

SAYING IT WITH DATA: NAEP AND NEW JERSEY'S STATE ASSESSMENTS

National Assessment of Educational Progress (NAEP)

NJDOE Summer Conference

*One Vision, Many Voices: Academic
Achievement for ALL Students*

Atlantic City, NJ
August 8, 2006

*Barbara Smey-Richman, EdD.
New Jersey NAEP Coordinator*

Hi! I'm from the "NJDOE" and I'm here to help!!



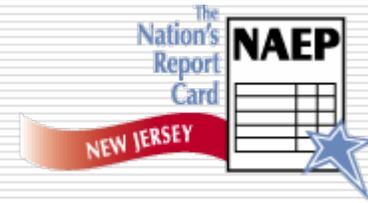
The Nation's Report Card
NAEP

NEW JERSEY

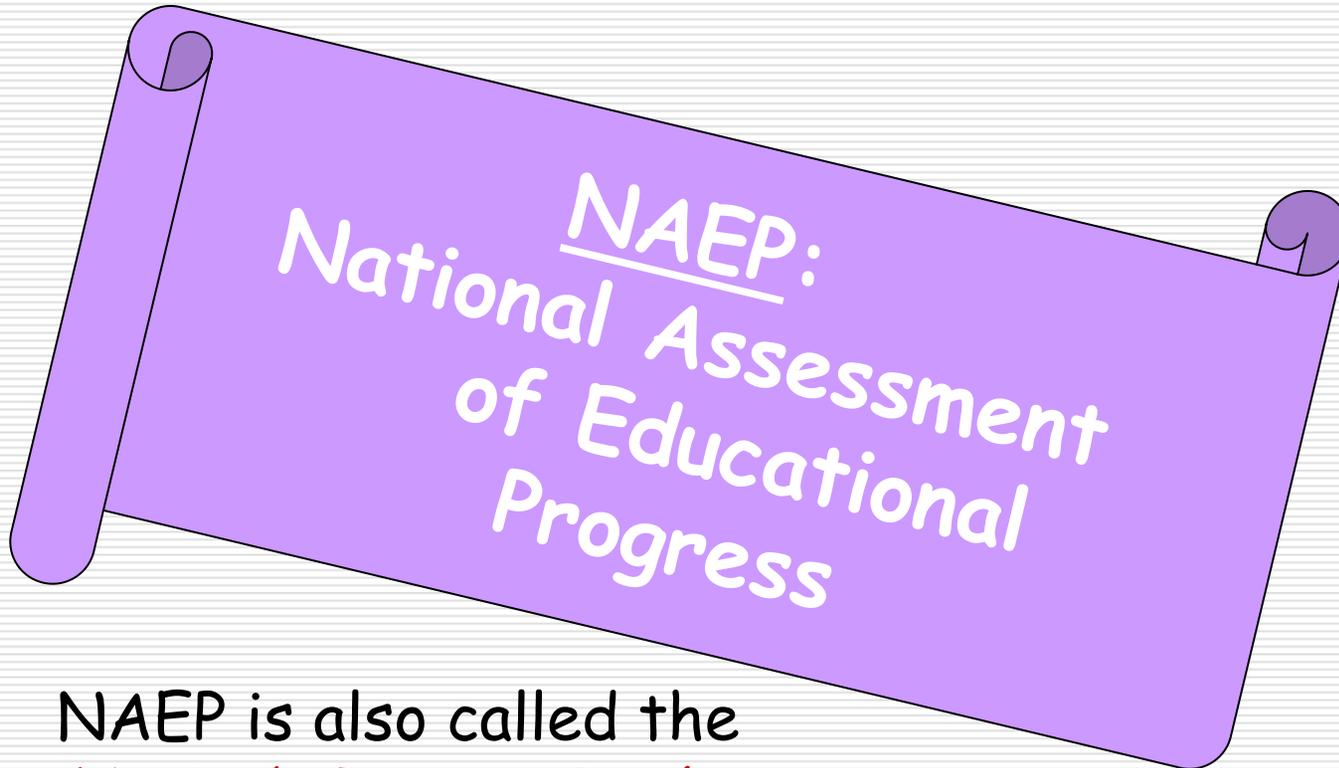


Barbara Smey-Richman, Ed.D.

*New Jersey
NAEP Coordinator*



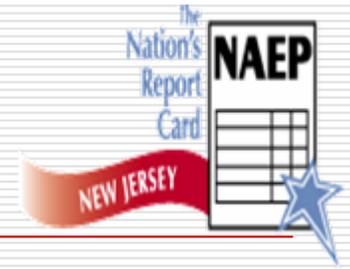
NAEP Means What?



NAEP is also called the
Nation's Report Card

NAEP
Rhymes
with
"Grape"

The Power of the Metaphor

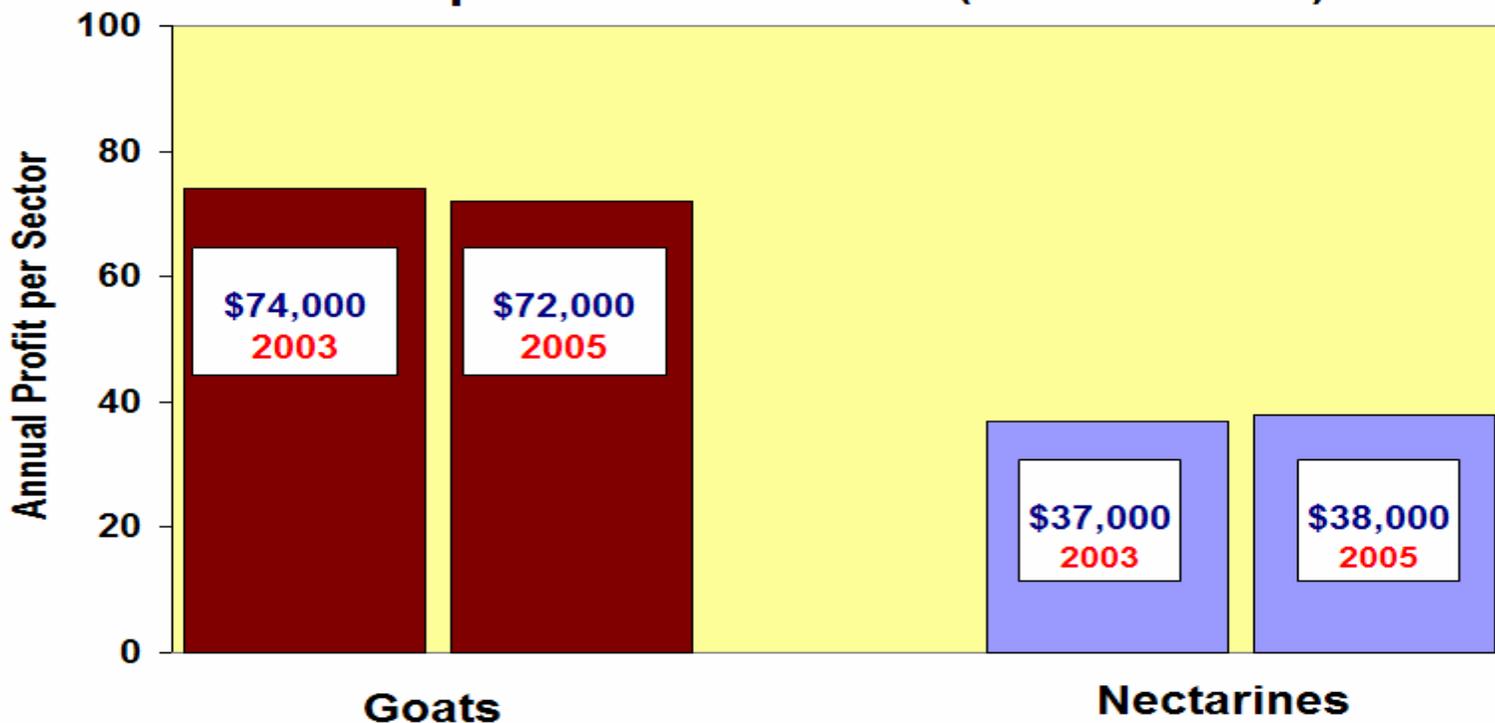


Comparing NAEP with
the New Jersey
tests is like
comparing?



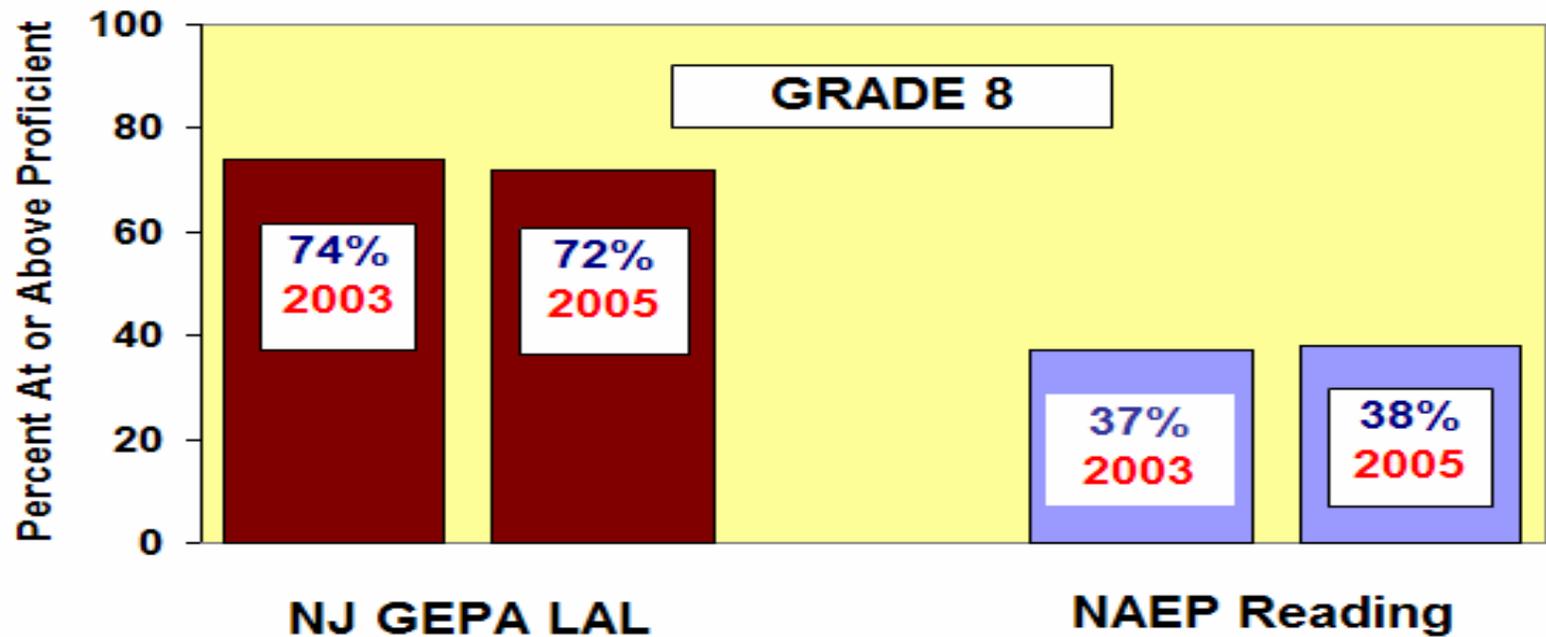
Metaphor: The Garden State

**Goats and Nectarines: Garden State Farm --
Comparison of Profits (2003 and 2005)**



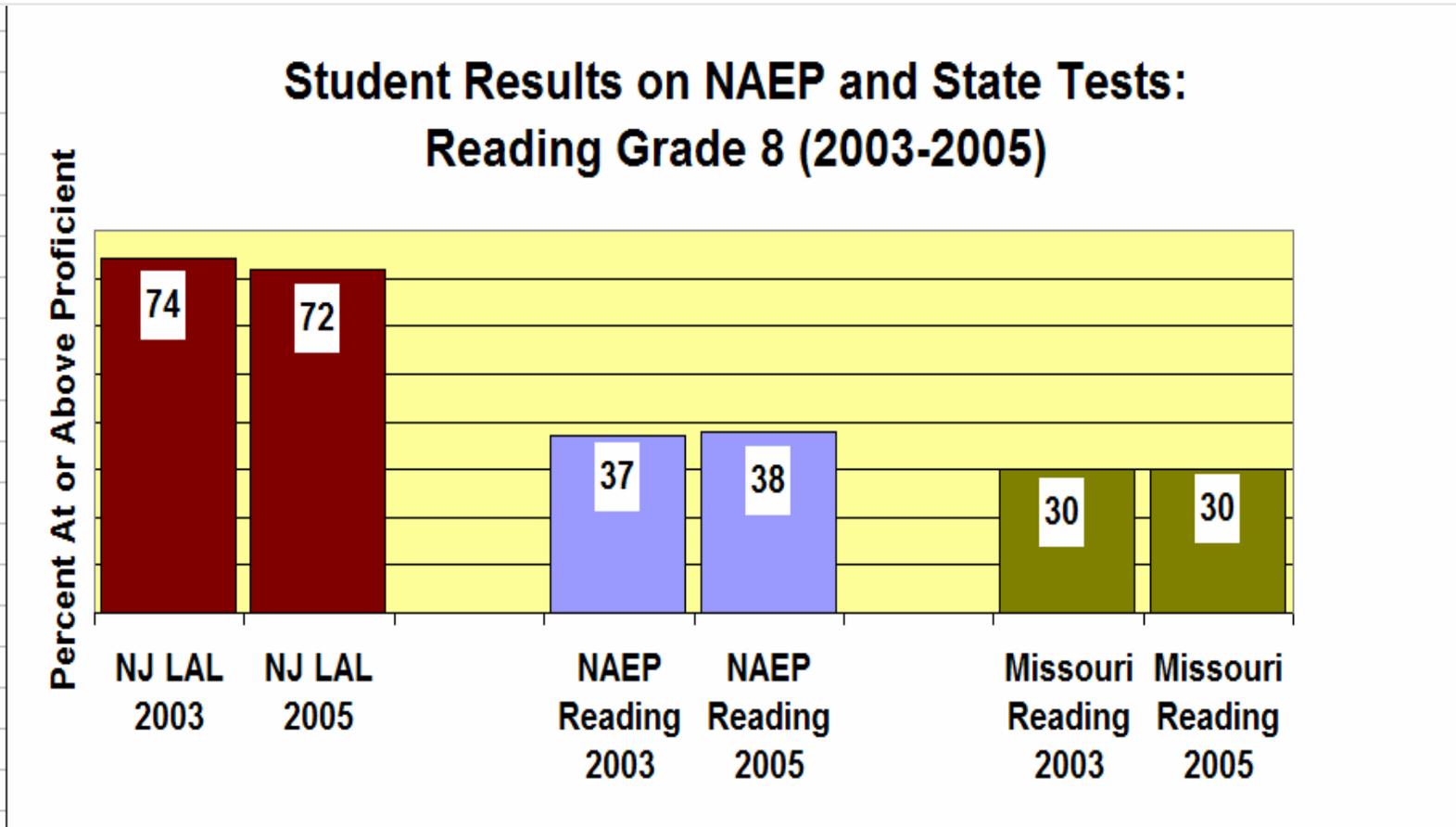
Surprise: Same Data/Different Story

NAEP and NJ GEPA Language Arts Literacy (LAL): Different Tests Yield Different Results (2003-2005)

