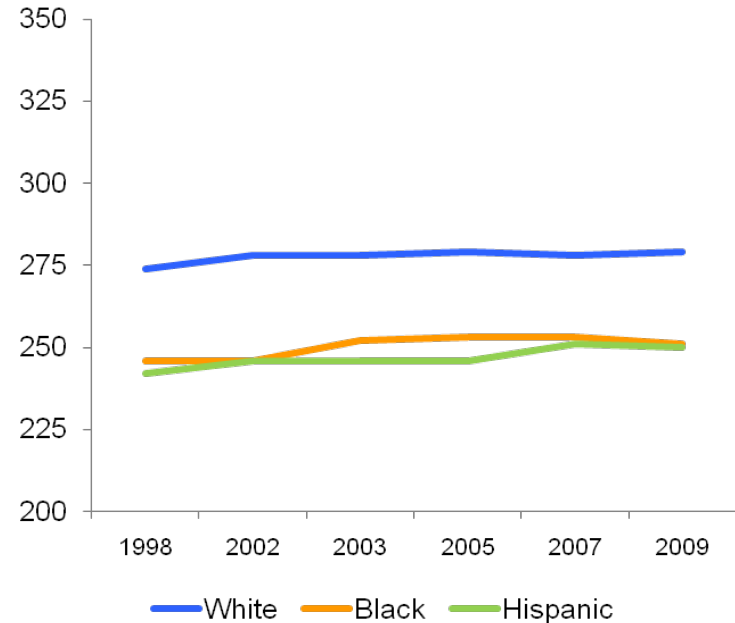
Summary: There was no significant closing of the gap in MA for Black, Hispanic, or poor students on math. See details at: http://nces.ed.gov/nationsreportcard/pdf/stt2009/2010454MA8.pdf

Summary: There was no significant closing of the gap in MA for Black, Hispanic, or poor students on reading. See details at: http://nces.ed.gov/nationsreportcard/pdf/stt2009/2010460MA8.pdf

In addition to inflating scores generally, reviewers often diverged wildly in their scoring, despite looking at the same materials. Because all scores were averaged, these outlier scores shaped final standings.
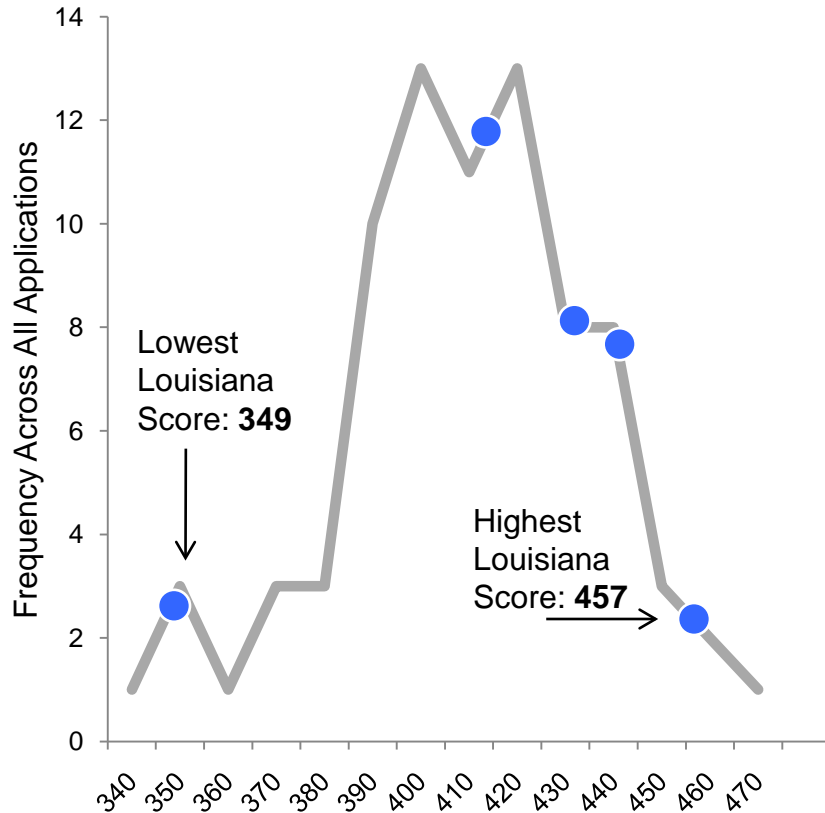
Finalist Application Score*

Individual reviewer score
○ Mean

▶ Six finalist applications saw particularly wide divergence among highest and lowest scores:

- Colorado
- District of Columbia
- Georgia
- Louisiana
- New York
- South Carolina

* Scores do not include all-or-nothing allocations for the STEM competitive priority.

Eleventh-place Louisiana was held back by a single score that was the third-lowest for any finalist; third-place Georgia benefitted significantly from the highest score assigned to a finalist.

