

Approaches in the United States to Benchmarking at the Subnational Level

Valena Plisko

National Center for Education Statistics (NCES)

Institute of Education Sciences

U.S. Department of Education

Presentation at the IEA General Assembly Meeting

October 2009

U.S. Interest in Subnational Benchmarking to International Standards Renewed

- High levels of interest in U.S. states for international benchmarking
 - Reports issued by national associations of state education leaders and governors (e.g., *Benchmarking for Success* white paper)
 - Contacts from states interested in obtaining PISA or TIMSS scores
- In spite of all the testing already required of states
 - All students in grades 3 through 8 in mathematics and reading
 - NAEP (National Assessment of Educational Progress) now mandatory for states

Economic Impetus

- States feeling economic competition from abroad

“Governors recognize that new economic realities mean it no longer matters how one U.S. state compares to another on a national test; what matters is how a state’s students compare to those in countries around the globe.” (*Benchmarking for Success* white paper)

“. . . the people of Ohio are sending us a clear message to . . . focus on the core issues . . . creating living-wage jobs, building an education system, from pre-school through college, that doesn’t just compete with our neighbors like Indiana and Kentucky, but rivals the best schools in the world...”(Gov. Ted Strickland, on election victory)

Not the First Time U.S. Interested in State Benchmarking

- In 1999, U.S. funded TIMSS participation for 13 states and 14 school districts and consortia of school districts for the same reasons states are asking to participate now:
 - Benchmarking achievement against international peers
 - Assessing rigor and effectiveness of educational programs in an international context
- Since 1999, U.S. has not funded subnational participation in international studies, and few states and districts have invested resources to obtain scores on their own
 - Exceptions are Indiana (TIMSS 2003), Massachusetts (TIMSS 2007), and Minnesota (TIMSS 2007)
 - Consortium of districts in southwest Pennsylvania used their 1999 TIMSS results to redesign curricula

New Dimension: Benchmarking “Standards”

- State interest in benchmarking standards
- One method of benchmarking standards is to compare assessments:
 - How does “basic” or “proficient” or “advanced” in my state compare with “low” or “intermediate” or “advanced” in other countries?
 - What percentage of students in my state reach various international benchmarks? How does that compare with other countries?
- Federal interest in common standards that are higher, clearer, fewer, and internationally benchmarked
 - Internationally benchmarked standards one of four core reforms in “Race to the Top” portion of American Recovery and Reinvestment Act of 2009

Current U.S. Efforts

Four potential approaches to leverage national participation in international assessments:

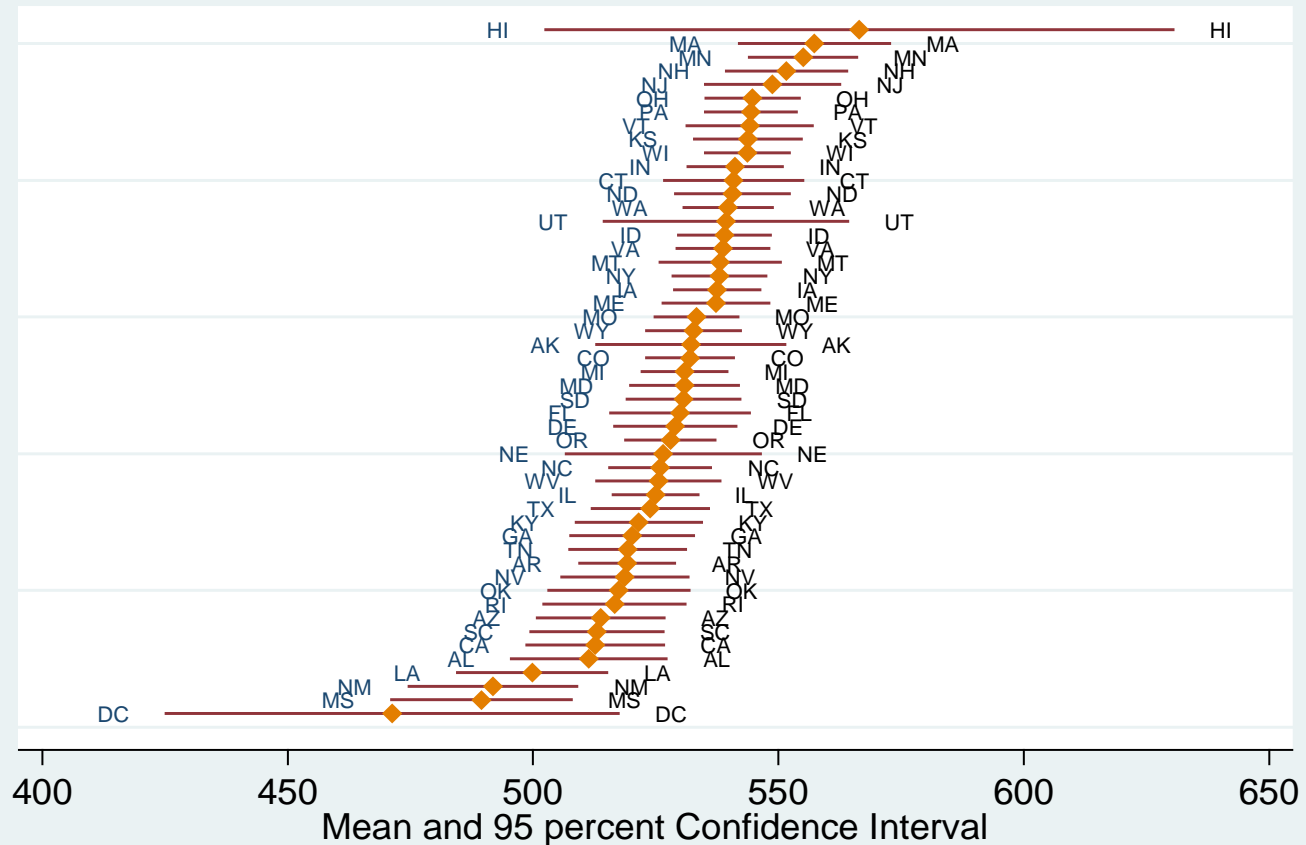
1. Small Area Estimation: leveraging small TIMSS samples in states with other, more broadly available data to estimate TIMSS equivalent scores
2. Spiraling Items: embedding international items in state assessments
3. Statistical Moderation: converting NAEP (National Assessment of Educational Progress) levels into TIMSS benchmarks
4. Spiraling Blocks: embedding blocks of TIMSS items into NAEP booklets

Small Area Estimation

- Method:
 - Use national TIMSS sample to estimate regression coefficients for variables available for all U.S. schools (e.g., school mean TIMSS score = function of school demographics, school score on state assessment, state score on NAEP, etc.)
 - Use regression coefficients to estimate TIMSS equivalent scores for all schools
 - Calculate state mean scores; estimate standard errors by repeated simulations
- Assumptions:
 - Relationships between explanatory variables and scores are consistent across states (e.g. relative performance by race/ethnicity the same in Hawaii and Illinois)
- Potential:
 - Will not produce precise state results in any but the largest states, but may help in designing more efficient samples