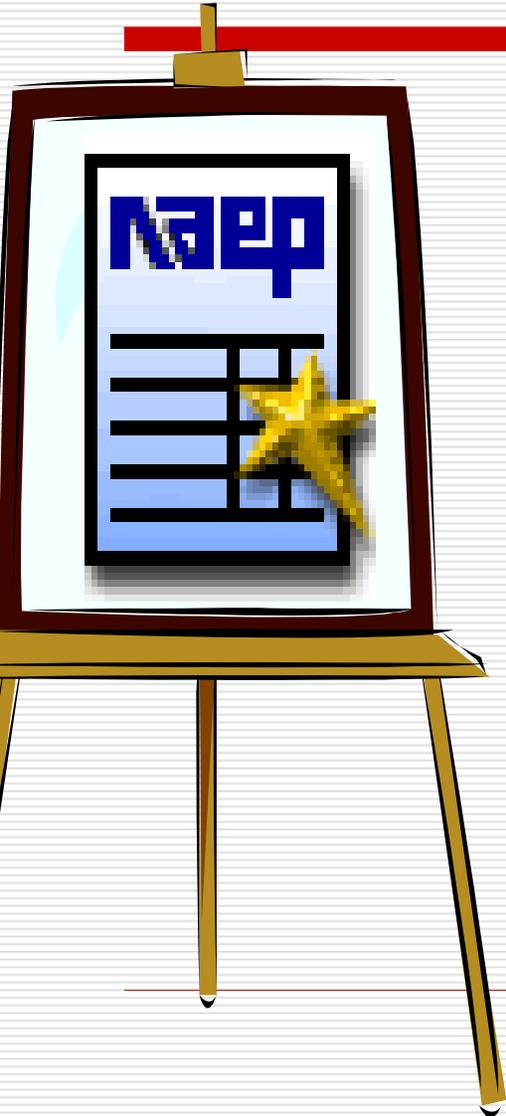


NAEP and State Tests

**Student Results on NAEP and State Tests:
 Reading Grade 8 (2003-2005)**

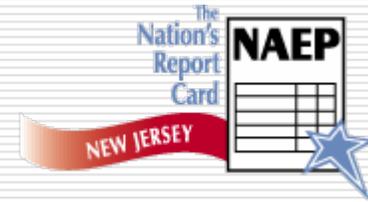


Session Themes:

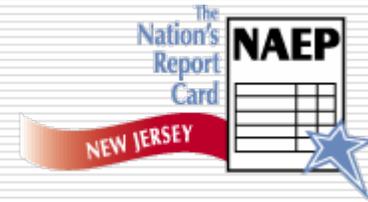


- NAEP Characteristics
 - NAEP and NJ Assessments
 - NJ Student Results
 - Achievement Gaps
 - Simpson's Paradox
 - Final Observations
 - Q & A
-

What is NAEP?



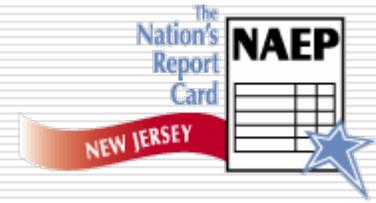
NAEP was authorized by Congress in 1969 to create a *reliable way* of determining areas of strengths and weaknesses in the American school system



What is NAEP?

NAEP was created to answer the question:

Are students in American schools learning what they should be learning?



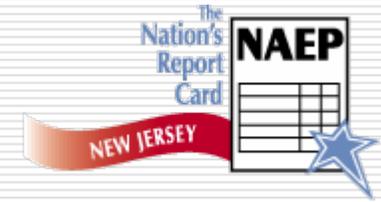
What is NAEP?

- NAEP's mission is to produce national and state-level results for student populations
 - NAEP does **not** provide individual student, school or district results
-

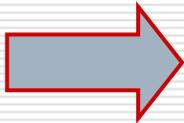
What's Tested?

- NAEP tests a variety of subjects
 - The primary assessments are:
 - Reading
 - Math
 - Writing
 - Science
 - NAEP also funds "special studies"
-

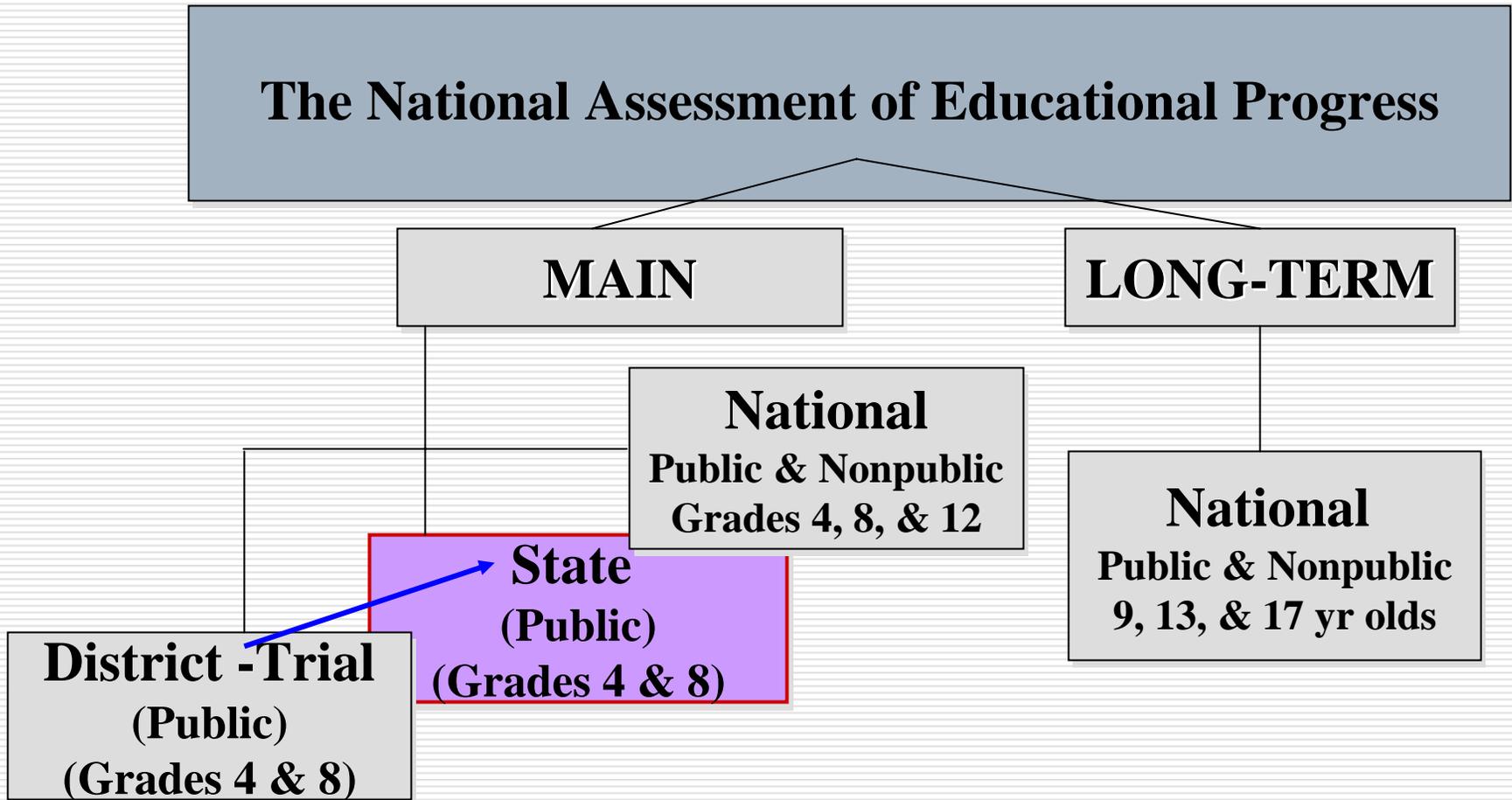
NAEP's Assessment Schedule

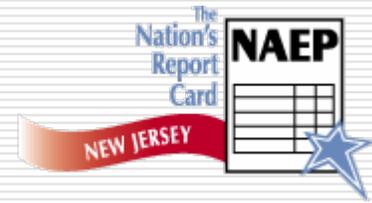


Year	National	State
2003	Reading, Mathematics	Reading, Mathematics
2004	Foreign Language, Long-term Trend <i>(Cancelled)</i>	
2005	Reading, Mathematics, Science	Reading, Mathematics, Science
2006	U.S. History, Civics, Economics	
2007	Reading, Mathematics, Writing	Reading, Mathematics, Writing
2008	Arts, Long-term Trend	
2009	Reading, Mathematics, Science	Reading, Mathematics, Science
2010	World History, Geography	
2011	Reading, Mathematics, Writing	Reading, Mathematics, Writing
2012	Civics, Foreign Language, Long-term Trend	



NAEP Has Many Components



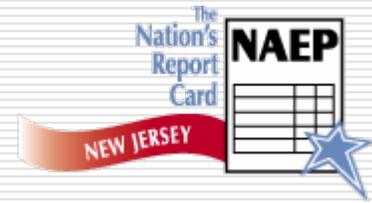


What is State NAEP?

State NAEP (1990) is an estimate of student performance for the

- 50 states
- Washington, DC
- Puerto Rico
- Department of Defense Education Activities

➤ State NAEP is a biennial assessment



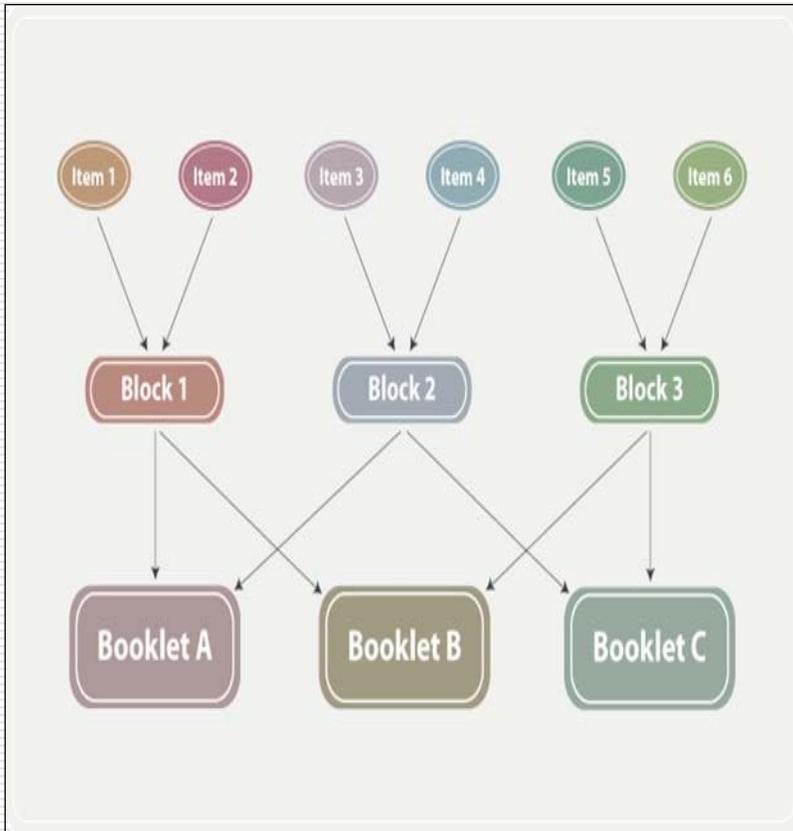
What is State NAEP?

- With NCLB, Title I funding became contingent upon school participation in NAEP at grades 4 and 8
 - Prior to 2003, school participation was voluntary
 - The NAEP data are not reported if the school-participation rate is low
-

Who's Tested?

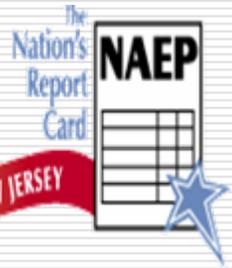
- NAEP uses a complicated **Multi-Stage Stratified Random Sampling** method
 - Sample sizes
 - Approximately 3,000 students are selected for each subject and grade level (same for all states)
 - NAEP selects about 3% of NJ's 100,000 students for each grade-level cohort
-

How are Students Assessed?



"Matrix Sampling"

- Each student answers only a small portion of the NAEP items
- Students in the same classroom receive different booklets



What Content is Tested?

- NAEP does not have curriculum content standards
 - NAEP's **assessment frameworks** are analogous to the New Jersey **test specifications**
 - The NAEP assessment frameworks include:
 - theoretical basis of the test
 - directions for item development
-

NAEP's Theoretical Basis *

➤ Context for Reading

- Literary Experience
- Information - retrieving information
- Perform a Task - following instructions

➤ Aspects of Reading

- General Understanding
- Interpretations
- Reader/Text Connection
- Context and Structure

* See the NAEP web site for further details.



What's New With NAEP?

- NAEP 2009 -- a totally new NAEP Reading Assessment
 - The current NAEP trend line will be broken in 2009
-

Check It Out!!



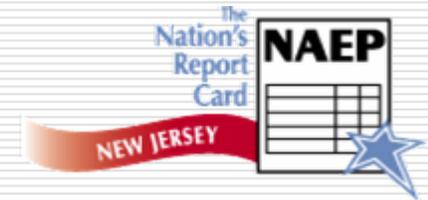
NAEP Releases Test Items!!

Obtain a copy of the
Questions Tool
Quick Reference at
this session

Add to Your Favorites List!

On the NAEP web site

<http://nationsreportcard.gov>



- Find sample questions (*NAEP Question Tool*)
- Review scoring guides
- Compare state groups (*NAEP Data Explorer*)
- View results by scores and achievement levels

[NJ NAEP Web pages:](http://www.nj.gov/njded/assessment/naep)

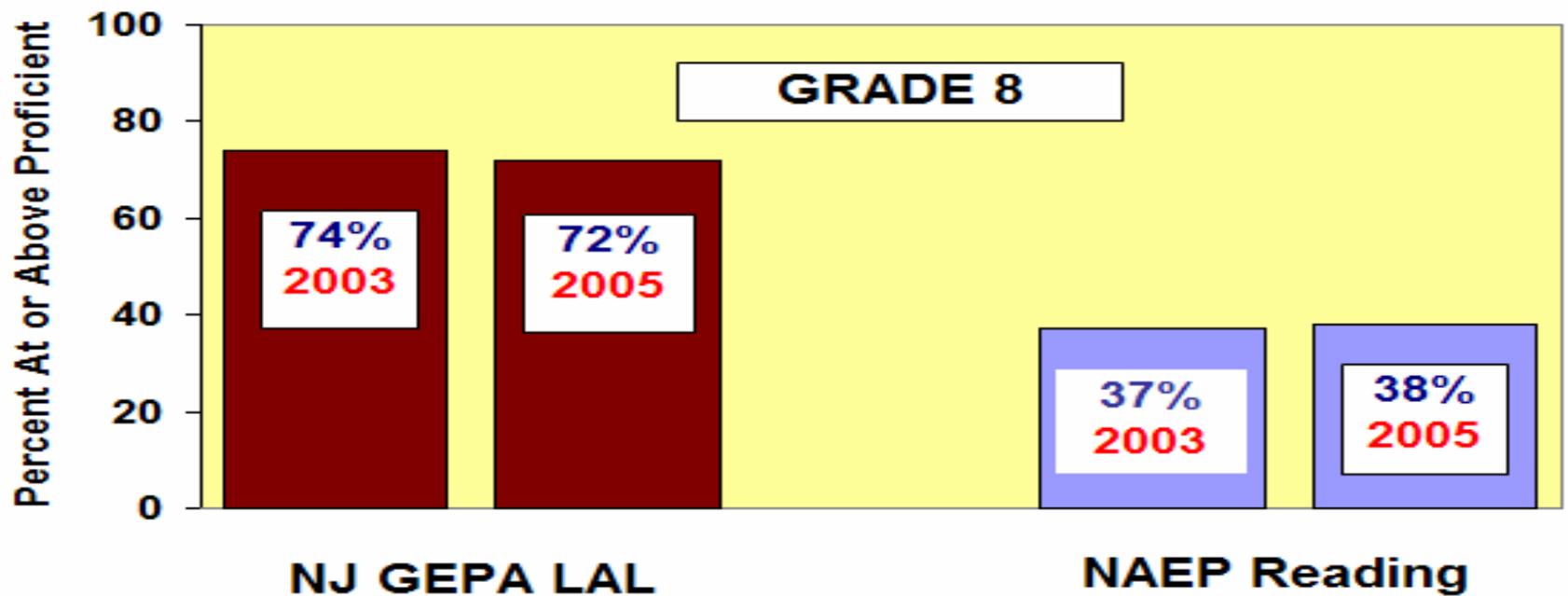
<http://www.nj.gov/njded/assessment/naep>



NAEP
and
New Jersey
Assessments

Getting Back to "Why" So Different?

