COMPARING ACHIEVEMENT BETWEEN OAKS AND SMARTER BALANCED

SUMMER ASSESSMENT INSTITUTE

DEREK BROWN, DIRECTOR OF ASSESSMENT

JON WIENS, MANAGER, ACCOUNTABILITY AND REPORTING

STEVE SLATER, MANAGER, PSYCHOMETRICS

2014-15 SMARTER BALANCED ASSESSMENT RESULTS

SMARTER BALANCED CUT SCORES

- The Level 3 standard is intended to represent the academic readiness threshold for success in non-remedial college courses.
- OUS and Oregon Community colleges have announced placement policies for high school students achieving at Level 3 and Level 4.
- Graduation Standards (i.e., Essential Skills) are likely to be lower.

SMARTER BALANCED TESTS ADMINISTERED IN 2014-15

- Participation: must respond to at least 5 items on the CAT and 1 item on the PT.
- A fairly large number of students started the assessment but were not counted as participants.

Subject	Test Name	Number of Tests Started	Number of Tests Completed
English Language Arts	Compute Adaptive (CAT)	285,345	283,011
	Performance Task (PT)	284,129	281,168
	Total Participants	282,037	
Mathematics	Compute Adaptive (CAT)	284,482	281,887
	Performance Task (PT)	283,258	282,807
	Total Participants	281,651	

PRELIMINARY PARTICIPATION

- Participation: must respond to at least 5 items on the CAT and 1 item on the PT.
- Rates include extended assessments.

PRELIMINARY Participation Rates

Grade	ELA	Math	Science
3	97.0	96.9	
4	97.6	97.4	
5	97.5	97.2	97.9
6	97.2	97.1	
7	96.9	96.6	
8	96.3	95.9	90.5
11	88.8	87.2	80.2
All Grades	96.0	95.5	89.6

PRELIMINARY Participation Rates

Student Group	ELA	Math
Economically Disadvantaged	96.8	96.4
English Learners	98.3	98.3
Students with Disabilities	93.7	93.1
American Indian/AK Native	95.9	95.5
Black/African American	93.4	92.6
Hispanic/Latino	97.4	97.0
Pacific Islander	97.0	96.5
Asian	97.0	96.6
White	95.5	95.0
Multi-racial	95.3	94.8

PRELIMINARY SMARTER RESULTS

- Scoring: ODE has reviewed the partially completed tests and arrived at the following rules:
  - All participants will receive a score.
  - Most students with partially completed tests will be at level 1.
- Data below shows includes partially completed tests.

English Language Arts

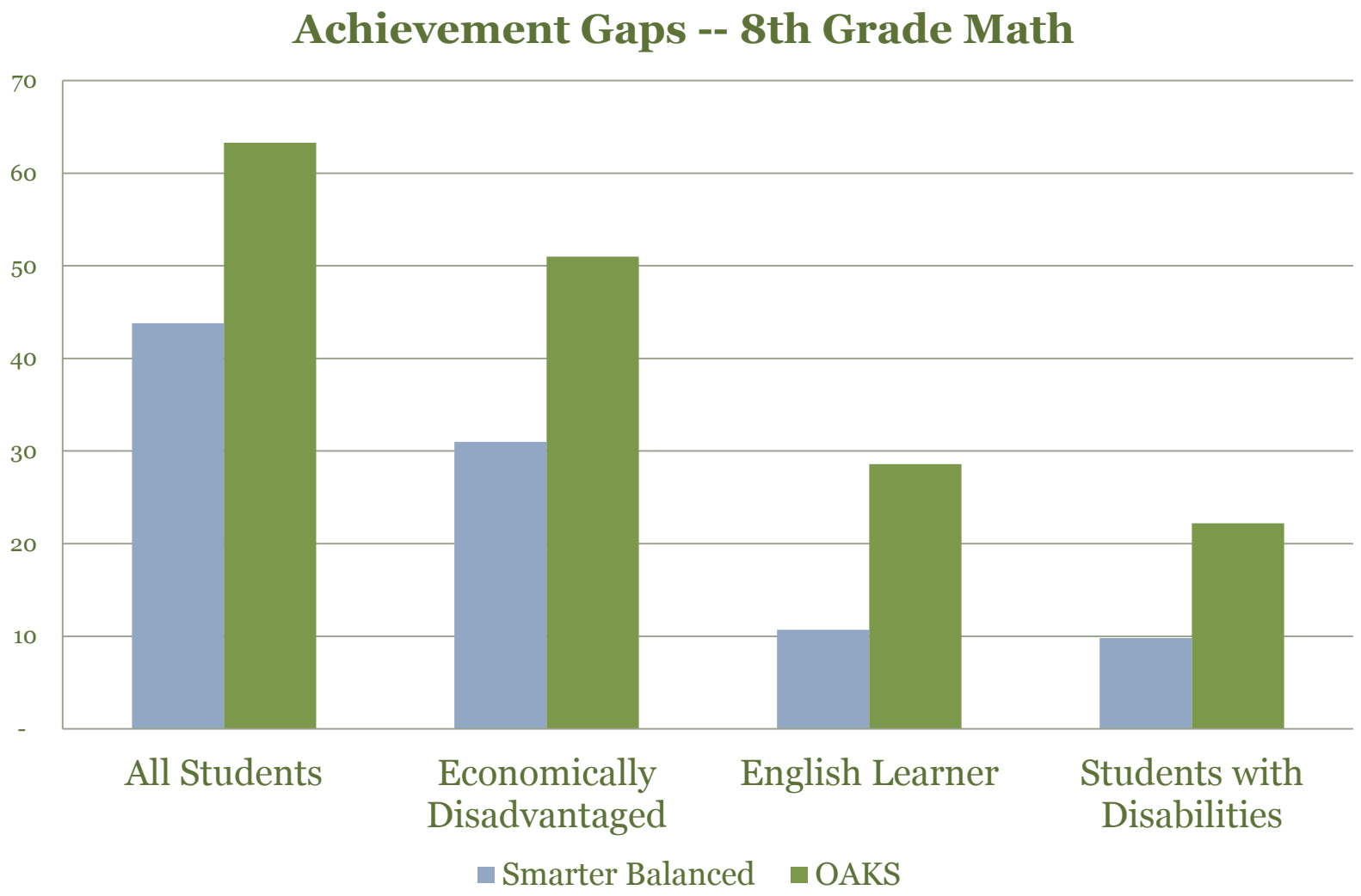
Grade	SB Tests (Expected 41K-42K)	Level 3 or Higher		
		Oregon %	Field Test %	Improve-ment
3	40,847	47	38	9
4	40,129	51	41	10
5	40,547	55	44	11
6	40,142	54	41	13
7	39,811	57	38	19
8	40,325	58	41	17
11	35,756	69	41	28

