Academic Growth and Gaps for Specific Student Subgroups on a State Accountability Reading Test

Joseph J. Stevens

University of Oregon

Presented at the National Conference on Student Assessment conference, San Diego, CA, June, 2018

NCAASE National Center on Assessment and Accountability for Special Education

Advancing research on growth measures, models, and policies for improved practice

http://www.ncaase.com/

Contact Information:

Joseph Stevens, Ph.D.
College of Education
5267 University of Oregon
Eugene, OR 97403
(541) 346-2445
stevensj@uoregon.edu

Presentation available on NCAASE web site: http://www.ncaase.com/

NCAASE National Center on Assessment and Accountability for Special Education

Advancing research on growth measures, models, and policies for improved practice

http://www.ncaase.com/

Presentation Purpose

- Discuss issues in estimating and understanding achievement gaps.

- Discuss importance of directly testing interaction effects (see Stevens & Schulte, 2016), i.e., more disaggregation of groups.

- Describe results of an ongoing study of reading achievement growth for students with learning disabilities (SLD) and English Learners (EL) on Arizona state reading test.

- Because of our short time, we only present some highlights:
  - Reading achievement over time for SLD vs. not-SLD and for EL vs. not-EL.
  - Reading achievement over time for the SLD-EL interaction effect.
  - Differences in SLD-EL subgroup performance (i.e., achievement gaps) expressed as effect sizes (ES).

NCAASE National Center on Assessment and Accountability for Special Education
Advancing research on growth measures, models, and policies for improved practice

http://www.ncaase.com/

3

Interactions of Disability Status and Student Characteristics

- Many studies do not directly test the interaction of SWD status and other student characteristics thought to be related to student performance (e.g., LD status and sex of student).

- In many published studies, these variables are included only as the singular effect of the variable not the actual interaction effect (e.g., combined characteristics like LD-male versus LD-female).

- This can be very misleading and may result in incorrect interpretations as well as incomplete understanding of group differences. See:

Stevens, J. J., & Schulte, A. C. (2016). The interaction of learning disability status and student demographic characteristics on mathematics growth. *Journal of Learning Disabilities*. DOI: 10.1177/0022219415618496

NCAASE National Center on Assessment and Accountability for Special Education
Advancing research on growth measures, models, and policies for improved practice

http://www.ncaase.com/

Interactions of Disability Status and Student Characteristics

- Our purpose in the current study, therefore, was to test true interactions of SLD status with several other student characteristics.

- We were also interested in the size of the achievement gap in these comparisons and in whether the achievement gap was increasing or decreasing over grades.

- In this presentation we present selected results showing SLD interactions with EL status, although we have also analyzed interactions of SLD status with economic disadvantage, and with Hispanic vs. White race/ethnicity.

NCAASE National Center on Assessment and Accountability for Special Education
Advancing research on growth measures, models, and policies for improved practice
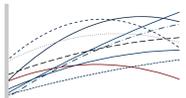
http://www.ncaase.com/

Analytic Methods

- We used hierarchical linear models (HLM) to test the interaction effects over Grades 3 to 6.

- Briefly the form of the two-level (grades and students) random intercepts and random slopes HLM model was:

Level-1 Model: $\qquad Y_{ti} = \pi_{0i} + \pi_{1i}*(\text{Time}_{ti}) + \pi_{2i}*(\text{Time}^2_{ti}) + e_{ti}$ $\qquad$ (1)

Level-2 Model: $\qquad \pi_{0i} = \beta_{00} + \beta_{01}*(\text{Predictor}_i) + r_{0i}$ $\qquad$ (2)

$\qquad\qquad\qquad\qquad \pi_{1i} = \beta_{10} + \beta_{11}*(\text{Predictor}_i) + r_{1i}$ $\qquad$ (3)

$\qquad\qquad\qquad\qquad \pi_{2i} = \beta_{20} + \beta_{21}*(\text{Predictor}_i) + r_{2i}$ $\qquad$ (4)

NCAASE National Center on Assessment and Accountability for Special Education
Advancing research on growth measures, models, and policies for improved practice
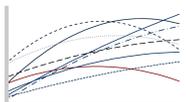
http://www.ncaase.com/

6

Method

- Student scores on the reading subtest of the Arizona Instrument to Measure Standards (AIMS) used for analyses.

- Sample details:
  - Sample size, N = 82,675 in Grade 3
  - Race/ethnicity composition in percent was 2.8 Asian, 5.6 Black, 43.3 Hispanic, 5.2 American Indian, 43.1 White.
  - 48.8% of the students were female; 12.9% were SWD; 19% were EL; 51.2% were economically disadvantaged.