NAEP and NJ GEPA Language Arts Literacy (LAL): Different Tests Yield Different Results (2003-2005)



Let's Take a Closer Look!



What are the similarities and differences?

NAEP's Definition of Reading

Reading is defined as an active process that involves

- Understanding written text
- Developing and interpreting meaning, and
- Using meaning as appropriate to the type of text, purpose, and situation

(NAEP 2009 Framework)

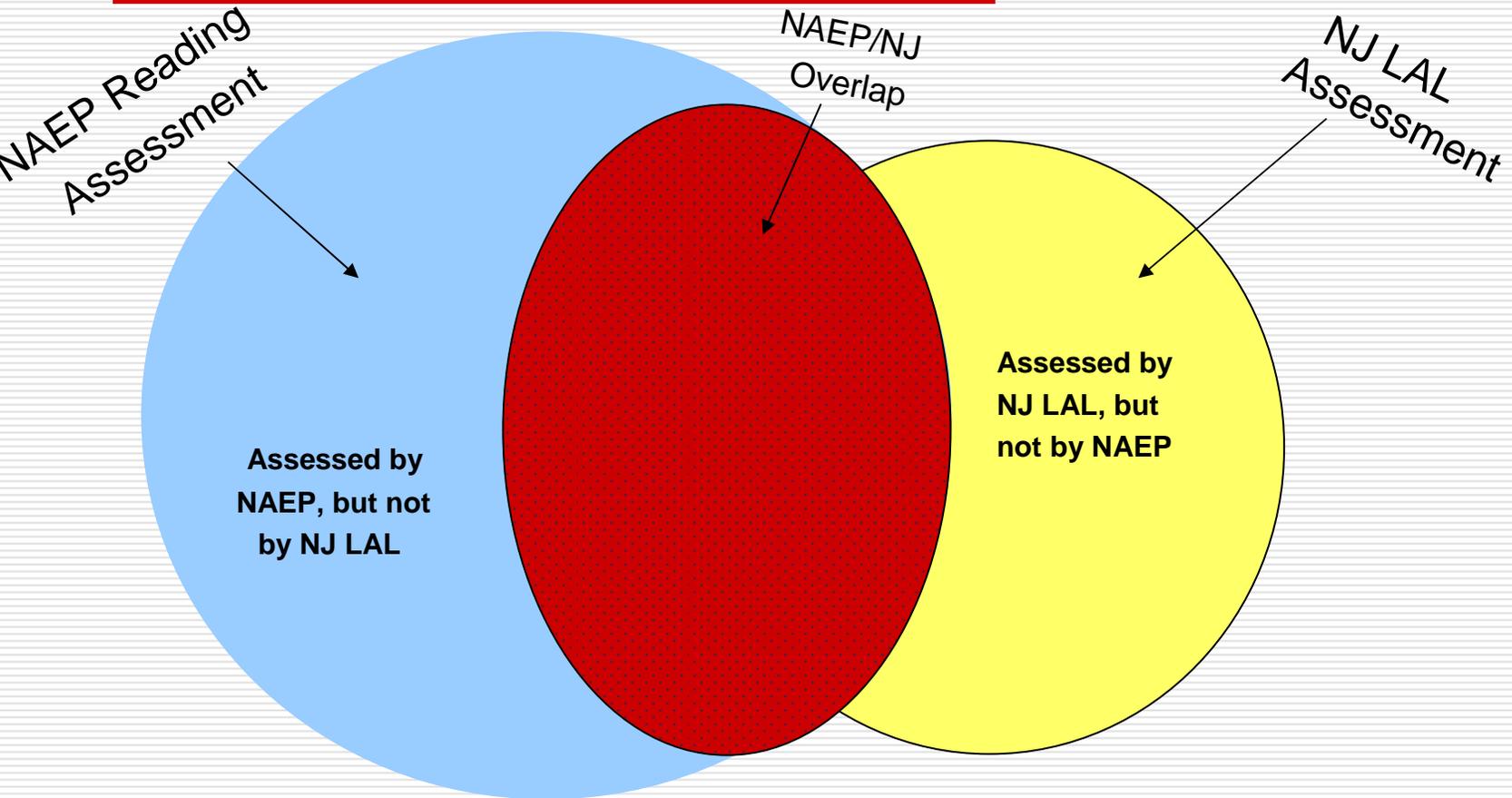
What's Being Tested?



Similarities?

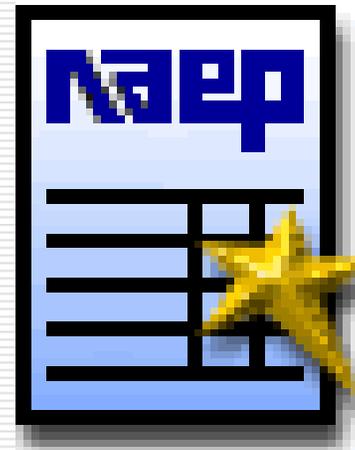
Don't NAEP and the NJ tests measure a common core of reading ability?

NAEP/NJ Overlap: Conceptual Model



This diagram is only an approximation of how the NAEP reading and New Jersey LAL tests relate to each other. The diagram is not drawn to "scale," and the extent to which the tests overlap is not based on an alignment study.

An Obvious Difference



Language Arts Literacy

- Combined Tests
- Combined Scores only

Reading and Writing

- Separate R/W Tests
 - Separate R/W Scores
-

Different Schools of Thought??

- NAEP's assessment frameworks are developed as a result of a *national-based process* and reflect a *national view*
 - New Jersey's content standards and test specifications for NJASK 4 and GEPA reflect the *New Jersey perspective*
-

An Important Question - -



Are the NJ state scores higher because NJ's instruction is focused on the LAL standards?

NAEP/NJ Differences

➤ Test length

- *Content questions take 50 minutes per student; total test would take 7 hours*

➤ Content and skills assessed

- *NAEP tests a wider domain of knowledge since the test is longer*

➤ Average scale score points for reading tests

- *NAEP has a 500 pts. scale; NJ, 300 pts. scale*

➤ Student motivation

- *NAEP has no consequences for the students*
 - *Are students focused and trying their best?*
-

NAEP/NJ Differences

➤ Standard-setting methods

- *NAEP's process for determining the achievement levels may be different from that used by the individual states*
- *Cut points vary according to the method used*

➤ Item difficulty and format

- *NAEP items may be more difficult*
 - *For NAEP, about half the time is spent on short/extended constructive response items*
 - *Also has multiple choice items*
-

NAEP's Achievement Levels

Level Descriptors

- **Basic:** Partial mastery of prerequisite knowledge/skills that are fundamental for proficient work (*At grade level*)
- **Proficient:** Solid academic performance -- demonstrated competency over challenging subject matter
- **Advanced:** Superior performance

+++++

[**Below Basic:** Incomplete knowledge/skills necessary for proficient work]

Play It Again Sam



1. NAEP and the NJ tests are inherently different tests
 2. Test develops caution against making comparisons
 3. Controversial Question: Can NAEP be used to verify overall trends found in state results?
-

What are the Implications?

Because of the inherent differences:

- Proficient performance is defined uniquely for each large scale assessment
- The definitions of “proficient” set by states and by NAEP have no observable agreement*

* Linn, Robert, *Large-Scale Assessment Conference*, San Antonio, TX, June 2005

Over Interpreting the Data



*They shouldn't say
that should they?*

By twelfth grade, Black students are typically four years behind white and Asian students. Black students are finishing high school with a **junior high school education***

*Thernstrom, A. & Thernstrom, S. (2004). *No Excuses: Closing the Racial Achievement Gap*

What's an Achievement Gap?

NAEP

Achievement Gap = A statistically significant difference in the scores between two groups

If not statistically significant, it is not a gap.

Johnny Cochran:

If it doesn't fit, you must acquit.

What's an Achievement Gap?

Same difference in scale scores, but different decisions

	Group 1 Scale Score	SE	Group 2 Scale Score	SE	Point Difference Between Groups	t- test
Test 1 GAP	210	2.5	200	3.0	10*	t = 2.56 p < .05
Test 2 Not a GAP	210	3.5	200	4.5	10	t = 1.75 p > .05

* Statistically significant



Pictures at An Exposition

Overall Results

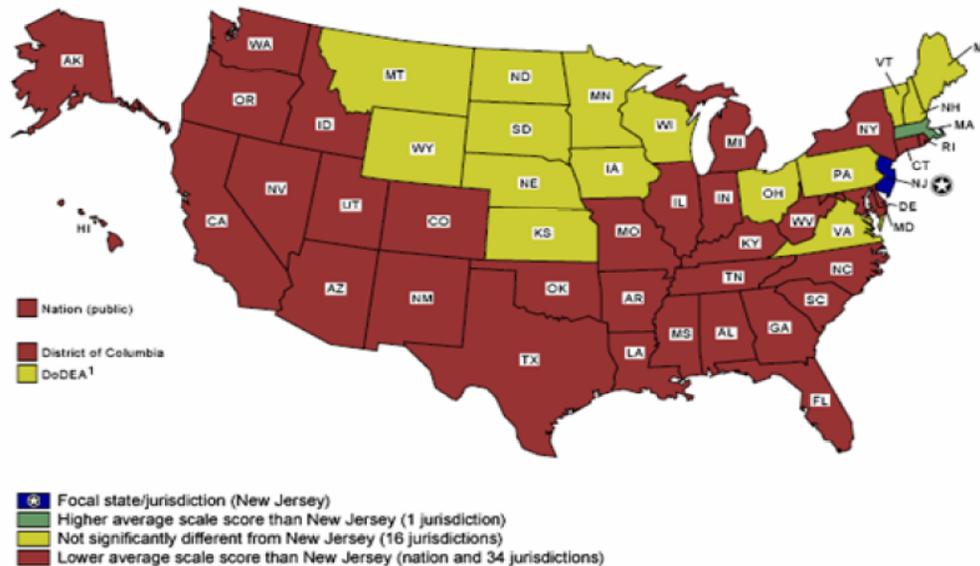


We Have
Reason
To Cheer!!

How Do NJ 8th Graders Compare?

Reading Average Scale Scores

New Jersey's average reading scale score compared with scores for the nation and other participating jurisdictions, grade 8 public schools: 2005



¹ Department of Defense Education Activity schools (domestic and overseas).
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 Reading Assessment.

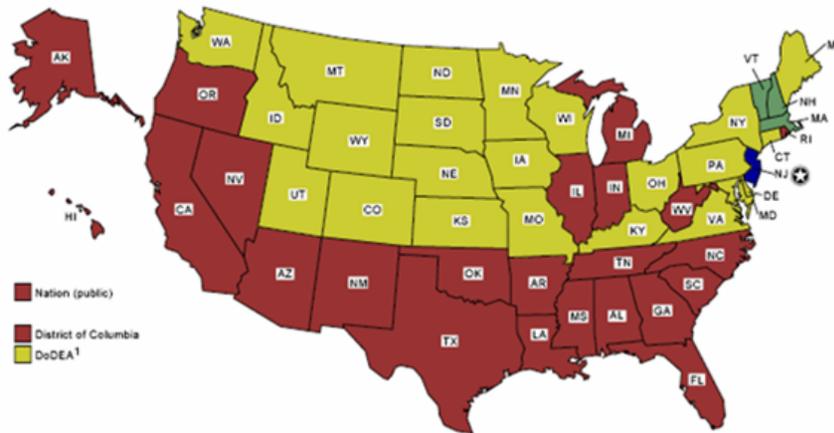
NJ scale scores were:

- ✓ Lower than 1
- ✓ Equal to 16
- ✓ Higher than 34

How Do NJ 4th Graders Compare?

Reading Average Scale Scores

New Jersey's average reading scale score compared with scores for the nation and other participating jurisdictions, grade 4 public schools: 2005



¹ Department of Defense Education Activity schools (domestic and overseas).

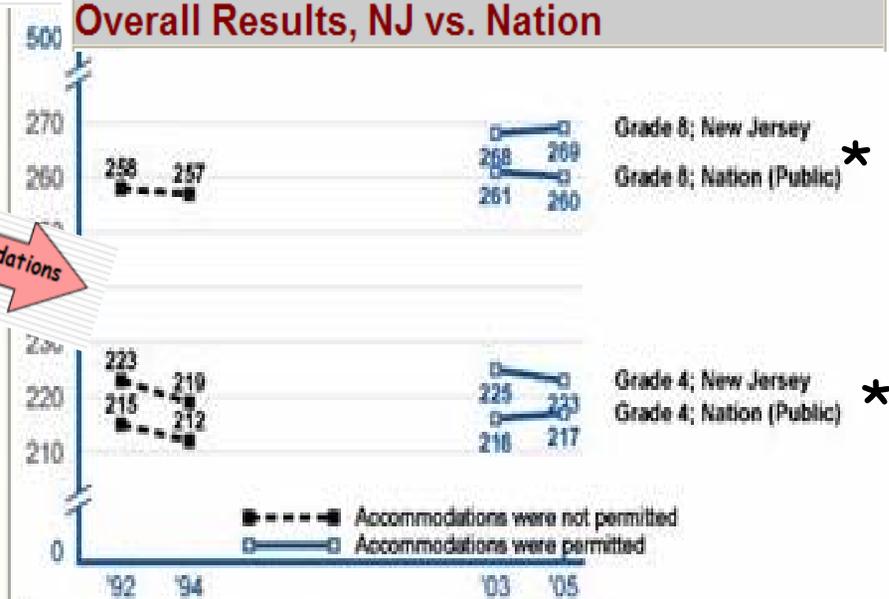
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 Reading Assessment.

NJ scale scores
were:

- ✓ Lower than 3
- ✓ Equal to 24
- ✓ Higher than 24

How Do NJ 4th and 8th Graders Compare to the Nation?

NAEP New Jersey Reading, Grade 4 and 8
Overall Results, NJ vs. Nation



At grades 4 and 8, NJ results were higher than the national public schools in 2003 and 2005

¹ Accommodations were not permitted for this assessment.

Note: The NAEP reading scale ranges from 0 to 500. Observed differences are not necessarily statistically significant.