Louisiana Reviewer Scores Compared to Distribution of All Finalist Scores

Lowest Louisiana Score: **349**

Highest Louisiana Score: **457**

Frequency Across All Applications

= LA Application Score

Georgia Reviewer Scores Compared to Distribution of All Finalist Scores

Lowest Georgia Score: **391**

Highest Georgia Score: **466**

= GA Application Score

* Scores do not include all-or-nothing allocations for the STEM competitive priority.
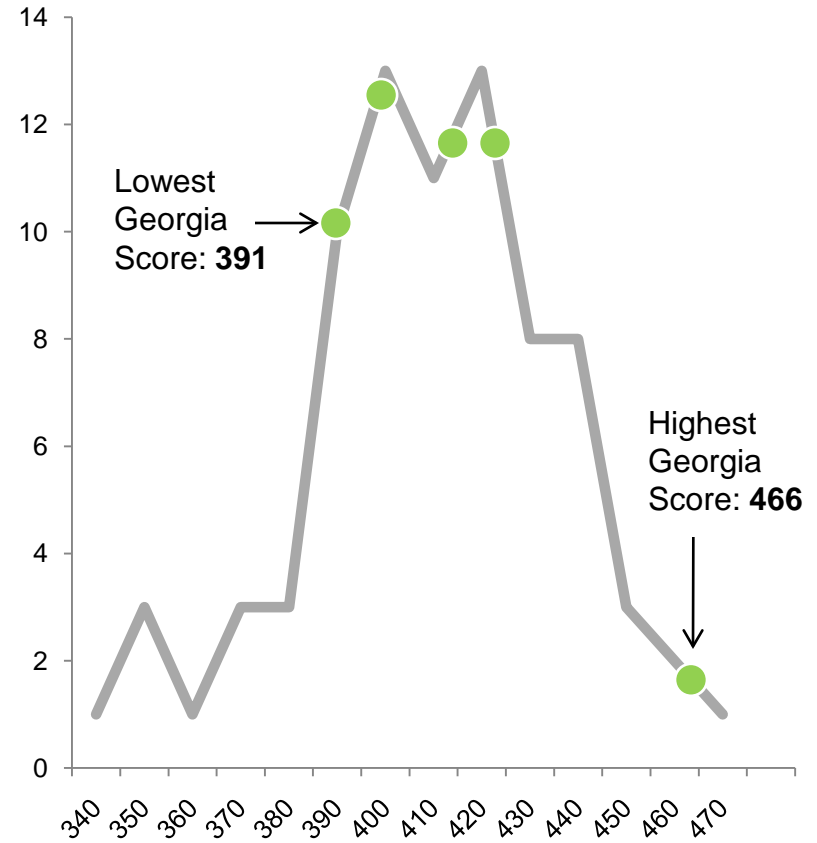
Compounding the problem, reviewers sometimes awarded points in a manner inconsistent with the scoring guidance, a problem which can be easily addressed through tighter oversight in Round 2.

▶ **Example 1:** Under criterion (C)(1): Fully Implementing a Statewide Longitudinal Data System, the scoring guidance directed reviewers to award states 2 points for every element of America COMPETES that they currently have in place. In its application, Georgia reported having only 10 of the 12 elements of America COMPETES, yet one reviewer awarded Georgia 24 points for (C)(1) and cited Georgia's completion of the Data Quality Campaign's 10 Essential Elements as evidence for the perfect score. *Note: This same reviewer gave Georgia an overall score of 466 points (prior to awarding points for STEM), 44 points more than any other reviewer of Georgia's application and the single highest score given by a reviewer to any finalist state.*

▶ **Example 2:** Under criterion (E)(1): Intervening in the Lowest-Achieving Schools and LEAs, the scoring guidance directed reviewers to award scores of 10, 5, or 0 points, depending upon the state's legal authority to intervene in low-achieving schools and LEAs. Reviewers for Colorado and Kentucky each awarded 8 points for (E)(1) despite the legal authority of both states to intervene in schools and LEAs, and despite the scoring guidance that did not designate 8 points as a valid score for (E)(1).

A scoring system in which the highest and lowest scores for each state were discarded would have resulted in significantly different outcomes for some states, most notably Louisiana and Georgia.

Actual RTTT Rankings	Actual RTTT Final Score	Hypothetical Score After Discarding the Highest and Lowest Scores	Change in Final Score	Change in Final Ranking	Hypothetical New Ranking
1. Delaware	454.6	**453.3**	(-1.3)	(+0)	*1*
▲ 2. Tennessee	444.2	**450**	(+5.8)	(+0)	*2*
▼ 3. Georgia	433.6	**427**	(-6.6)	(-4)	*7*
4. Florida	431.4	**431**	(-0.4)	(+1)	*3*
▲ 5. Illinois	423.8	**429**	(+5.2)	(+1)	*4*
▲ 6. South Carolina	423.2	**428.7**	(+5.5)	(+1)	*5*
7. Pennsylvania	420	**422.3**	(+2.3)	(-2)	*9*
8. Rhode Island	419	**415.7**	(-3.3)	(-2)	*10 –tie*
▲ 9. Kentucky	418.8	**425.3**	(+6.5)	(+1)	*8*
▼ 10. Ohio	418.6	**412.3**	(-6.3)	(-2)	*12–tie*
▲ 11. Louisiana	418.2	**428.3**	(+10.1)	(+5)	*6*
12. North Carolina	414	**415.7**	(+1.7)	(+2)	*10–tie*
13. Massachusetts	411.4	**412.3**	(+0.9)	(+1)	*12–tie*
14. Colorado	409.6	**409.7**	(+0.1)	(+0)	*14*
15. New York	408.6	**408**	(-.6)	(+0)	*15*
16. DC	402.4	**407**	(+4.6)	(+0)	*16*