- We examined attrition of the sample over grades; compared to Grade 3, 94% were present in Grade 4, 91% in Grade 5, and 87% in Grade 6.
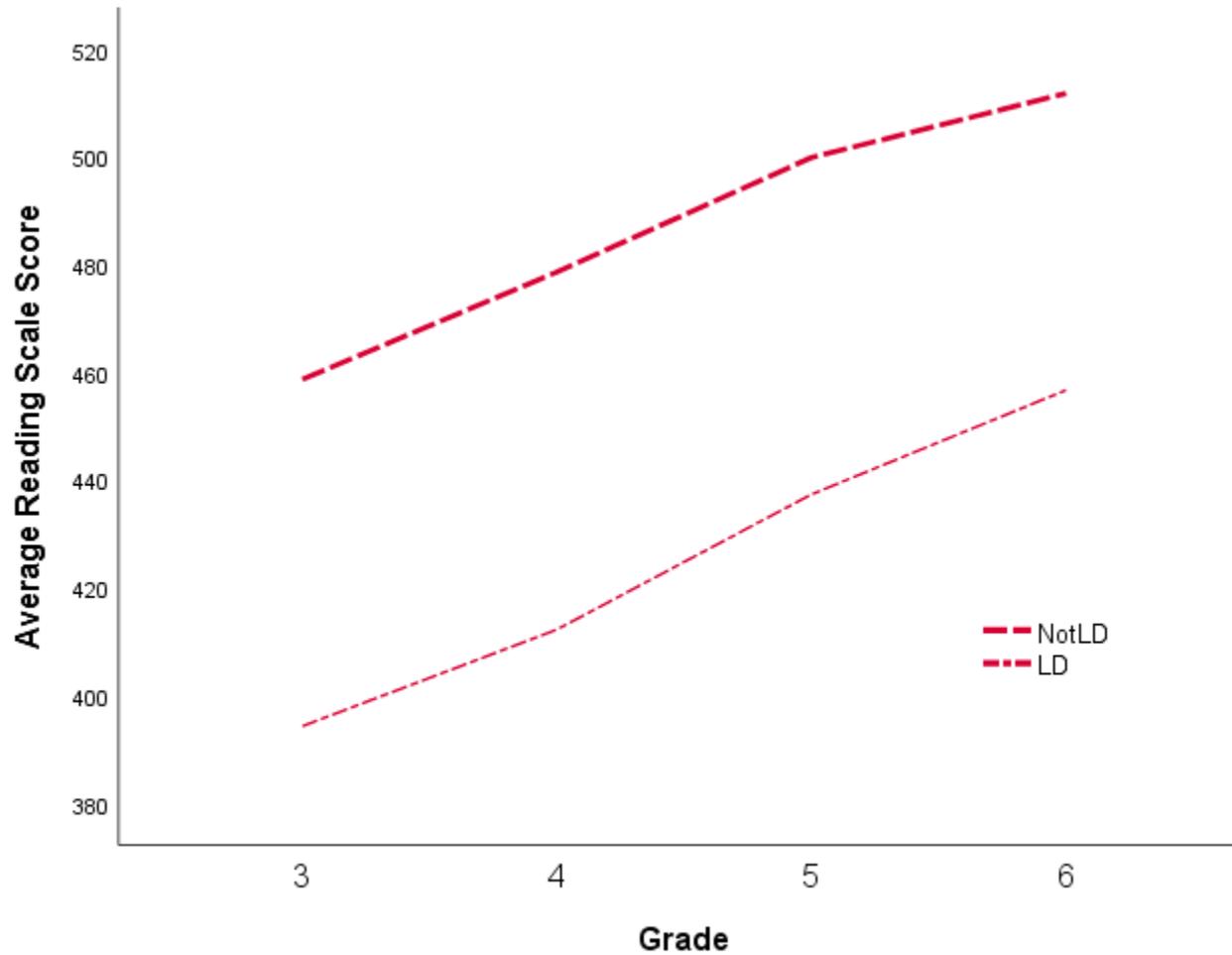
Further details on sample, methods and procedures available on request from the first author.

NCAASE National Center on Assessment and Accountability for Special Education
Advancing research on growth measures, models, and policies for improved practice