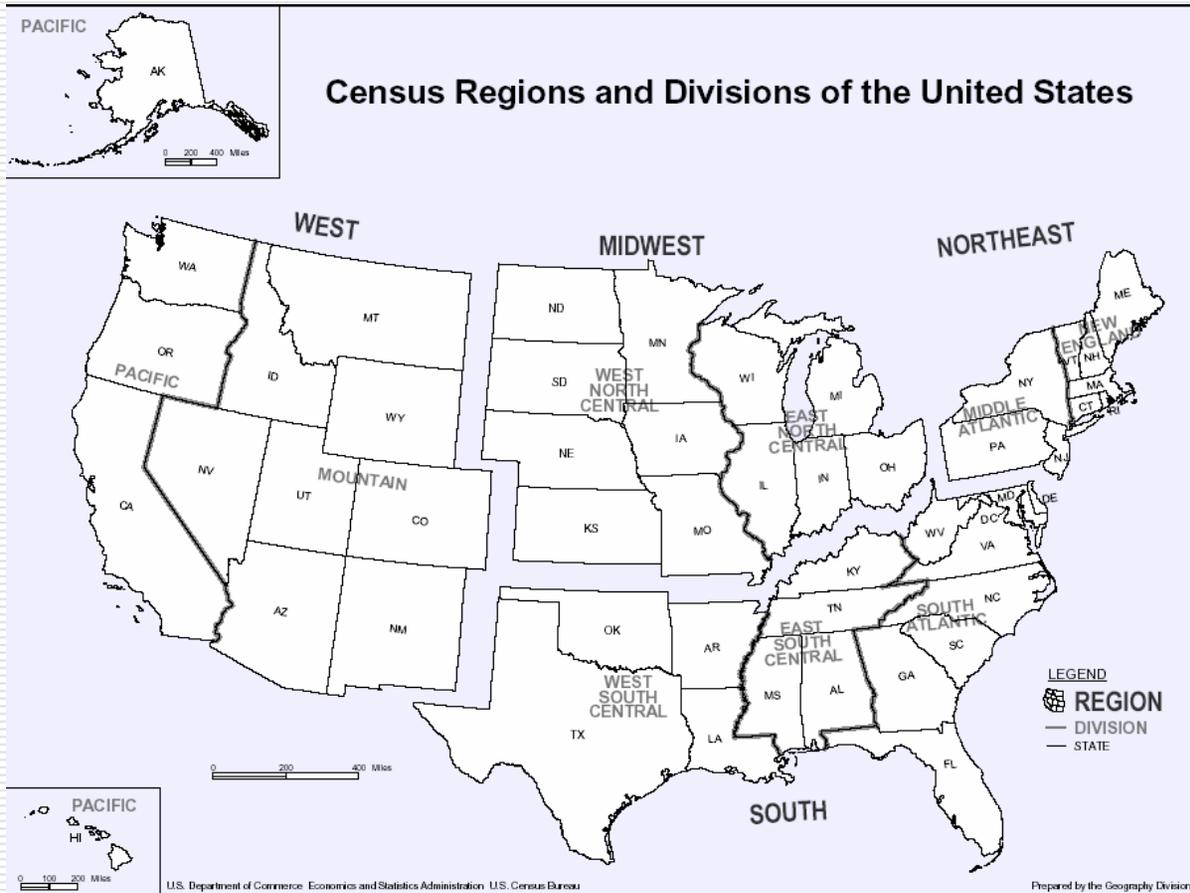
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 Reading Assessment.

* Statistically significant

Comparing Jurisdictions With Similar Demographics

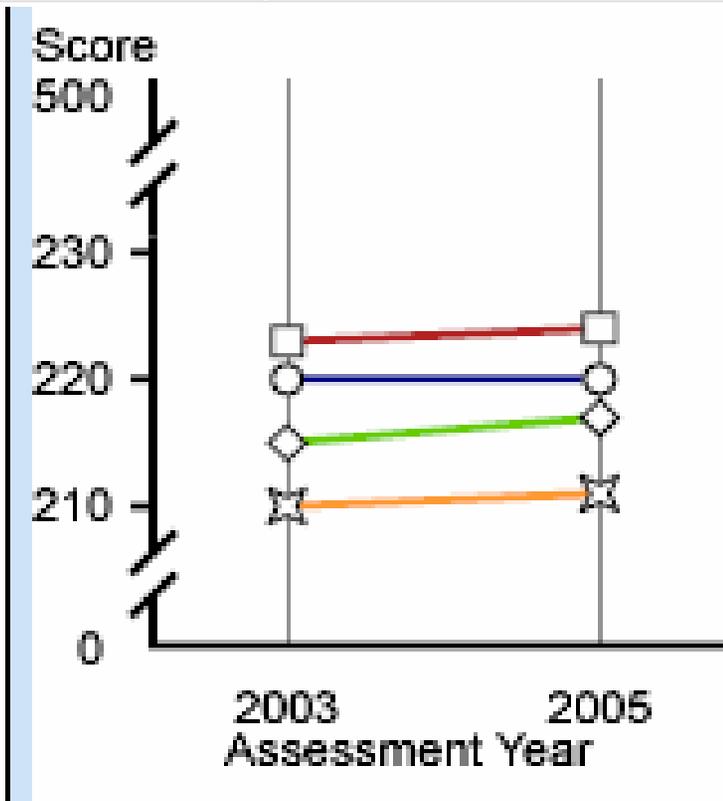


NJ Is Like States in Northeast?



How Do the Nation's Census Regions Compare? Grade 4

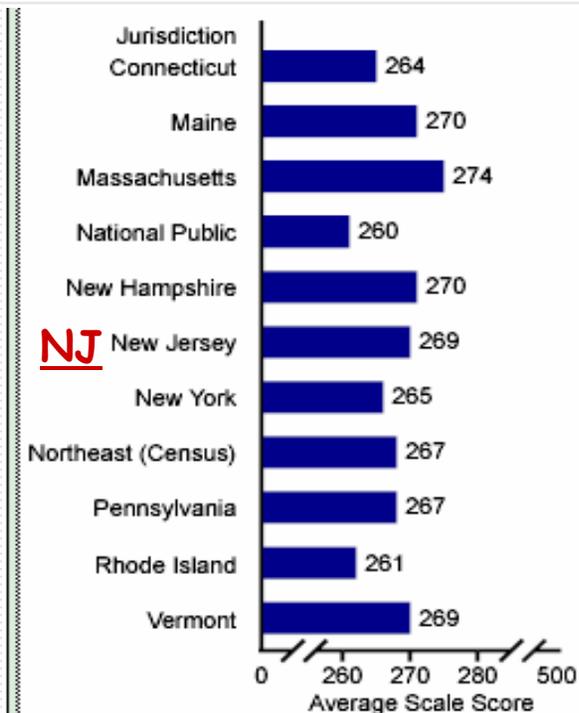
NE Compared to Other Regions



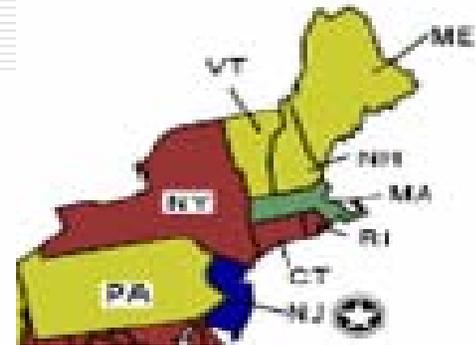
NAEP 2003 & 2005

	MW	NE	S	W
Midwest (MW)		<	>	>
Northeast (NE)	>		>	>
South (S)	<	<		>
West (W)	<	<	<	

How Do NJ 8th Graders Compare to State in the Northeast? (2005)



NOTE: The NAEP Reading scale ranges from 0 to 500. Observed difference
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Educational Progress (NAEP), 2005 Reading Assessment.

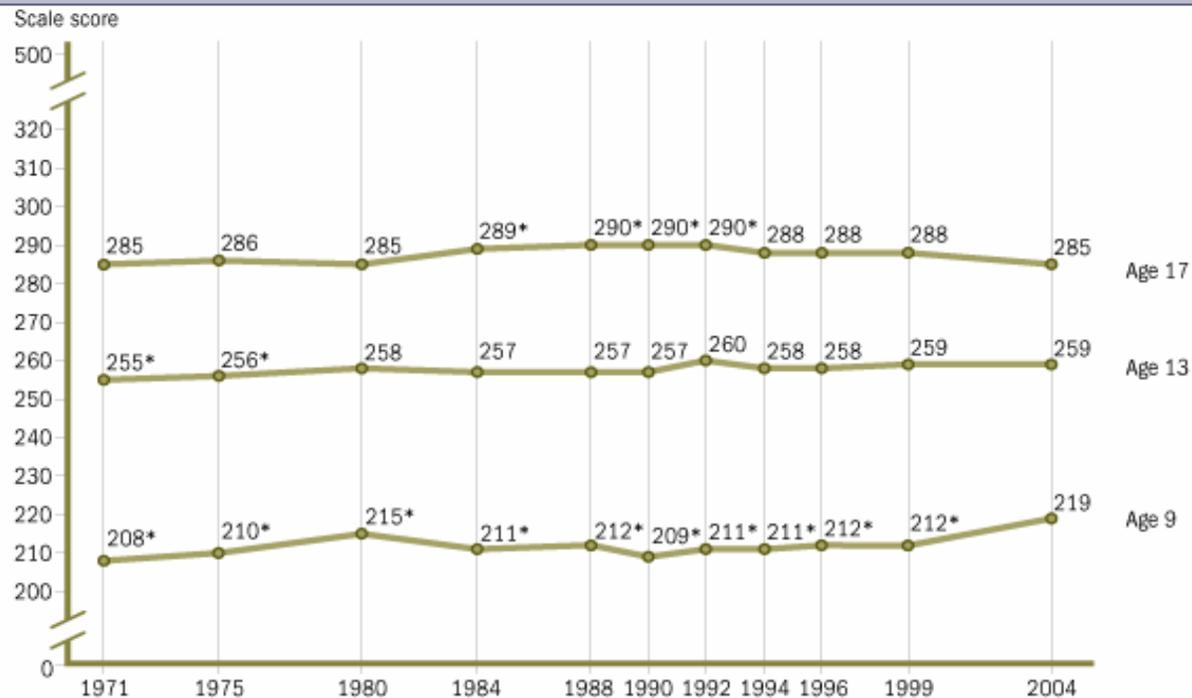


- Blue star icon: Focal state/jurisdiction (New Jersey)
- Green box: Higher average scale score than New Jersey (1 jurisdiction)
- Yellow box: Not significantly different from New Jersey (16 jurisdictions)
- Red box: Lower average scale score than New Jersey (nation and 34 jurisdictions)

Note: New Jersey's average scale (269) score was not different from the average scale score for the Northeast Census Region (267).

Are the Nation's Reading Scores Improving? (Long-term Trend) (2004)

Trends in average reading scale scores for students ages 9, 13, and 17: 1971–2004 



View data with standard errors for [age 9](#), [age 13](#), and [age 17](#).

Key Findings:

9-yr olds: The 2004 score is higher than previous years

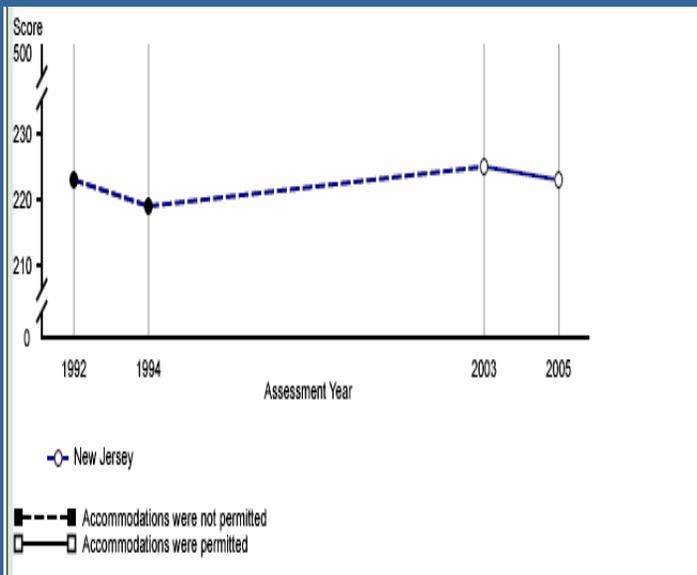
13-yr olds: The scale score in 2004 is **higher than** in 1971 and in 1975, no different for subsequent years.

* Significantly different from 2004.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), selected years, 1971–2004 Long-Term Trend Reading Assessments.

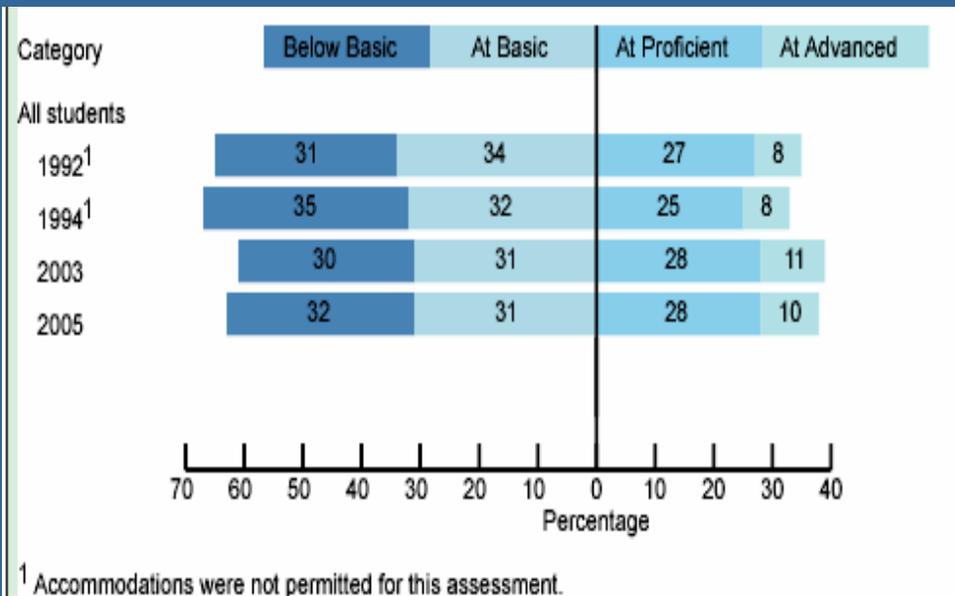
Did the NJ NAEP Reading Scores Improve from 2003 to 2005?

Grade 4



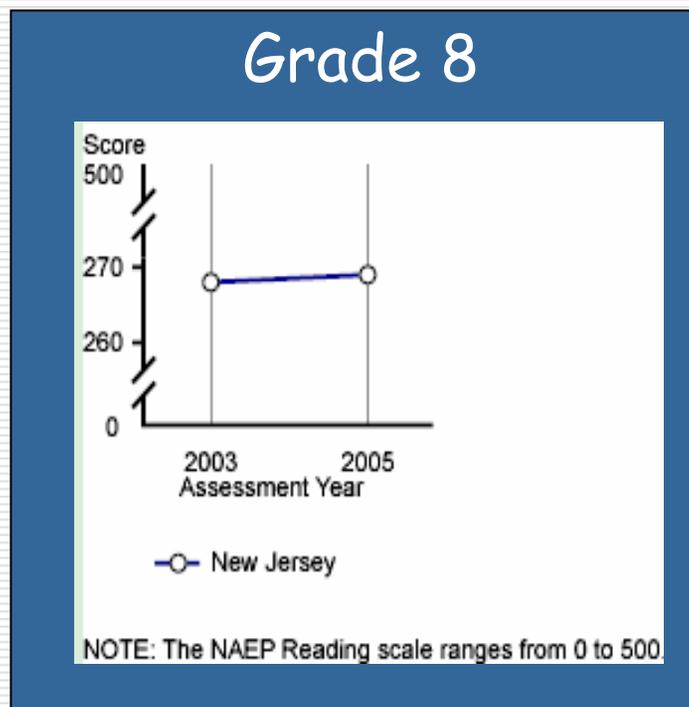
No significant change in scales scores from 2003 to 2005.

