State of California **Department of Education**

LAST MINUTE MEMORANDUM

DATE: July 6, 2004

TO: MEMBERS, STATE BOARD OF EDUCATION

FROM: Geno Flores, Deputy Superintendent
Assessment and Accountability Branch

RE: Item No. 5

SUBJECT: Information regarding the National Assessment of Educational Progress (NAEP)

Presenters will include Charles Smith, Executive Director, National Assessment Governing Board; Dr. Steve Lazer, Vice President of Assessment Development and Executive Director of NAEP, Educational Testing Service; and Dr. Eric Zilbert, Education Research and Evaluation Consultant and California NAEP Coordinator, for California Department of Education.

The presentation will provide an overview of the National Assessment of Educational Progress and a comparison of NAEP to the Standardized Testing and Reporting (STAR) Program.

Attachment 1: The National Assessment of Educational Progress (NAEP)
An overview of NAEP and a comparison to the California STAR Language Arts Results (6 Pages)

The National Assessment of Educational Progress

An Overview of NAEP and a Comparison to the California STAR Language Arts Results

Presented by:

Charles Smith
Executive Director
National Assessment Governing Board

Stephen Lazer
Vice President, Assessment ETS
NAEP Executive Director

Goals of NAEP

- Report on what students know and can do in a variety of broadly defined subject areas
- Use methodologies appropriate to broad content measurement
- Provide a common yardstick against which all states are compared
- Provide solid information on educational progress

What NAEP is NOT

- A test of an individual student
- Something that gives individual children scores
- Something used for school building accountability

Program Constraints

- Need to cover broad content areas, and report on subscales
- Need to use constructed response items (roughly half the items on the grade 4 reading assessment, and a higher percentage at grade 8)
  - less on math
- Aggregate testing time
  - If one student took the entire assessment, would be extremely long
  - Therefore, no one student takes the entire assessment

Designs That Help With These Constraints

- Student Sampling: that is, use of representative samples of students
- Item Sampling: that is, a design in which no one individual student takes the entire assessment

Student Sampling

- Representative samples at the national and state levels
- State NAEP samples are generally 100-200 schools for each grade assessed
- In CA over-sampling in LA and San Diego have boosted state sample. For grade 4 reading:
  - In 1998, 1,713 out of 451,069 students participated in NAEP (less than 1%)
  - In 2002, 4,016 out of 491,510 students participated in NAEP (less than 1%)
  - Approximately 12,400 students will be in the 2005 sample

Why do NAEP and STAR not appear to show the same trends?

- Different instruments based on different frameworks
- Different instruments use different scales when reporting scores
- **Neither of these necessarily represents a problem with either measure**

Different frameworks

- NAEP and STAR measure different things
- NAEP is a reading assessment, while STAR combines reading and English language arts scores
- This may lead to different trends

Different assessments

- STAR is mainly a multiple-choice measure
- NAEP reading a mixed instrument made up largely of open-ended exercises (roughly 50% at grades 4 and slightly higher at grade 8)

4th Grade NAEP Examples

- Do you think this story was exciting? Use an example from the story to explain why or why not.
- What was one of the most important lessons that Cory and Elisa learned from their experience?

4th Grade STAR Examples

- **These passages are *best* described as**

A modern-day science fiction stories.

B fables about animals who learn a lesson.

C myths that explain how things began.

D fairy tales from two different countries.

4th Grade STAR Examples

- ***Both* the snake and the lizard change into men when they**

A are in water.

B see water.

C drink water.

D say “water.”

The Issue

- While there are reasons why these assessments tell different stories
  - Different frameworks
  - Different instruments
- The evidence indicates STAR and NAEP reading trends do not appear meaningfully different when the scales are standardized

NAEP vs. STAR

- Different assessments use different scores when reporting performance
- Different assessments use different scales
  - SAT-9 = 0 – 999 points
  - NAEP = 0 - 500 points

Grade 4 NAEP Reading vs. STAR Reading

- Stanford 9 1998 = 627
  - 40th percentile
- Stanford 9 2002 = 638
  - 50th percentile
- California NAEP 1998 = 202
  - National average = 213
- California NAEP 2002 = 206
  - National average = 217

Creating a common metric

- We must standardize the scales to allow comparisons across different measures
- We standardize the scales by creating a common metric called the effect size
- The effect size is calculated by dividing the average scale score by the standard deviation for the assessment

NAEP vs. STAR

- 1998 – 2002 NAEP as compared to SAT-9
  - Grade 4 gains are almost identical (.09 vs .11)
  - Grade 8 changes were very similar (-.06 vs +.03)
- The California NAEP changes are not statistically significant
- The changes are small for both programs

Conclusions

- Performance on NAEP and SAT-9 were comparable between 1998 and 2002 for both grades 4 and 8 in reading
- While differences exist, both exams provide valid and reliable scores
- Differences may continue to exist as long as the frameworks are different and the instruments are constructed differently
- In 2005 it will be possible to compare NAEP with the California Standards Tests (CSTs) over a 3 year period