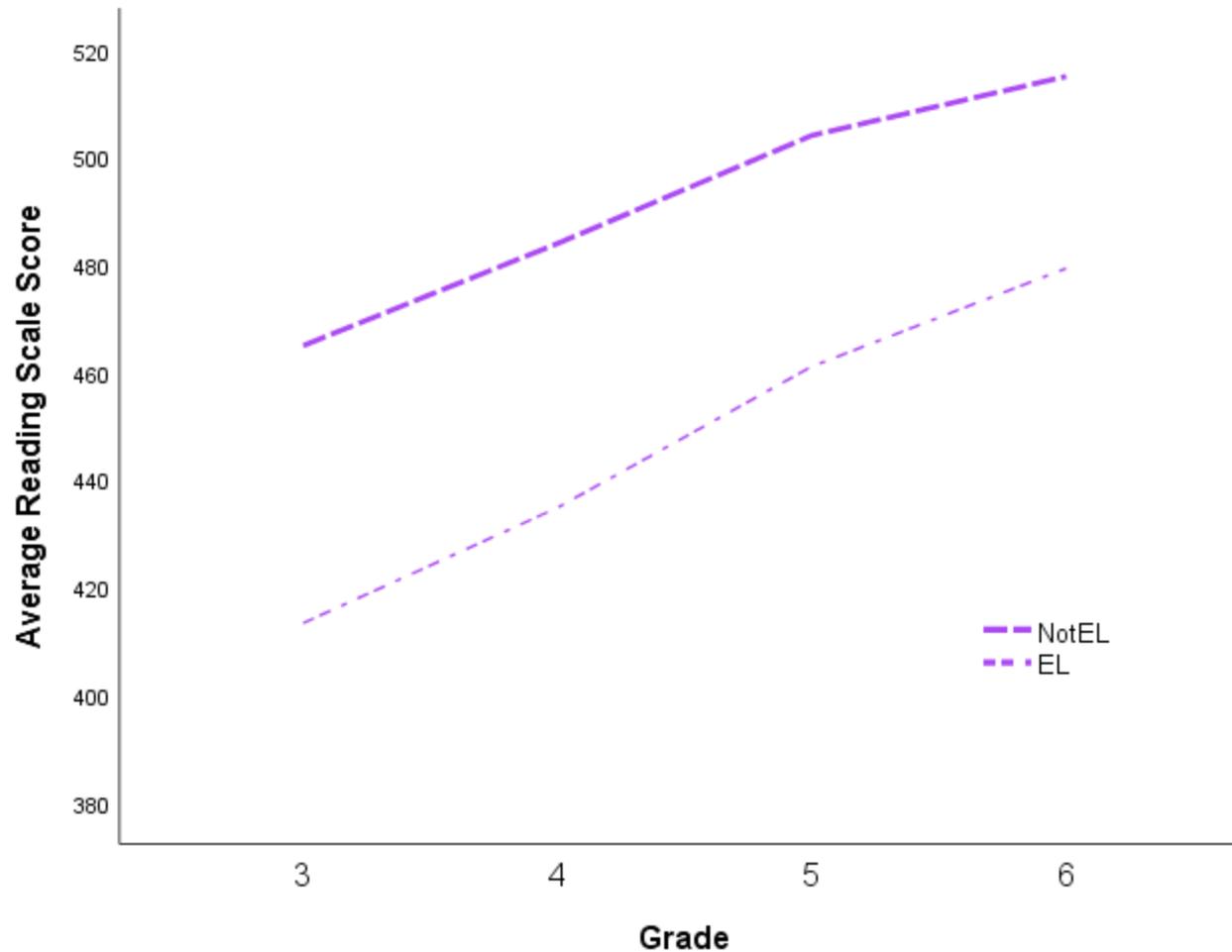http://www.ncaase.com/

Results

- In all the analyses we conducted, the interaction effect was statistically significant (i.e., LD X EL, LD X ECD, EL X ECD, LD X Hispanic, EL X Hispanic).

- For brevity, we only present graphical displays of the key results here for illustration.

- We then provide some summaries of the size of achievement gaps expressed as effect sizes (ES).

NCAASE National Center on Assessment and Accountability for Special Education
Advancing research on growth measures, models, and policies for improved practice