Grade 4

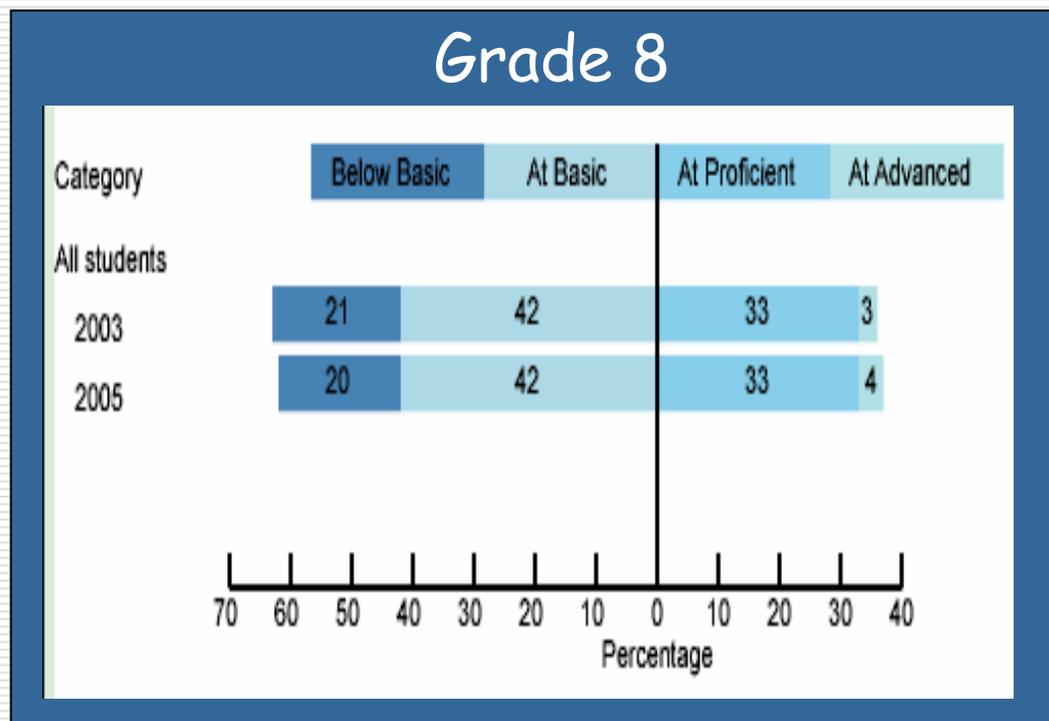


No significant change in the percentage At or Above Proficient from 2003 to 2005.

Did the NJ NAEP Reading Scores for Improve from 2003 to 2005?

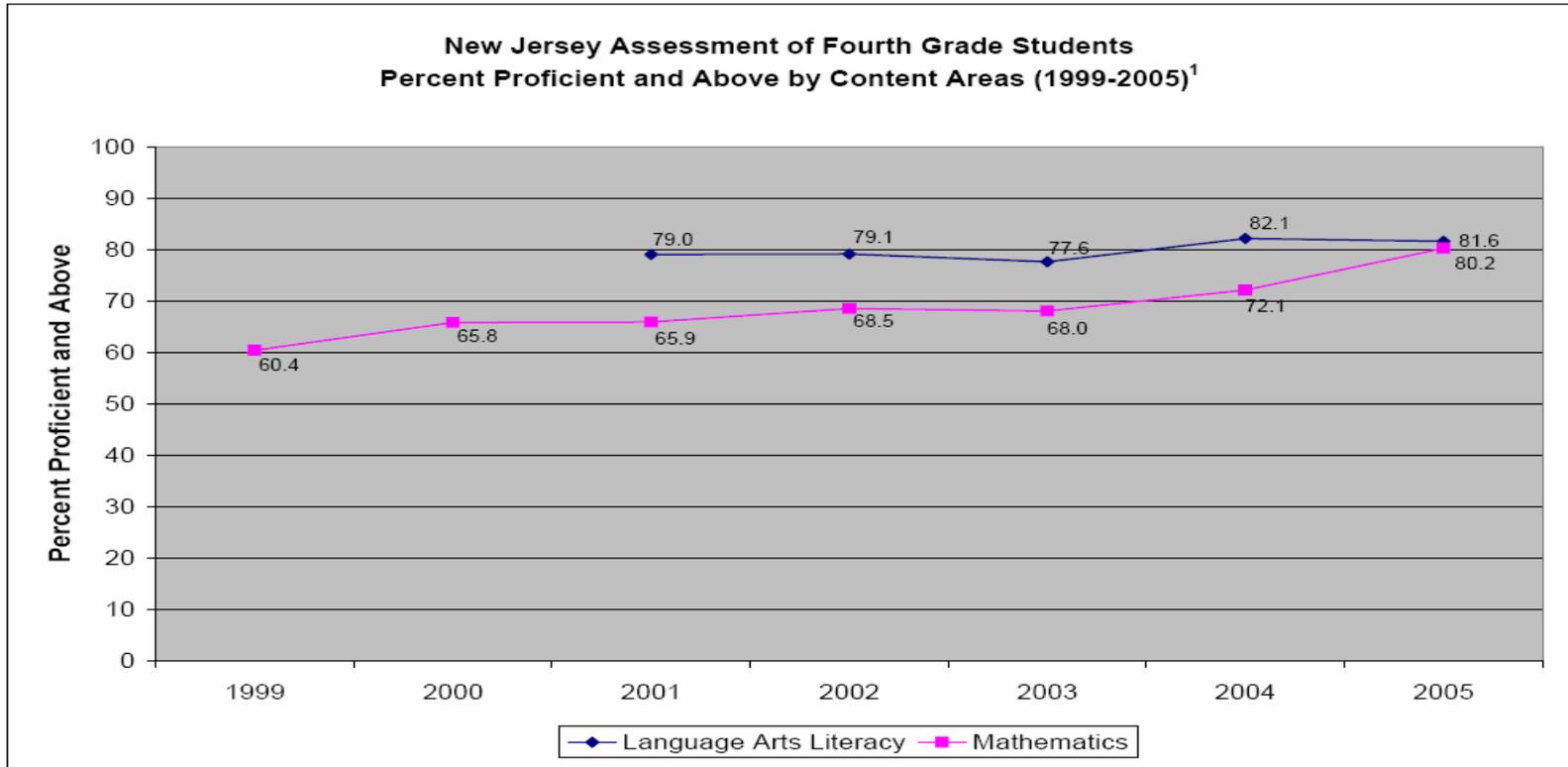


No significant change in scales scores from 2003 to 2005.



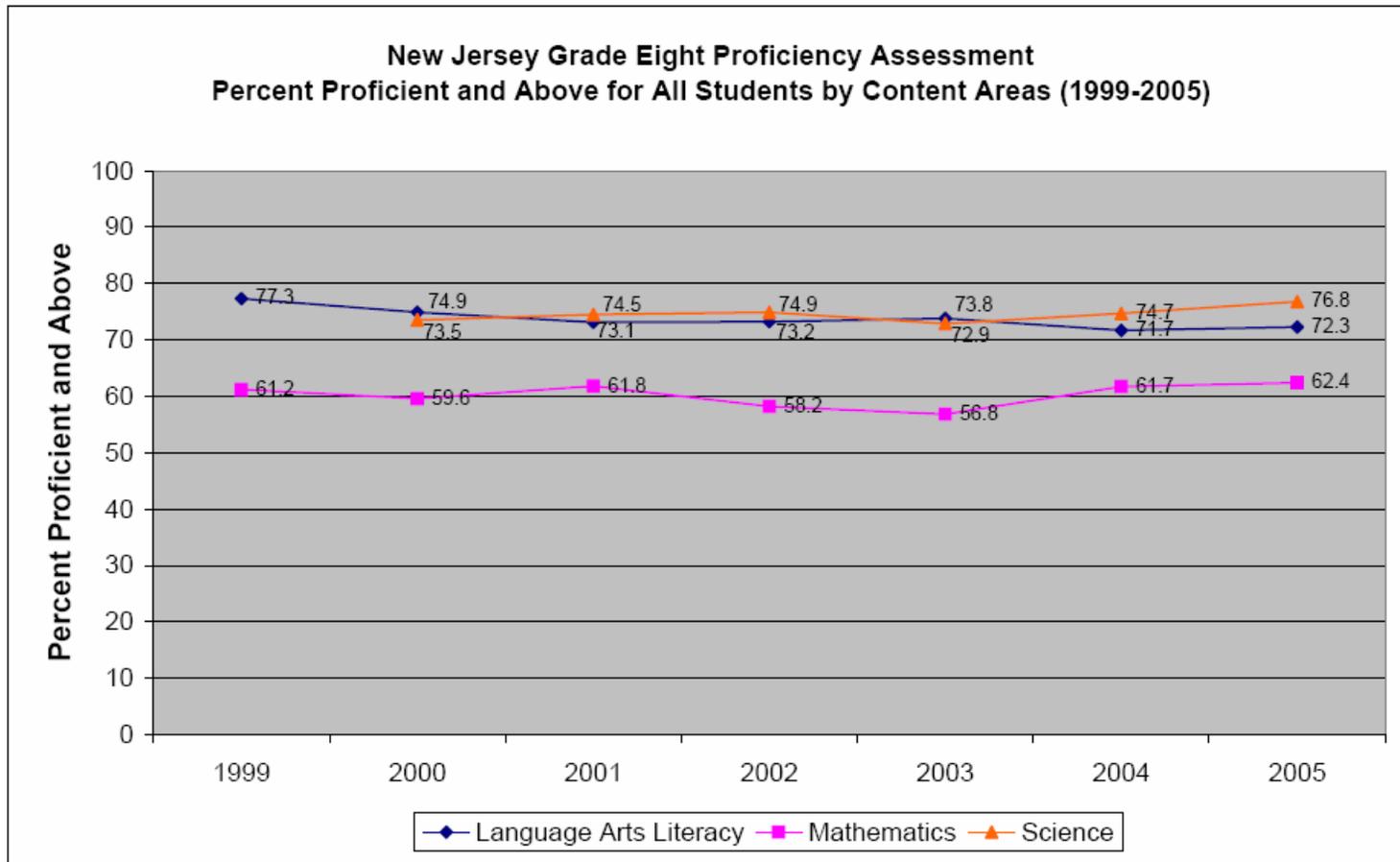
No significant change in the percentage At or Above Proficient from 2003 to 2005.

What's the Trends for NJASK 4?

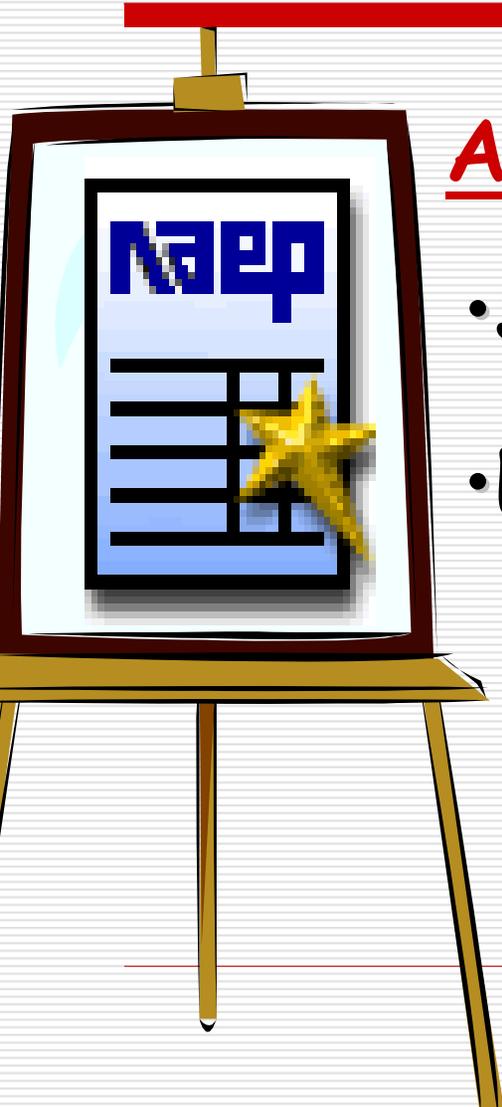


¹Elementary School Proficiency Assessment (ESPA), 1999-2002; New Jersey Assessment of Skills and Knowledge (NJ ASK4), 2003-2005. For Language Arts Literacy, 2001 was the standard setting year for the ESPA and NJ ASK4. For Mathematics, 1999 was the standard setting year for the ESPA and NJ ASK4.

What's the Trends for GEPA?



Pictures at an Exposition



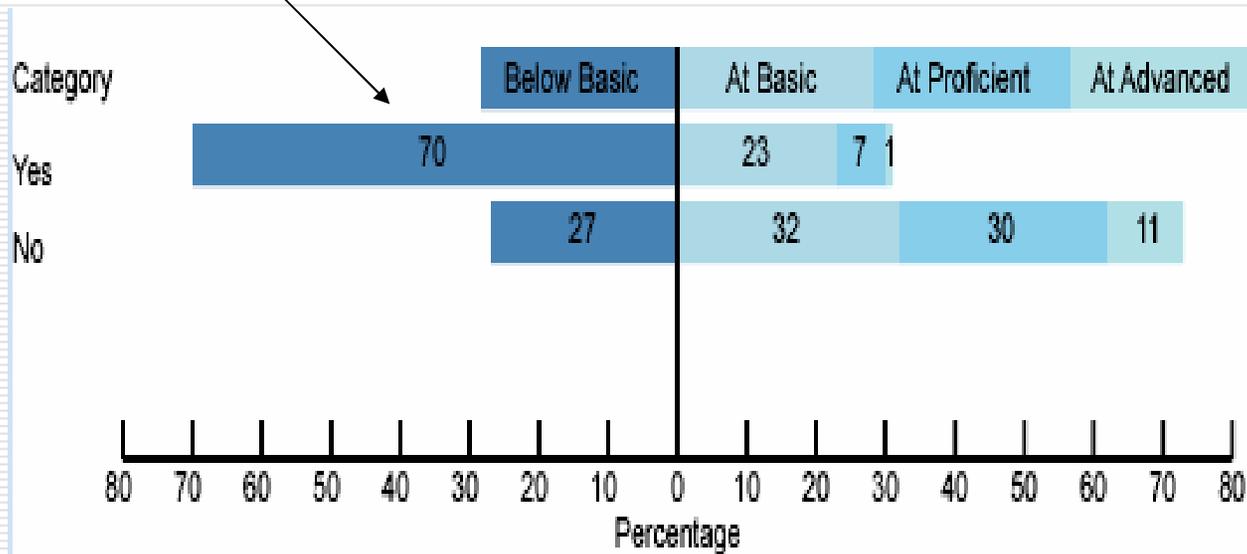
Achievement for ALL Students

- Students with Disabilities (SD)
 - English Language Learner (ELL)
-

NAEP SD/Regular Education for NJ, Grade 4, (2005)

Students with Disabilities (SD)

Below Basic

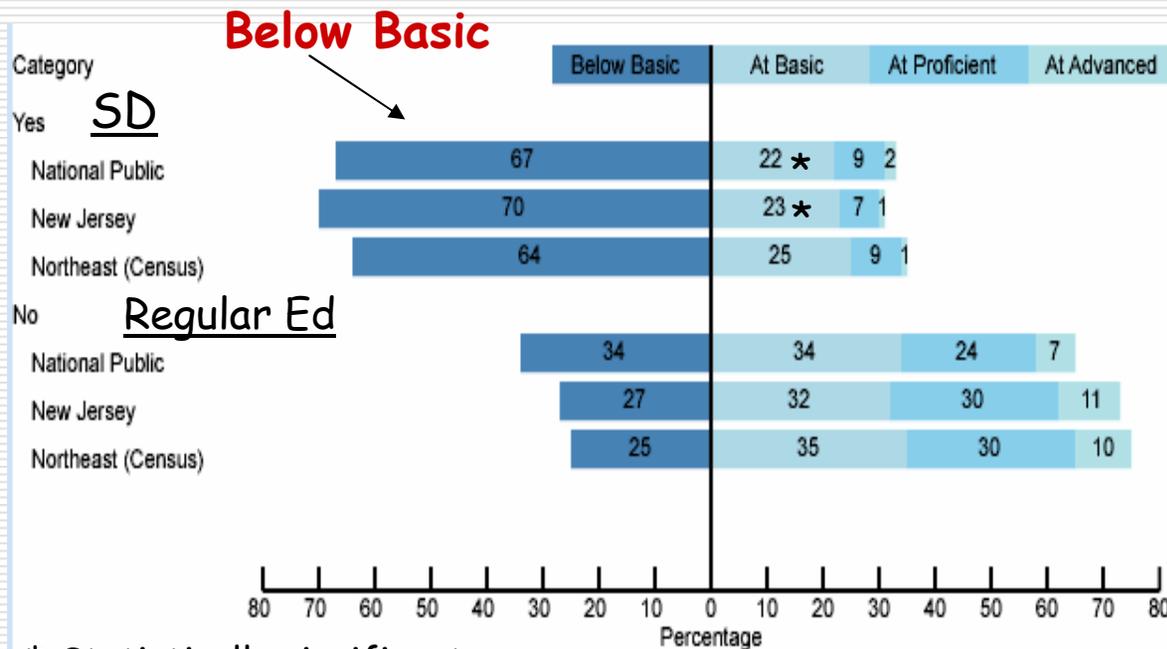


	REGULAR
SD Below Basic	<
SD Basic	<
SD Proficient	<
SD Advanced	<

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 Reading Assessment.