Mathematics

Grade	Students (Expected 41K-42K)	Level 3 or Higher		
		Oregon %	Field Test %	Improve-ment
3	41,169	47	39	8
4	40,316	45	37	8
5	40,435	42	33	9
6	40,363	39	33	6
7	39,855	44	33	11
8	40,417	44	32	12
11	35,482	31	33	(2)

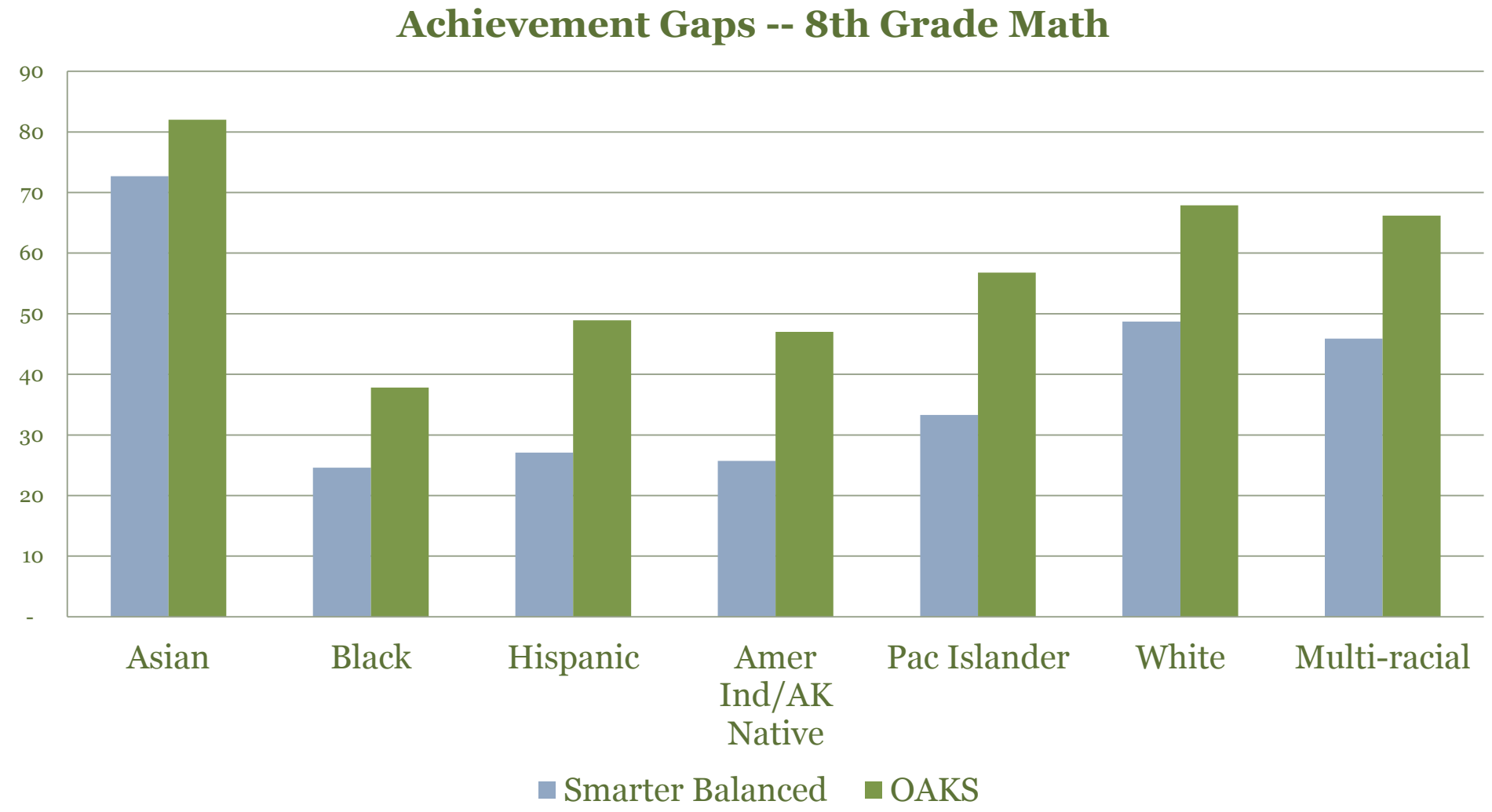
SMARTER BALANCED ACHIEVEMENT GAPS

We are also looking at Achievement Gaps on Smarter versus OAKS.



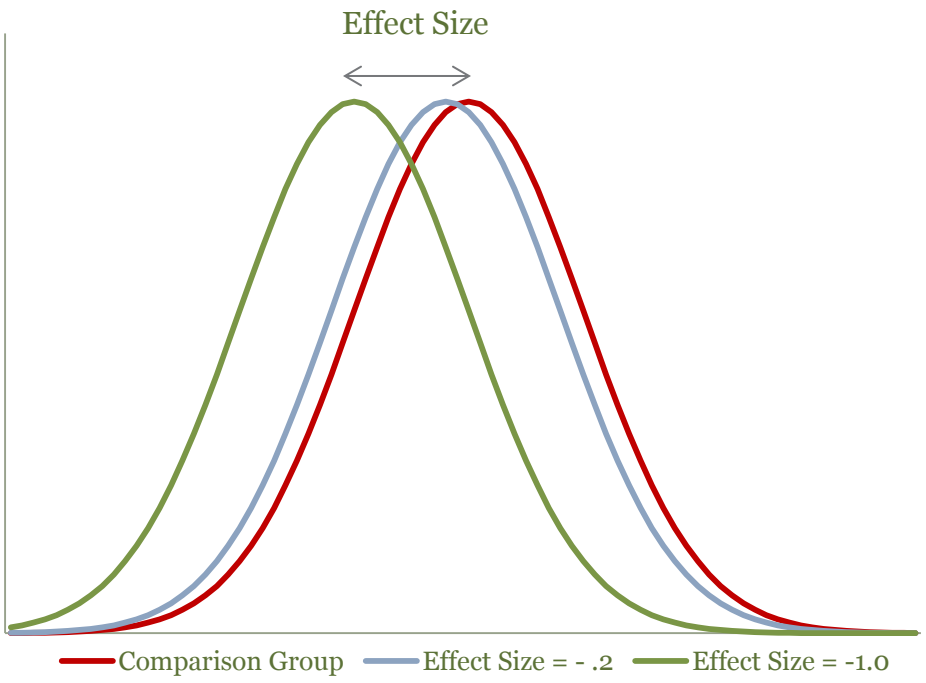
SMARTER BALANCED ACHIEVEMENT GAPS

We are also looking at Achievement Gaps on Smarter versus OAKS.



EFFECT SIZES

- While we can compare percentage meeting on two different assessments to compare achievement gaps, there are more reliable options.
- Effect Size – computes a "standardized" achievement gap that can be compared across assessments.
  - Uses average scores for subgroups.
  - Divides by the standard deviation for an assessment (measure of the spread of scores)
  - Creates a measure of how different the score distribution is for two different groups.



INTERPRETING EFFECT SIZES

- One Interpretation – what would the score percentile be for the "average" (50th percentile) student in one group, if placed in another group?
  - Data here are preliminary, and may change when some student's tests are rescored upwards.

