

# State Achievement Comparisons: Is the Time Right?


Rebecca Zwick  
ETS and UC Santa Barbara

National Research Council, Washington DC  
December 11, 2009



# Overview

- I. Changing views of state achievement comparisons
- II. Common Core State Standards Initiative: Opportunities and challenges for assessment
- III. Recommendations for meaningful achievement comparisons



Should academic  
achievement be compared  
across states,  
and if so, how?

A selective historical romp...



## Original NAEP Design (1960' s)

“ ... the original NAEP design by deliberate plan made it difficult if not impossible to link assessment results to state programs ... ”

-----*Messick, Beaton, & Lord, 1983, p. 12*

# A New Climate in the 1980's

- 1983 - *A Nation at Risk*:  
“If an unfriendly foreign power had attempted to impose on America the mediocre educational performance that exists today, we might well have viewed it as an act of war.”
- 1984: Reagan's State of the Union address called for comparing achievement for states and schools:  
“Without standards and competition there can be no champions, no records broken, no excellence ...”

## “Wall charts,” 1984-1991



- US DoE compared states in terms of educational resources and educational performance, including average SAT and ACT test scores.
- Test content not aligned with K-12 instruction
- Only college-going students take tests
- States differ vastly in participation rates



## Alexander-James review of NAEP (1987)

“The problem with [the Secretary of Education’s] wall chart is that its sources of performance ... information are wretchedly inadequate. We know why ... there are simply no other comparable data available.” (p. 6)



## Alexander-James review of NAEP

“The single most important change recommended by the Study Group is that [NAEP] collect representative data on achievement in each of the fifty states and the District of Columbia.” (*p. 11*)



# New NAEP Legislation

- 1988: Trial State Assessment authorized: 37 states administered grade 8 math assessment in 1990
- 1996: NAEP state assessments no longer considered "trial."

# Voluntary National Test

Born, 1997


Died, 1999






## NCLB Act of 2001 said:

- States that receive Title I funding must participate in state NAEP in reading and mathematics at grades 4 and 8 every two years.



NCLB also led to comparisons of states based on their own assessment programs, despite differences in

Standards  
Tests  
Definitions of Proficiency



Now: Common Standards Initiative  
creates both opportunities and  
challenges for state achievement  
comparisons . . .



# Opportunities

Economies of scale in test development (and Race to the Top funding) could mean improved assessments:

- Better alignment with standards
- Richer and more innovative test questions
- Better and more timely reporting
- Better diagnostics, instructionally relevant feedback (formative assessment)



# Opportunities (continued)

## Better research:

- Piloting of test items
- Validity studies
- Fairness studies
  - Is the test equally valid for all ethnic, language, socioeconomic, and gender groups?
  - Is there evidence of differential item functioning?
- Accommodations research

# Challenges

How can we assure that state achievement comparisons are:

- Fair
- Valid
- Useful for improving teaching and learning?





## Challenges: Technical issues

Can a common test be developed that allows the claim that state results are “on the same scale?”

- Can enough common items be developed in each area to assure a strong linkage across states?
- Will the inclusion of state-specific items affect performance on common items (context effects)?
- Will scores be useful for improving instruction?



## Technical issues (continued)

Will it be feasible for the states to adhere to common procedures?

- Test preparation
- Administration
- Accommodations
- Student exclusion (exemption from testing or reporting requirements)



## Challenges: Interpretation issues

- Common standards don't imply uniform opportunity to learn: curriculum and instruction still differ
- States also differ in terms of demographic characteristics, per-pupil expenditures, teacher licensing and training, class size, student retention policies, and more ...



## National Academy of Education review of Alexander-James report:

“We are concerned about the emphasis [in the report] on state-by-state comparisons of average test scores. Many factors influence the relative rankings of states ... Simple comparisons are ripe for abuse and are unlikely to inform meaningful school improvement efforts.” (Glaser et al., 1987, p. 59)

Recommendations for  
Meaningful Achievement  
Comparisons  
Across States





# 1. Recommendations on Contextual Information

- Differences across states (e.g. in resources, instruction, and educational policies) need to be documented to provide a context for comparisons.
- Ensure that data on all relevant characteristics is collected in a rigorous and comparable fashion.



## 2. Recommendations on the Test Instrument

- Decide early on scoring and reporting models.
- Develop an adequate pool of items in each area of interest.
- Ensure uniformity of format and instructions.



## 2. Recommendations on the Test Instrument (continued)

- Put unique state items in separate section after common items to minimize context effects.
  - Does item position really make a difference?  
See NAEP Reading Anomaly (1986)!
- Take advantage of available funds to conduct research that can improve the test.



### 3. Recommendations on Testing Policies

Implement rigorous and detailed protocols for:

- Test preparation and motivational activities
- Exclusion and accommodation rules for ELLs and students with disabilities
- Test administration (dates, timing, number of sessions, setting, computerization, etc.)
- Test scoring
- Test security



### 3. Recommendations on Testing Policies (continued)

- Offer high-quality training to all teachers and school administrators in the interpretation and use of test scores (long overdue!)



## Conclusion

To promote meaningful state comparisons:

- Provide as much context as possible
- Make the test instrument and testing policies as similar as possible
- Invest in research and professional development



## References

- Alexander , L. , & James, H. T . (1987). *The Nation's Report Card*. Cambridge, MA: National Academy of Education.
- Glaser , R. (1987). A review of the report by a committee of the National Academy of Education, included in above.
- Messick, S. , Beaton, A. , & Lord, F . (1983). *A New Design for a New Era*. Princeton, NJ: NAEP and ETS.