NAEP SD/Regular Education for NJ & Other Jurisdictions, Grade 4 (2005)

New Jersey, Nation, and Northeast



At Basic

Northeast Region was higher than NJ and Nation

No other comparisons were significant

* Statistically significant

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 Reading Assessment.

ELL for NJ Not Reported

- The NAEP ELL results for NJ were **unreliable** because:
 - Small sample sizes
 - High Standard Errors (SE)
 - NAEP sample sizes for NAEP are small due to exclusionary rules used since 1992
 - **NAEP 2007** -- exclusionary rule will be updated and aligned with the NJ rules
-

Pictures at an Exposition



Achievement for ALL Students

Students from
Low-Income
Families

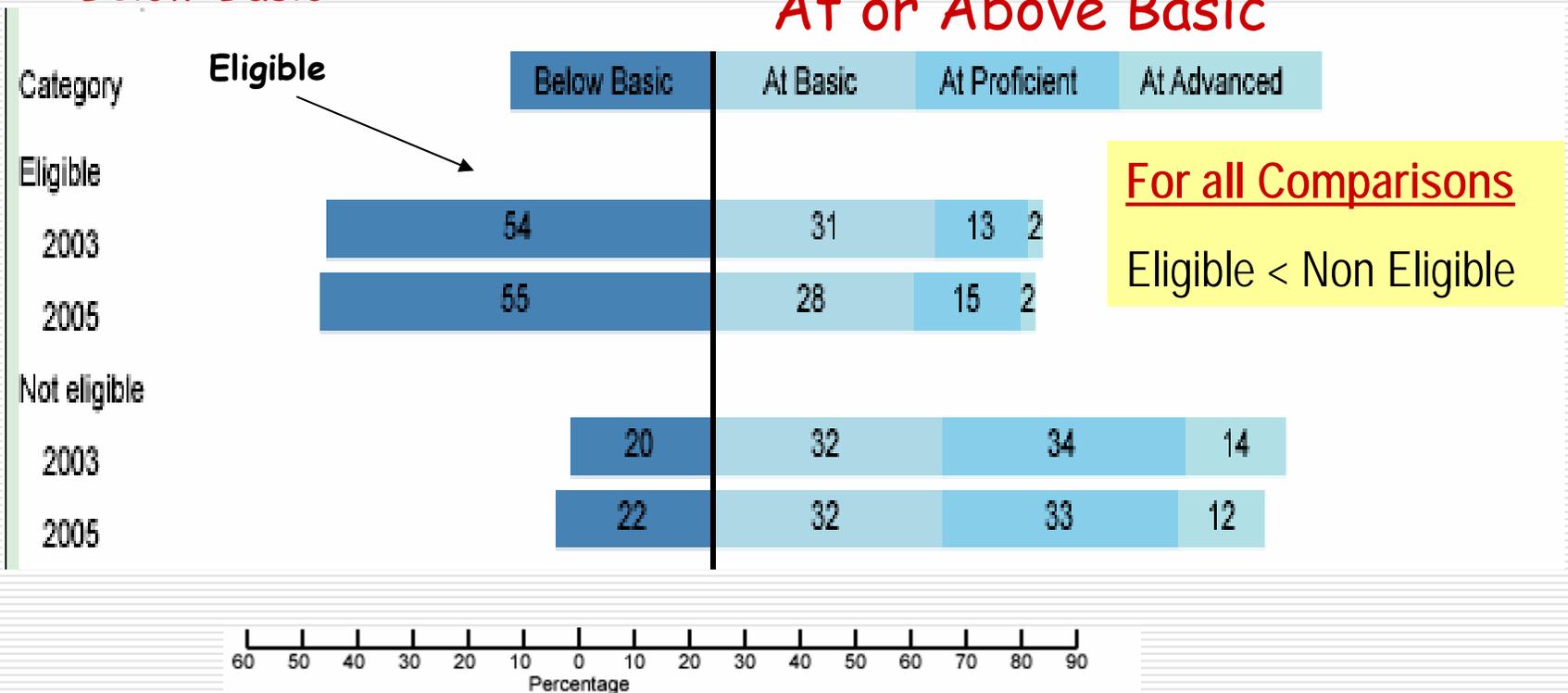
(Free/Reduced Lunch)

Low-Income Gap in NJ NAEP, At/Above Basic, Grade 4 (2003 and 2005)

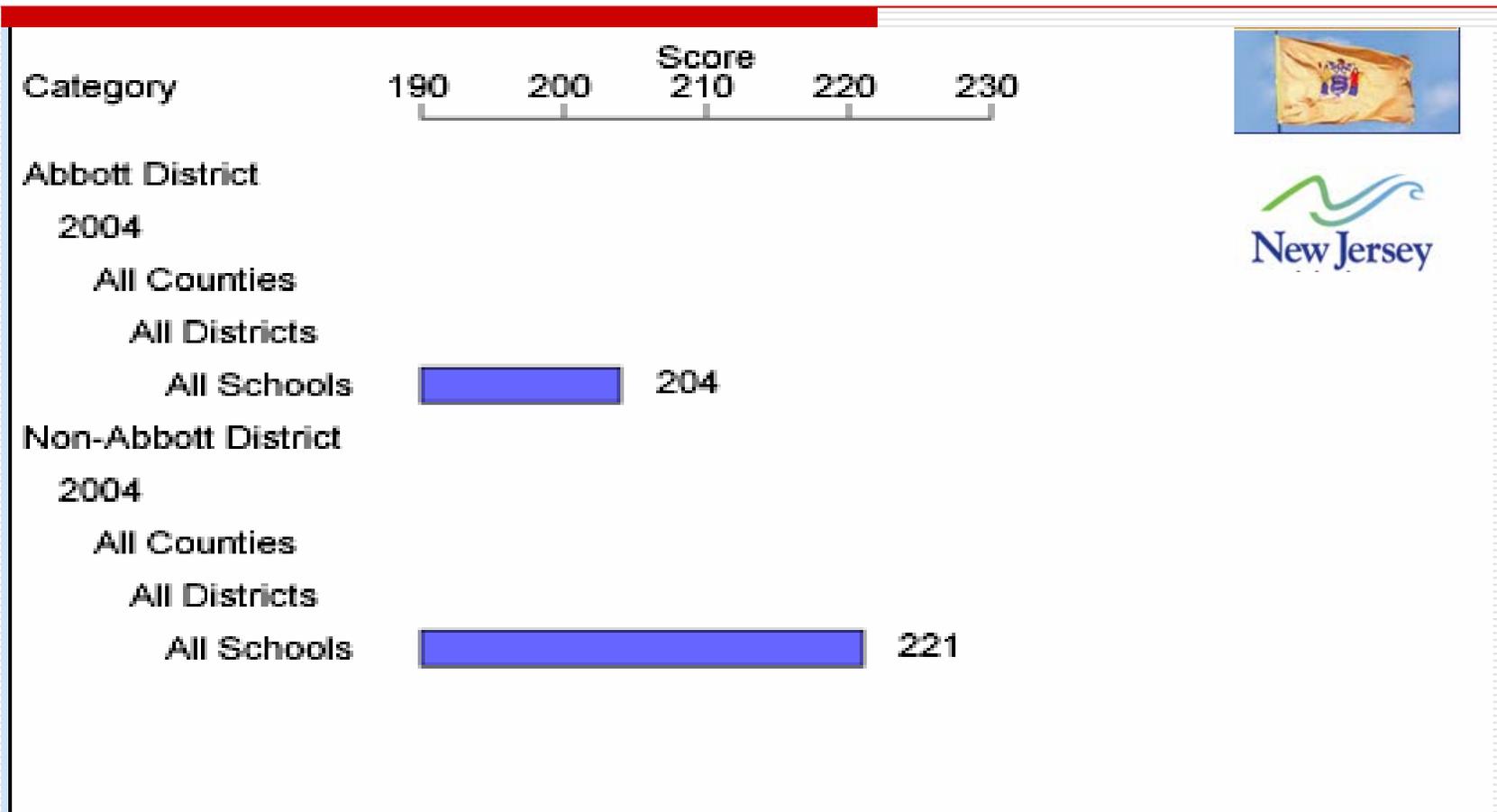
NJ Eligible for Free or Reduced Lunch

Below Basic

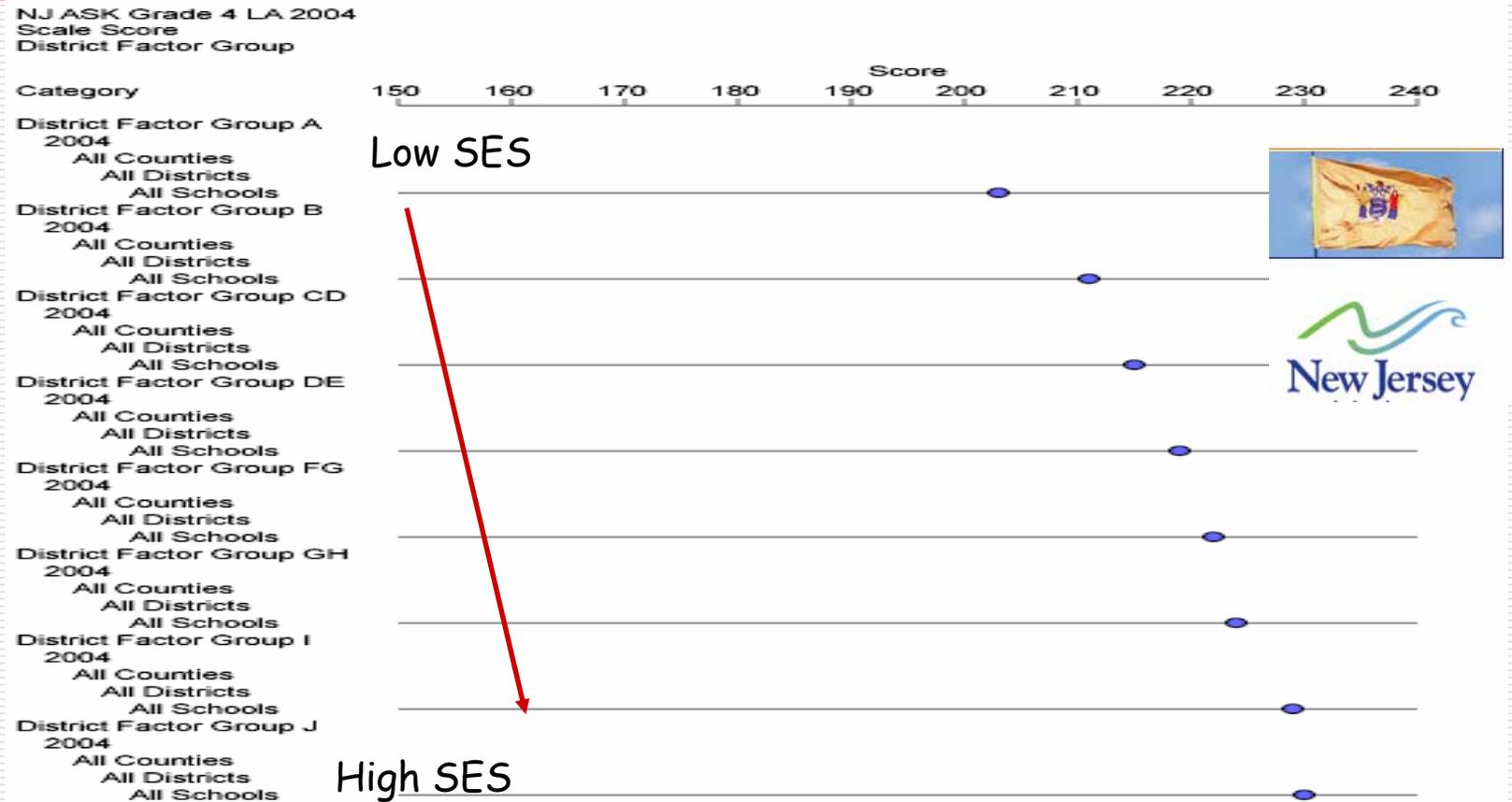
At or Above Basic



Low-Income's Impact on NJASK 4 Abbott/Non-Abbott Districts (2004)