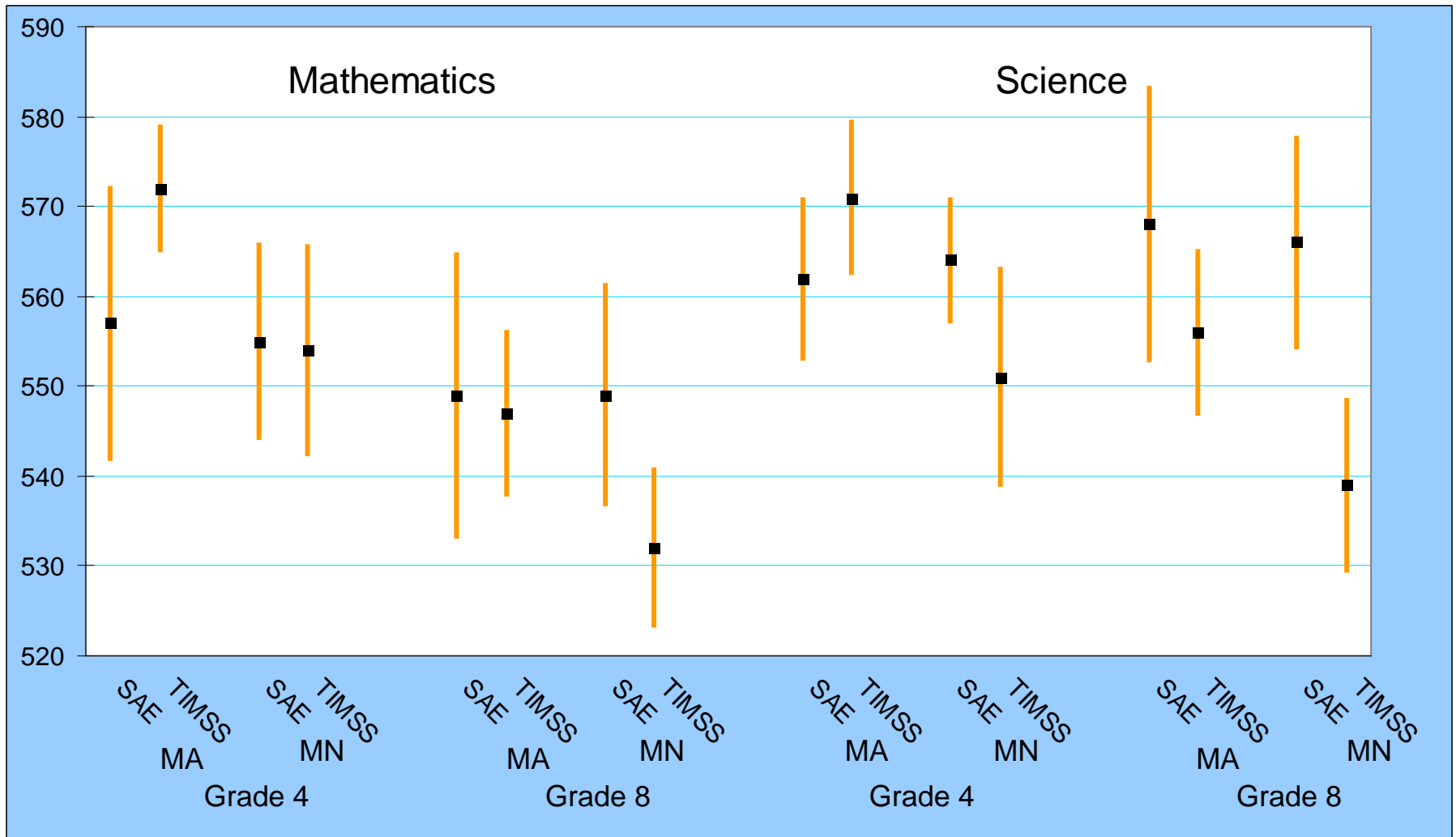
Small Area Estimation Preliminary Results

Estimated TIMSS State Means and Confidence Intervals
4th Grade Mathematics



SOURCE: Sherman, Dan. (in process).

Small Area Estimation Preliminary Results: Comparisons for States in TIMSS 2007



Spiraling Items

- Method:
 - Embed international items in state assessments.
 - Compare item parameters obtained in state assessment context with item parameters from the standard international assessment to calculate a linking equation between the state and international assessment
 - Use linking equation to estimate an equivalent score (linking error expected to be large)
 - Use equivalent score to benchmark states' proficiency levels against international benchmarks and estimate percentage of students reaching international benchmarks
- Assumptions:
 - Selected items represent full assessment
 - International items have similar parameters in two different settings (high stakes state assessment and low stakes international assessment)
 - International assessment bodies okay use of items
- Potential:
 - Unknown, still very early in development.