▲▼ **Indicates potential gain/loss of 5 points or more in final score.**

Recommendations to the Department of Education to Improve Scoring Precision and Reliability in Round 2

Recommendation	But Keep in Mind…
1. Round 1 Ratings Assessment Conduct a full assessment of each reviewer's ratings and comments in Round 1 to determine fidelity to rubric. Reviewers with significant instances of inaccuracy or deviation from guidance should not be invited to review again.	The Department of Education must avoid the perception that any reviewer is being penalized for political reasons. Using a comparative analysis of scoring should prevent even an appearance of impropriety.
2. Enhance Reviewer Training and Norming Reviewers selected for Round 2 should receive significantly more training—at least three full days—and should participate in norming sessions with other reviewers. Particular attention should be paid to: • Treating each application section independently, and not allowing impressions from one item to influence judgment of another. • Using a consistent approach to awarding points, whether that is starting from full credit and deducting for deficiencies or starting at zero and building up. • Assigning all levels of points appropriately and distinguishing between proposal elements of "high," "medium," and "low" quality in each section. • Justifying scores exclusively with evidence rooted in rubric, not considering external information or personal beliefs.	Reviewers must still have latitude to make professional judgments. Additional norming should not have the effect of unduly pressing reviewers to endorse policies or points of view that are not consistent with the contest scoring rubric.

Recommendations to the Department of Education to Improve Scoring Precision and Reliability in Round 2 (continued)

Recommendation	But Keep in Mind…
3. Regular Monitoring of Reviewers Processes for monitoring work by reviewers should be made more robust. For example, professional staff at the Department of Education should regularly spot-check reviewer work and offer feedback. In cases where scores for the same application differ widely across reviewers, the Department of Education should assign two additional reviews, for a total of seven.	Reviewers assigned after the start of deliberations will miss some dialogue. For this reason, they must be particularly capable.
4. Establish Scoring Controls There should be basic score controls across applications. Even when the five reviewers on a panel agree on scores for a given state, there should be procedures to ensure that their scoring approach was comparable to the approach taken by other panels. The same policies and accomplishments should score similarly across states. Each application should be benchmarked against 1-2 others for the sake of comparison. Reviewers should be given an opportunity to reconsider their scores accordingly.	Norming across applications will take additional time. The process cannot become so involved that it loses focus.
5. Drop Outlier Reviews Outlier reviews should not have an outsized effect on final scores. The Department of Education should automatically drop the highest and lowest final scores from each panel. If two extra reviewers are added, the two highest and two lowest scores could be dropped.	Reviewers whose opinions differ from colleagues should not feel like their feedback is not valuable because their score could be dropped. The process must allow for reviewers with sound arguments to persuade others.
6. Remember Secretarial Discretion Though the opinion of application reviewers must be weighted heavily in final decisions, Secretary Duncan retains final authority over awards and should be prepared to intervene in Round 2 if best attempts to strengthen the process are insufficient.	There could be enormous political fallout if the Secretary is forced to deviate from peer scores in selecting winners. This is a last resort.

Race to the Top Myths and Round 1 Results

Lessons for State Applicants

Findings and Recommendations on the Scoring Process

▶ **Appendices**

Appendix 1: "High"/"Medium"/"Low" Scoring Ranges Provided in the Official RTTT Scoring Guidance

Criterion Point Value	Quality of Applicant's Response		
	"Low"	"Medium"	"High"
45	0 – 12	13 – 33	34 – 45
40	0 – 10	11 – 29	30 – 40
35	0 – 9	10 – 25	26 – 35
30	0 – 8	9 – 21	22 – 30
25	0 – 7	8 – 18	19 – 25
21	0 – 5	6 – 15	16 – 21
20	0 – 5	6 – 14	15 – 20
15	0 – 4	5 – 10	11 – 15
14	0 – 4	5 – 9	10 – 14
10	0 – 2	3 – 7	8 – 10
7	0 – 2	3 – 4	5 – 7
5	0 – 1	2 – 3	4 – 5

Appendix 2: Top-Scoring Applications, by Section

Section	Section Topic	Top-Scoring States
A	State Success Factors	1. **Delaware** 2. Kentucky 3. **Tennessee**
B	Standards and Assessments	T1. Ohio T1. Florida **T1. Delaware** T1. Illinois
C	Data Systems to Support Instruction	1. **Delaware** 2. **Tennessee** 3. Kentucky
D	Great Teachers and Leaders	1. Louisiana 2. Rhode Island 3. **Delaware**
E	Turning Around the Lowest-Achieving Schools	1. DC 2. Illinois 3. **Tennessee**
F	General	1. DC 2. Florida 3. Georgia

For More Information

www.tntp.org