NJASK 4 & the District Factor Groups



SOURCE: N.J Department of Education, NJASK, 2004 LA Assessments

From the Instructional Data Management System, www.idms.com

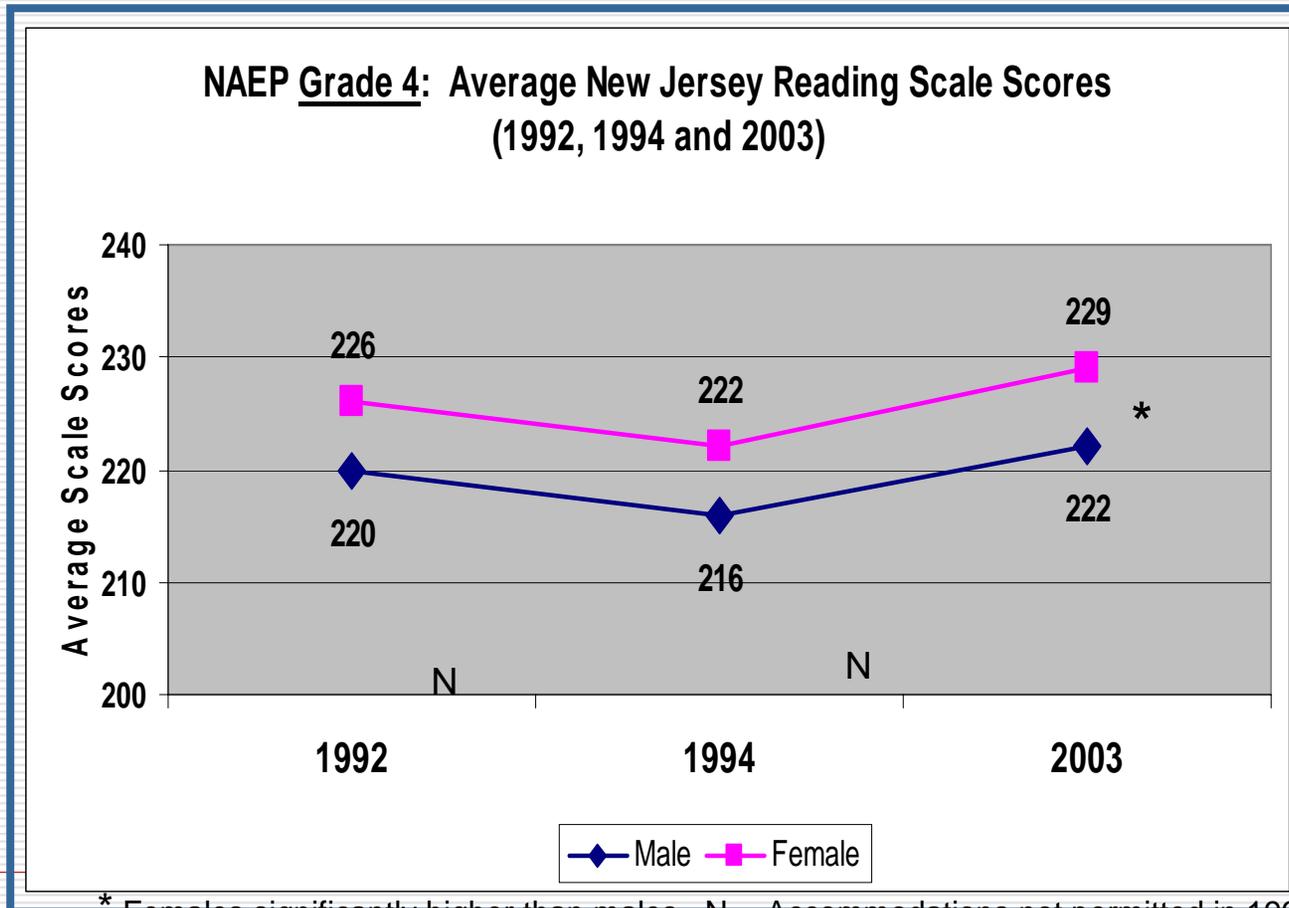
Pictures at an Exposition

Achievement for ALL Students

Gender



Gender Achievement Gap



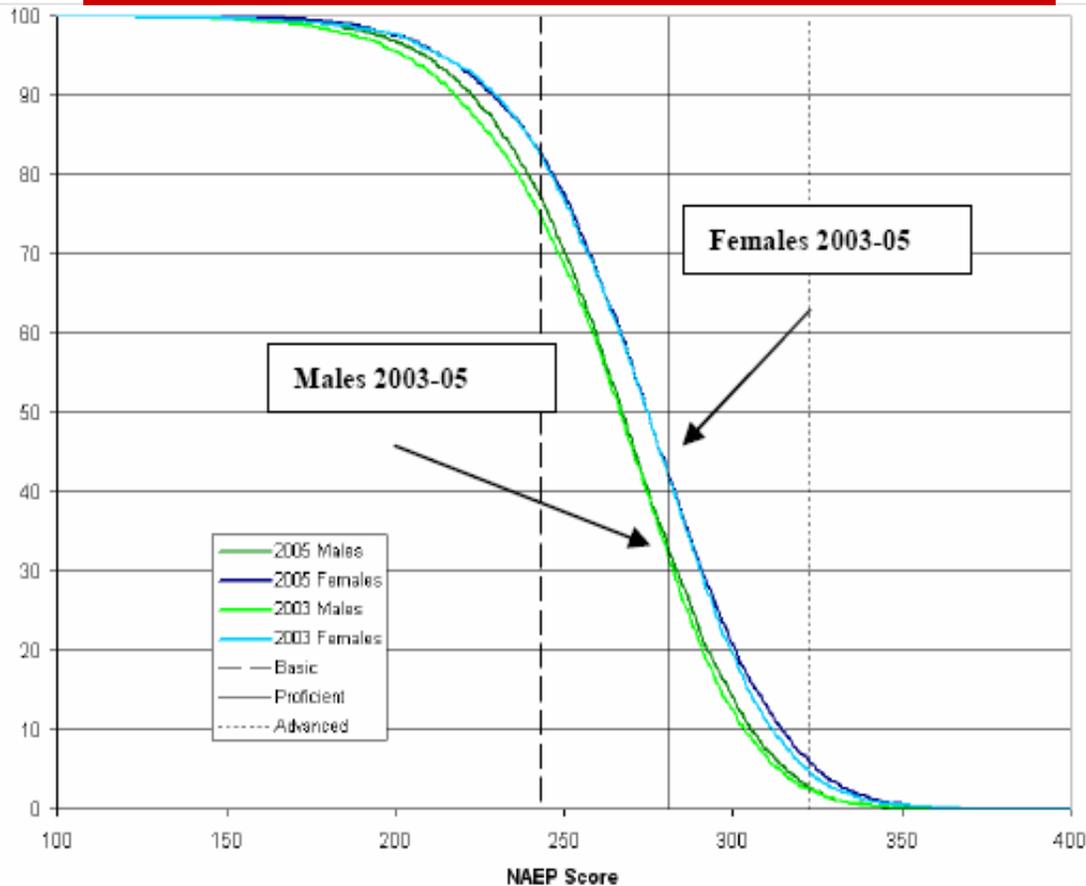
New Jersey

*Girls > Boys
(7 pts) in 2003*

*No Gap Change
from 1992 to
2005*

* Females significantly higher than males. N -- Accommodations not permitted in 1992 and 1994,

Gender Gap/Gap Change from 2003 to 2005, Grade 8

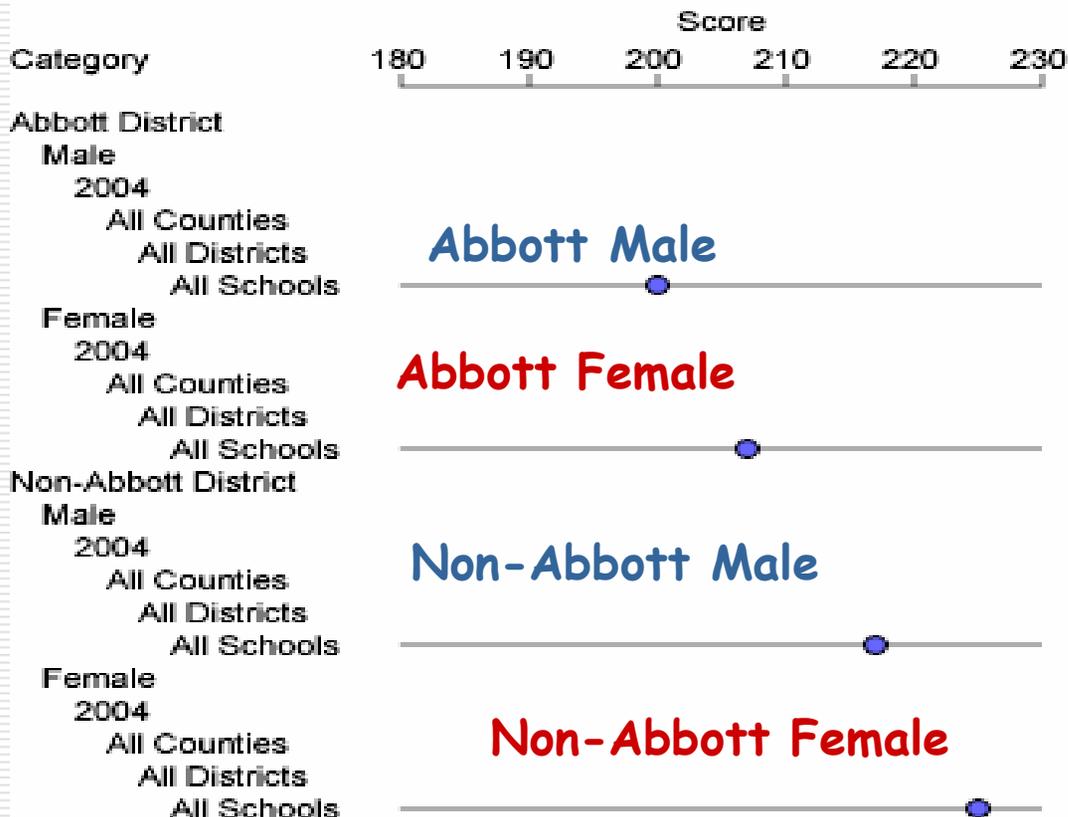


At or Above
Proficient
2003 - 2005

➤ Females > Males

No change in
gender gap
from 2003 to
2005

Gender Gap: NJASK 4 - Abbott/Non-Abbott (2004)



SOURCE: N.J. Department of Education, NJASK, 2003 and 2004 LA Assessments

From the Instructional Data Management System, www.idms.com

Gender Gap: NJ, US, NE, Grade 8

Scale Scores, Grade 8 Reading, 2005

	Females	Males	Scale Score Difference	Male/Female Gap
New Jersey	273	266	-8	Significant
National Public	266	255	-11	Significant
Northeast Census Region	272	263	-9	Significant

Pictures at an Exposition

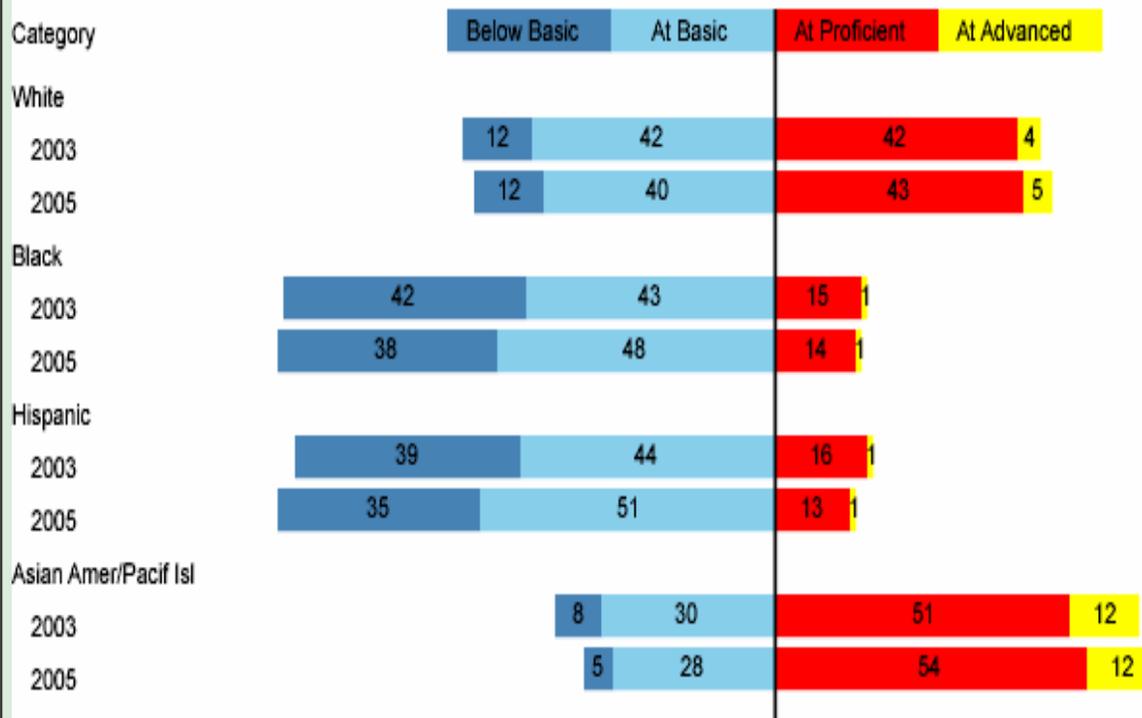


Achievement for ALL Students

Race/Ethnicity

Race Gap: NAEP Achievement Levels Grade 8 (2003 and 2005)

Percentages of students at each achievement level for reading, grade 8
 Race/ethnicity used in NAEP reports after 2001 [SDRACE]
 New Jersey, 2003 and 2005

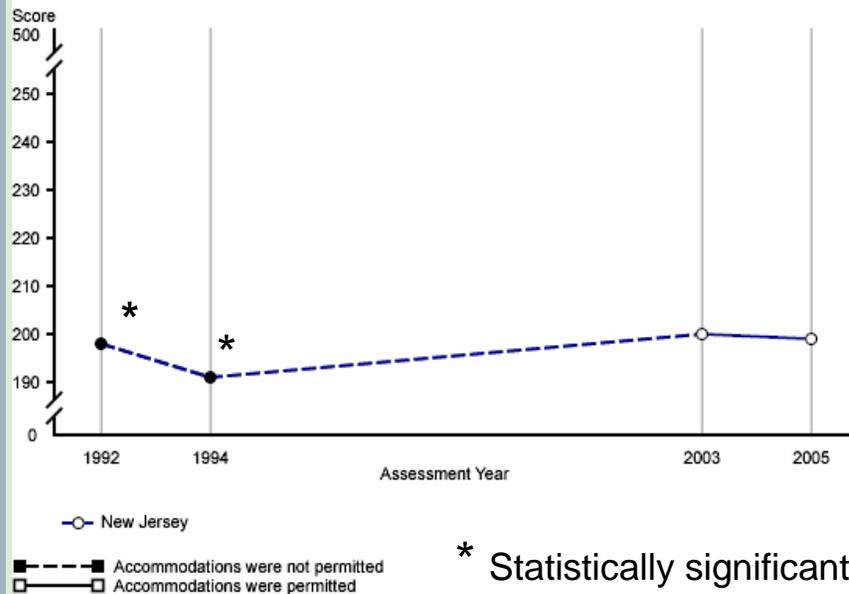


NAEP 2005 --Significance Percent At or Above Proficient

	W	B	H	A
White (48%)		>	>	<
Black (15%)	<		=	<
Hispanic (14%)	<	=		<
Asian (66%)	>	>	>	>

Race (1992-2005): Hispanics, Grade 4

Average scale scores for reading, grade 4
Race/ethnicity used in NAEP reports after 2001 [SDRACE] = Black
New Jersey, 1992, 1994, 2003 and 2005



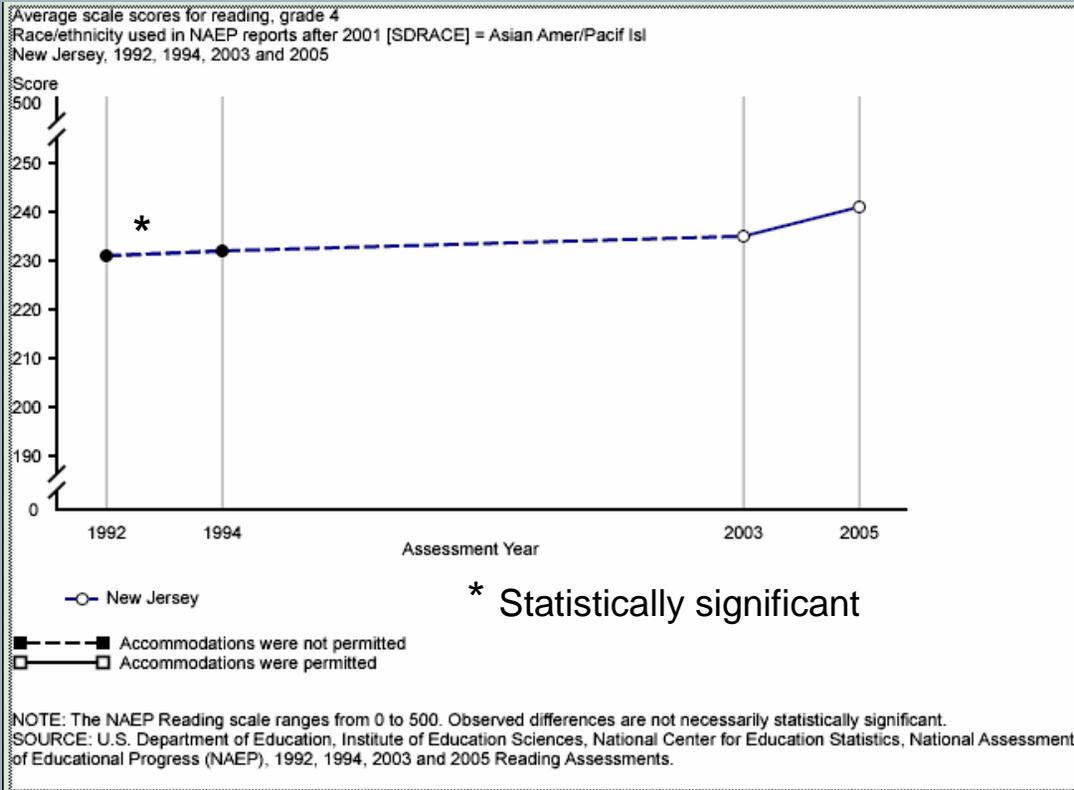
NOTE: The NAEP Reading scale ranges from 0 to 500. Observed differences are not necessarily statistically significant.
SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 1992, 1994, 2003 and 2005 Reading Assessments.