Statistical Moderation

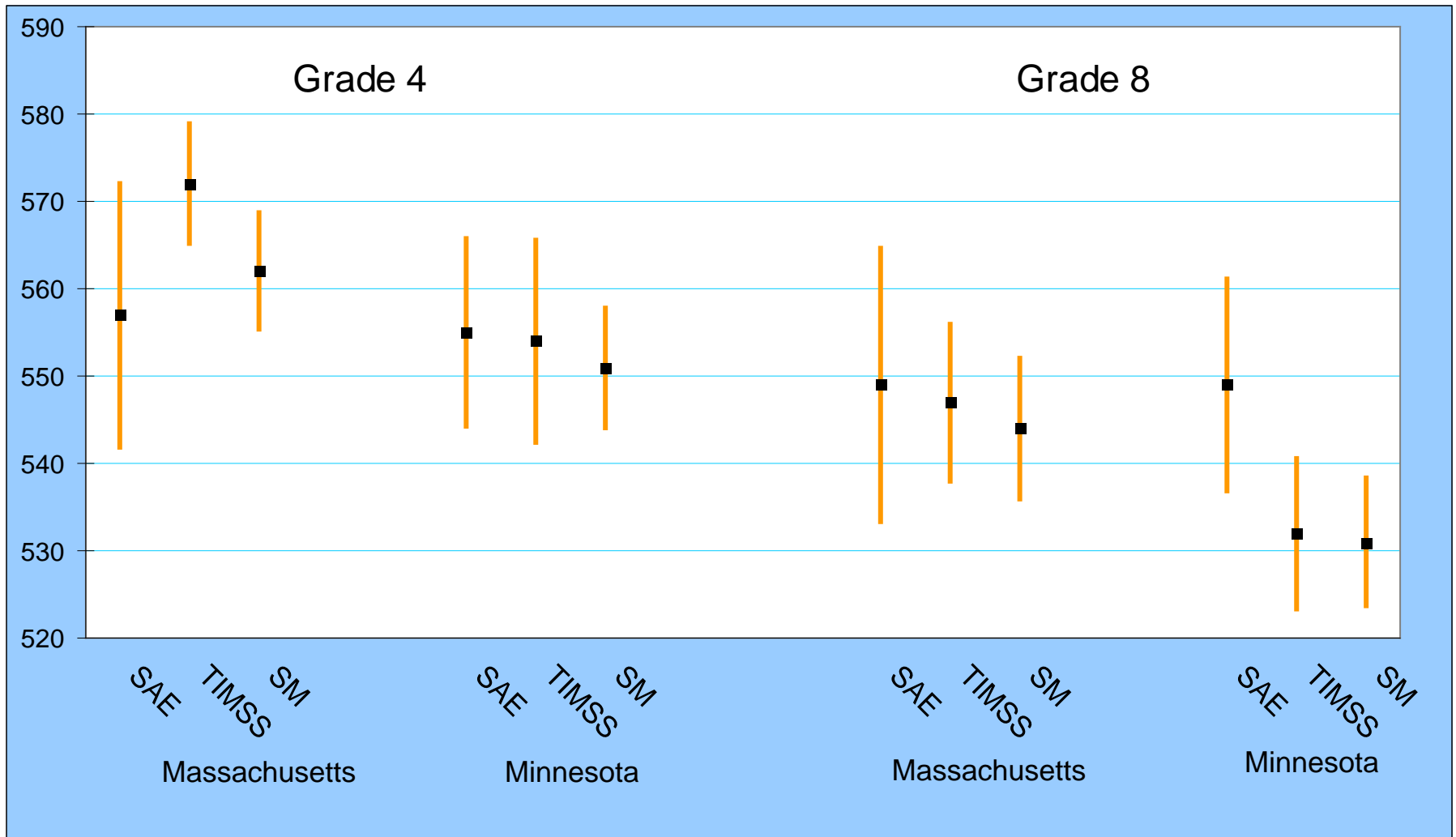
- Method:
 - NAEP-TIMSS: Convert mean and standard deviation of NAEP scale to TIMSS scale
 - State-TIMSS: Convert state to NAEP, then NAEP to TIMSS
 - Use obtained TIMSS equivalent scores to benchmark NAEP and state proficiency levels against TIMSS international benchmarks and estimate percentage of students reaching international benchmarks
- Assumptions:
 - TIMSS, NAEP, and state assessments are all assessing similar things
- Potential:
 - Gary Phillips (American Institutes for Research) has published several reports using this method
(see <http://www.air.org/news/documents/AIRInternationalBenchmarks2009.pdf>)

Statistical Moderation Results

Table 3: International Grades for States in 2007 Mathematics, Grade 4^a

State	Estimated TIMSS Mean	International Grade
Massachusetts	572	B
Minnesota	554	B
New Jersey	552	B
New Hampshire	552	B
Kansas	551	B
Vermont	546	B-
North Dakota	544	C+
Indiana	543	C+
Ohio	542	C+
Wisconsin	541	C+
Pennsylvania	540	C+

Statistical Moderation Results: 2007 Comparisons with TIMSS and Small Area Estimation Results



Spiraling Blocks

- Method:
 - Embed TIMSS blocks into NAEP booklets
 - Compare item parameters obtained in NAEP context with item parameters from the standard TIMSS administration to calculate a linking equation between NAEP and TIMSS
 - Use linking equation to estimate a TIMSS-equivalent score
 - Use TIMSS-equivalent score to estimate percentage of students in states reaching TIMSS international benchmarks
- Assumptions:
 - TIMSS items have similar parameters when administered in the standard TIMSS administration as when administered instead with NAEP items in NAEP-length booklets (50 minutes) during the NAEP administration window (January through March)
 - Contingent on expanded sample and budget increase
 - International assessment bodies okay use of items
- Potential:
 - Unknown, still very early in development, study planned for 2011

Conclusion

- U.S. trying to help states fulfill international benchmarking goals without dramatically increasing cost or burden to schools and students already doing lots of testing
- Several approaches in developmental stages
- Looking for advice and would like to hear experiences of other countries