Effect Sizes – Grade 8
Where would the average student in a group place if he/she were in the White Group, expressed as a percentile

	ELA/Reading		Mathematics	
Student Group	**OAKS**	**Smarter**	**OAKS**	**Smarter**
Economically Disadvantaged	32	38	34	35
English Learners	4	13	10	13
Students with Disabilities	12	20	14	16
American Indian/AK Native	29	36	30	30
Black/African American	24	31	23	23
Hispanic/Latino	28	36	32	33
Asian	58	60	74	71

GROWTH

- We have run the growth model to evaluate the transition from OAKS to Smarter Balanced.

PRELIMINARY Median Growth Percentile – All Grades				
	ELA/Reading		Mathematics	
Student Group	OAKS	Smarter	OAKS	Smarter
Economically Disadvantaged	47	47	47	46
English Learners	47	53	47	44
Students with Disabilities	43	40	42	42
American Indian/AK Native	45	39	48	43
Black/African American	45	42	43	42
Hispanic/Latino	47	51	47	46
Pacific Islander	45	52	50	49
Asian	58	69	61	62
White	51	49	50	51
Multi-racial	50	50	50	51

LINKING THE SMARTER AND OAKS SCALES

WHY LINK THE SCALES?

- Requirements for Essential Skills are set forth in OAR 581-022-0615:
  - (6) The Superintendent of Public Instruction shall establish an Assessment of Essential Skills Review Panel (AESRP) to make recommendations.
  - (7) The AESRP shall work toward the goal of a system with a high degree of technical adequacy and equivalent rigor between assessment options as practicable.
  - (14)(b) Students may use achievement standards adopted in their 9th through 12th grade years that are equal to or lower than the achievement standards approved as of March 1 of the students' 8th grade year.
- This rule requires that we find scores on Smarter Balanced that correspond to the old "meets" cuts on OAKS.
- We are working with AESRP to set these equivalents.

GUIDING PRINCIPLES

We are using the following principles as we work on the linking between the two sets of tests:

- A student who met the Essential Skills standard under the OAKS assessment should also meet the standard under the Smarter Balanced assessment

- The level of performance set by the state Board of Education for the Essential Skills requirement should be comparable

LINKING METHODOLOGIES

ODE Staff have been looking at three different methods to link the tests:

- Method 1: Match the statewide distribution of students on both assessments (OAKS 2013-14 and Smarter Balanced 2014-15)
- Method 2: Compare pairs of scores of students who "double tested" in 2014 on the Smarter Balanced field test and OAKS operational assessment
- Method 3: Embed OAKS items in Smarter Balanced operational assessments and link the two scales through item difficulty parameters

LINKING METHOD 1

Match the statewide distribution of students on both assessments (OAKS 2013-14 and Smarter Balanced 2014-15)

Advantages:

- Student motivation may have been more comparable than with Method 2 (field test direct link),
- Large sample size allows the comparison of linking results across student groups

Limitations:

- Unknown effect of "opt-outs" (estimated at 8-10%)
- Unknown effect of one additional year of common core instruction
- Unknown effect of multiple opportunities on OAKS vs. a single opportunity on Smarter Balanced
- Students who already met the Essential Skills on OAKS may have had less incentive to perform well on the Smarter Balanced assessments

LINKING METHOD 2

Compare pairs of scores of students who "double tested" in 2014 on the Smarter Balanced field test and OAKS operational assessment.

Advantage:

- Common student design controls for population differences that might exist in the Method 1 (assumption of randomly equivalent groups is not critical, since there is only one group)

Limitations:

- Unknown effect of lower motivation on the Smarter Balanced field test (scores did not count and were not returned)
- Reading scores (Claim 1) and writing scores (Claim 2) on the Smarter Balanced assessment were based on relatively few items, resulting in lower precision
- Relatively small sample size (1000-1300) reduces the ability to compare linking results across student groups

LINKING METHOD 3

Embed OAKS items in Smarter Balanced operational assessments and link the two scales through item difficulty parameters.

Advantage:

- Reduces the effect of differential motivation

Limitations:

- OAKS employs a limited set of item types (multiple choice and graphic response), limiting generalizability to the Smarter Balanced measures (i.e., student familiarity with OAKS item types, and conversely, unfamiliarity with Smarter Balanced item types that are not included in the analysis, may bias the result)
- Relatively small sample size (N ~ 750) reduces the ability to compare linking results across student groups

EARLY INDICATIONS FROM LINKING

- The Assessment of Essential Skills Review Panel (AESRP) is reviewing the linking data.
  - Method 1 (matching distribution) and Method 2 (analysis of double-tests) are producing similar results.
  - Method 3 (item embedding) is producing somewhat higher Smarter Equivalents.
- The AESRP will meet over the coming month to review the data and make recommendations for the State Board at their September 17th meeting.
- We are looking for adoption no later than the October Board meeting.