Good News:

Scores higher
in 2005 than in
1992 & 1994.

No change from
2003 to 2005.

Race (1992-2005): Asian/PI, Grade 4

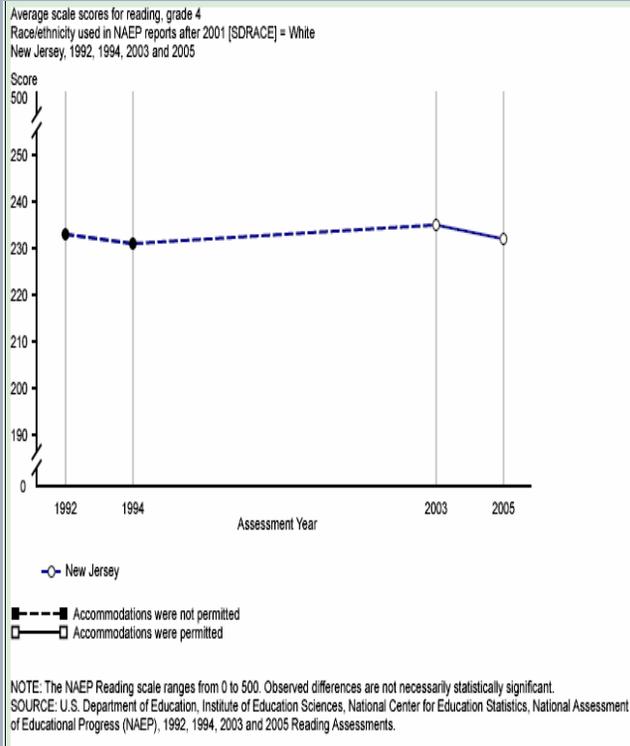


Good News:

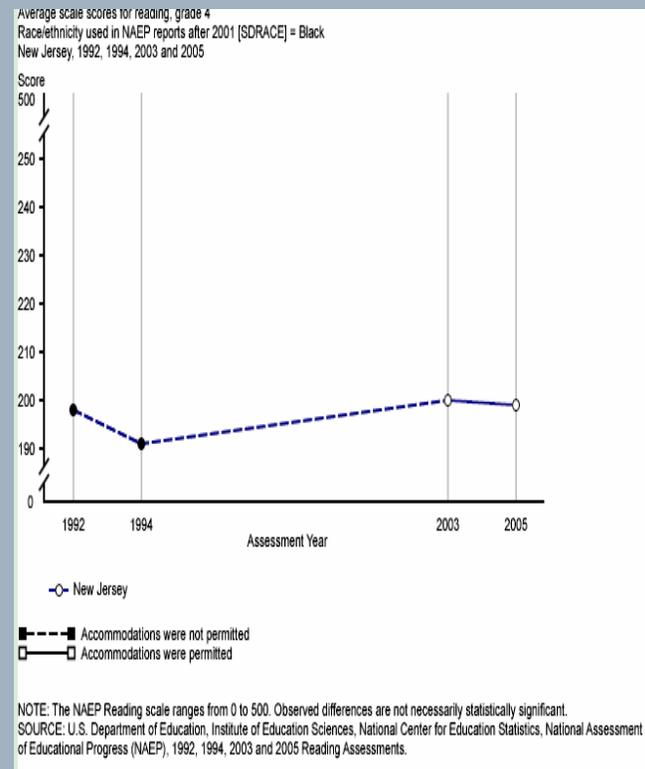
Scores higher
in 2005 than
in 1992

No change
from 2003 to
2005

Race (1992-2005): White and Black Results, Grade 4



White Students – No Change

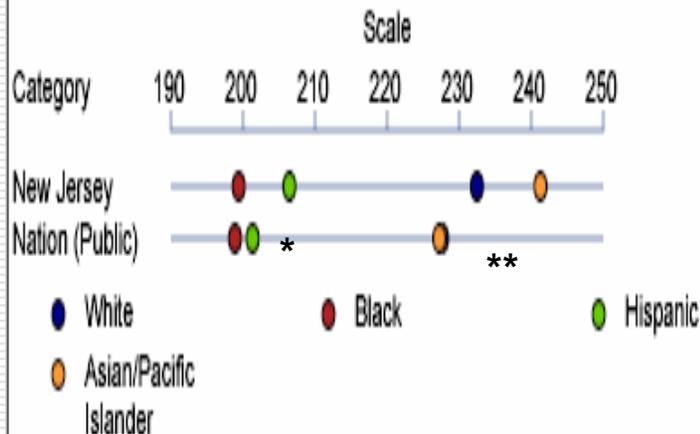


Black Students – No Change

Race Gaps: NJ and Nation, Grade 4

NAEP New Jersey Reading, Grade 4
Average Scale Score
For Race/Ethnicity, 2005

2005



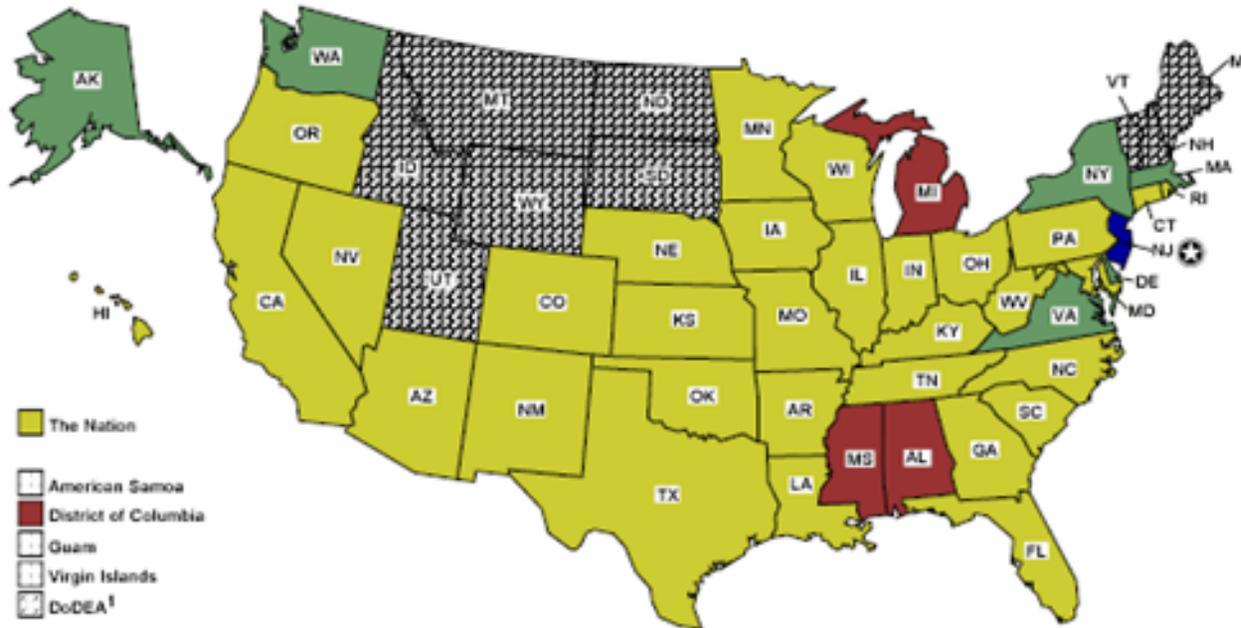
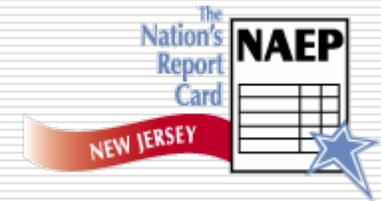
Average Scale Scores 2005	National Public
NJ White (232)	>
NJ Black (199)	=
NJ Hispanic (206)	>
NJ Asian/PI (241)	>

Note: The NAEP reading scale ranges from 0 to 500. Observed differences are not necessarily statistically significant.

SOURCE: U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP), 2005 Reading Assessment.

* Statistically significant

How Do NJ's Black Students Compare?



Grade 4,
2005

Only **6 states** have higher achievement for Black students

- Focal state/jurisdiction (New Jersey)
- Has a higher average scale score than the focal state/jurisdiction
- Is not significantly different from the focal state/jurisdiction
- Has a lower average scale score than the focal state/jurisdiction
- Did not participate or did not meet minimum participation rates
- Was not selected for comparison
- Sample size is insufficient to permit a reliable estimate

Why Flat Trend Data?



Digging Further?

Sometimes digging behind the "averages" can uncover an entirely different story



Bracey, G.W. (February 2004) *Simpson's Paradox and Other Statistical Mysteries*. *American School Board Journal*

Simpson's Paradox

- Simpson's Paradox is when the aggregate group score shows one pattern, but the subgroup scores show a different pattern
 - This phenomenon is seen in NAEP and SAT scores overtime
 - Subgroups show significant gains over time, but the overall national or state averages are flat
-

Simpson's Paradox: NJ's Trend Data for Race, Grade 4

		% of Testing Pop.	Av. Scale Score	Gains
All NJ Students	1992	100%	223	0
	2005	100%	223	
White	1992	69%	233	-1
	2005	58%	232	
Black	1992	16%	198	+1
	2005	17%	199	
Hispanic	1992	11%	195	+11
	2005	16%	206	
Asian	1992	4%	231	+10
	2005	8%	241	

- **Why no gains overall?**
- **Whites** -- Proportion of high scoring whites decreased, pulling down the average
- **Hispanics** -- Scores jumped, but averages are relatively low (206)
- **Asians** -- Scored jumped, but still proportion of the total population is low

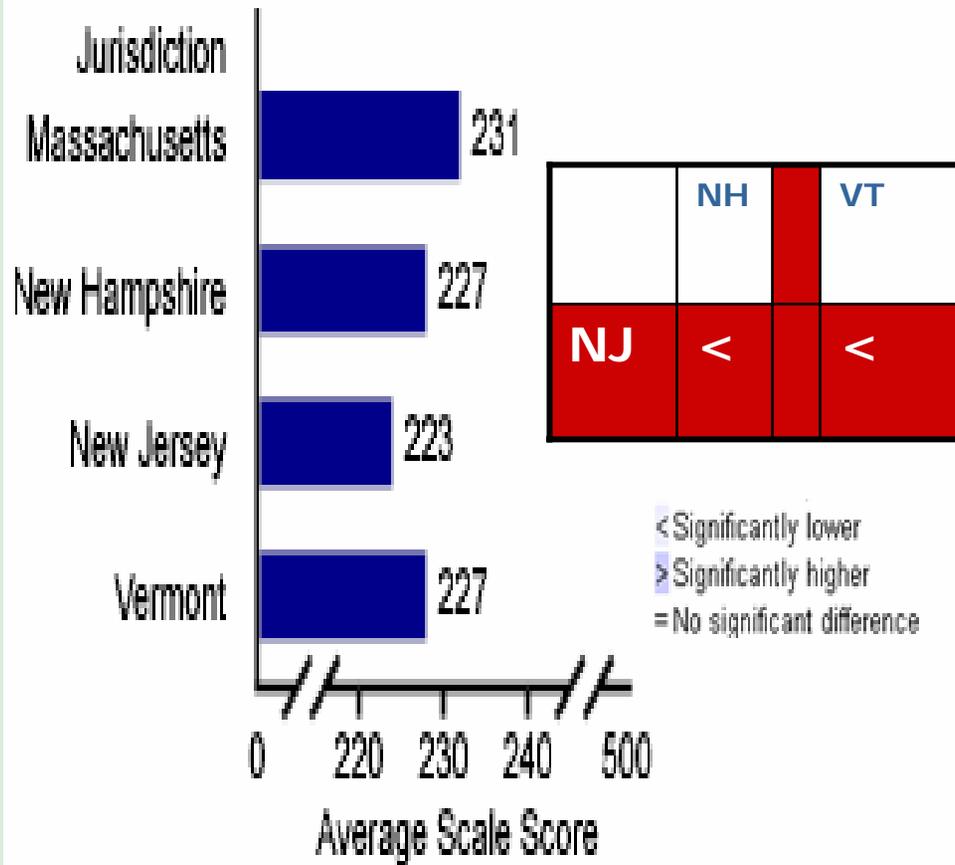
Simpson's Paradox

Another example of a Simpson's Paradox was provided by Howard Wainer

On the Academic Performance of New Jersey's Public School Children: Fourth and Eighth Grade Mathematics in 1992
(<http://epaa.asu.edu/epaa/v2n10.html>)

Examined performance in New Jersey in comparison to other states and the nation

Simpson's Paradox: NJ Compared to VT and NH, Grade 4 (2005)



➤ It appears that 4th grade students in NH, and VT do better in reading than their counterparts in New Jersey

Simpson's Paradox: NJ Compared to VT and NH, Grade 4 (2005)

	White	Black	Hispanic
New Jersey	58%	17%	16%
New Hampshire	94%	1%	2%
Vermont	96%	1%	1%

- It is caused by the **differences** in the **proportional representation** of the ethnic groups
 - Each state's mean score is a product of the mean score within each ethnic group and its proportional representation in the population
-

Simpson's Paradox: NJ Compared to VT and NH, Grade 4 (2005)

When we examine the average scale scores by ethnic group we find

	White	Black	Hispanic
New Jersey	232	199	206
New Hampshire	228	++	++
Vermont	227	++	++

++ Reporting standards not met

Simpson's Paradox

- Overall national and state results are informative at a general level
 - The **proportional representation** of subpopulations should be considered
 - To truly understand what students are learning, one must examine how the subgroups are performing
-

What Have We Learned?



This session
began and ended
with analyses of
state-level results

What Have We Learned?

NAEP's original question:

Are students in American schools
learning what they should be
learning?

What do you think?

What Can NAEP Tell Us?

What did we learn about

- NJ's overall reading performance?
 - NJ's subgroup reading performance?
 - Students with Disabilities
 - Students from Low-Income Families
 - Males and Females
 - White, Black, Hispanic and Asian/Pacific Islander students
-

Limitations to the NAEP Data

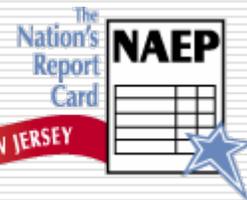
NAEP Data cannot tell us:

- Why gaps exist
 - Policies for closing gaps
 - Suggestions for interventions to improve student learning
-

Where to Look for Answers -

Educational Research, for example:

- Exemplary schools -- successful with low-performing students
 - What Works - a knowledge-base on effective schools and successful practices in the classroom
-



CONTACT INFORMATION

NJ NAEP Web pages: <http://www.nj.gov/njded/assessment/naep>

Barbara Smey-Richman, EdD
New Jersey Department of Education
Office of Evaluation and Assessment
P.O. Box 500
Trenton, NJ 08625-0500

Tel: (609) 984-1540

Fax: (609) 984-6032

Email: Barbara.Smey-Richman@doe.state.nj.us

Time to Hit the Beach?



Thanks for Your Interest in NAEP!!