http://www.ncaase.com/

8

Single Variable Comparison of LD status

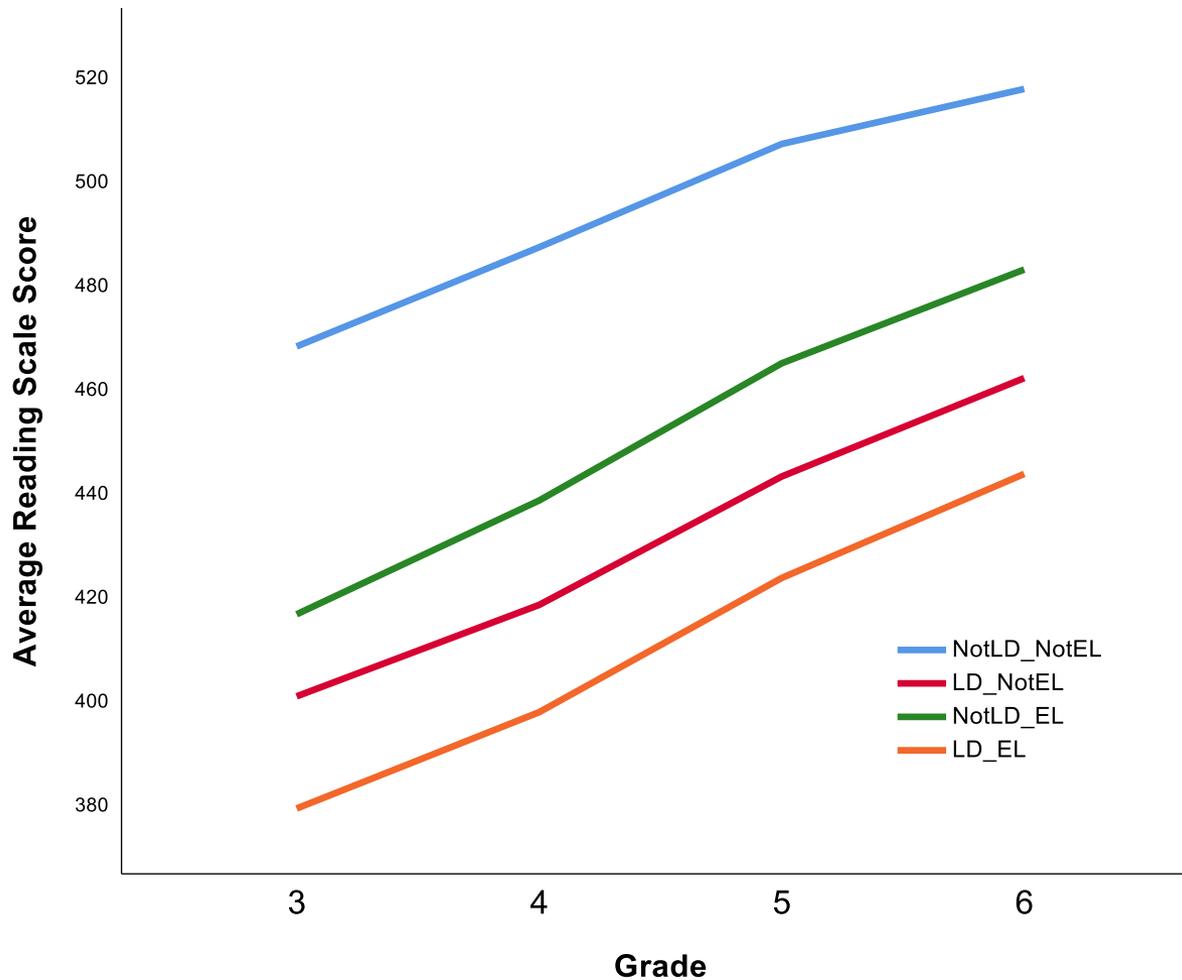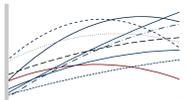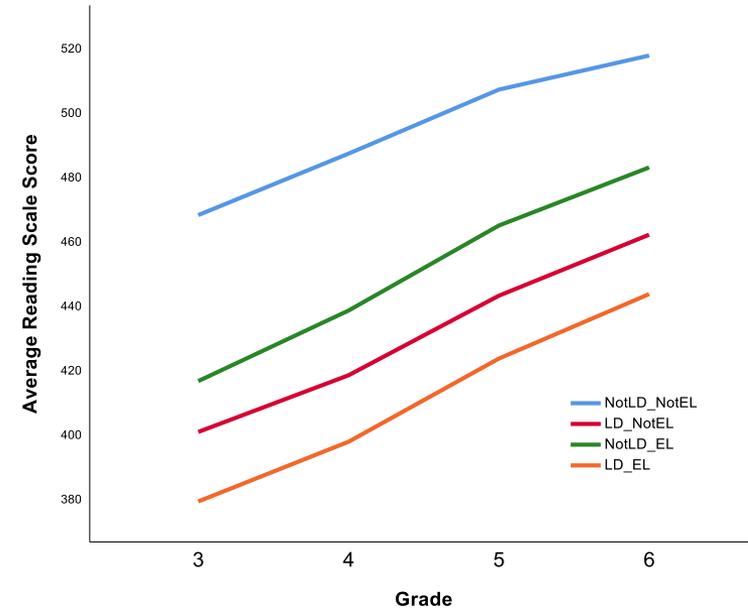# Single Variable Comparison of EL status

Figure 1. Three-way interaction of SLD status, EL status, and grade for the Arizona sample.

Results

- Note generally parallel growth over grades.

- Some closing of the gap for NotLD-EL students.

- Good news: all students groups are progressing in a similar way over grades.

- Bad news is that largely the gap is not closing.

- But let's also compare the "single variable" results shown earlier to these interaction results.
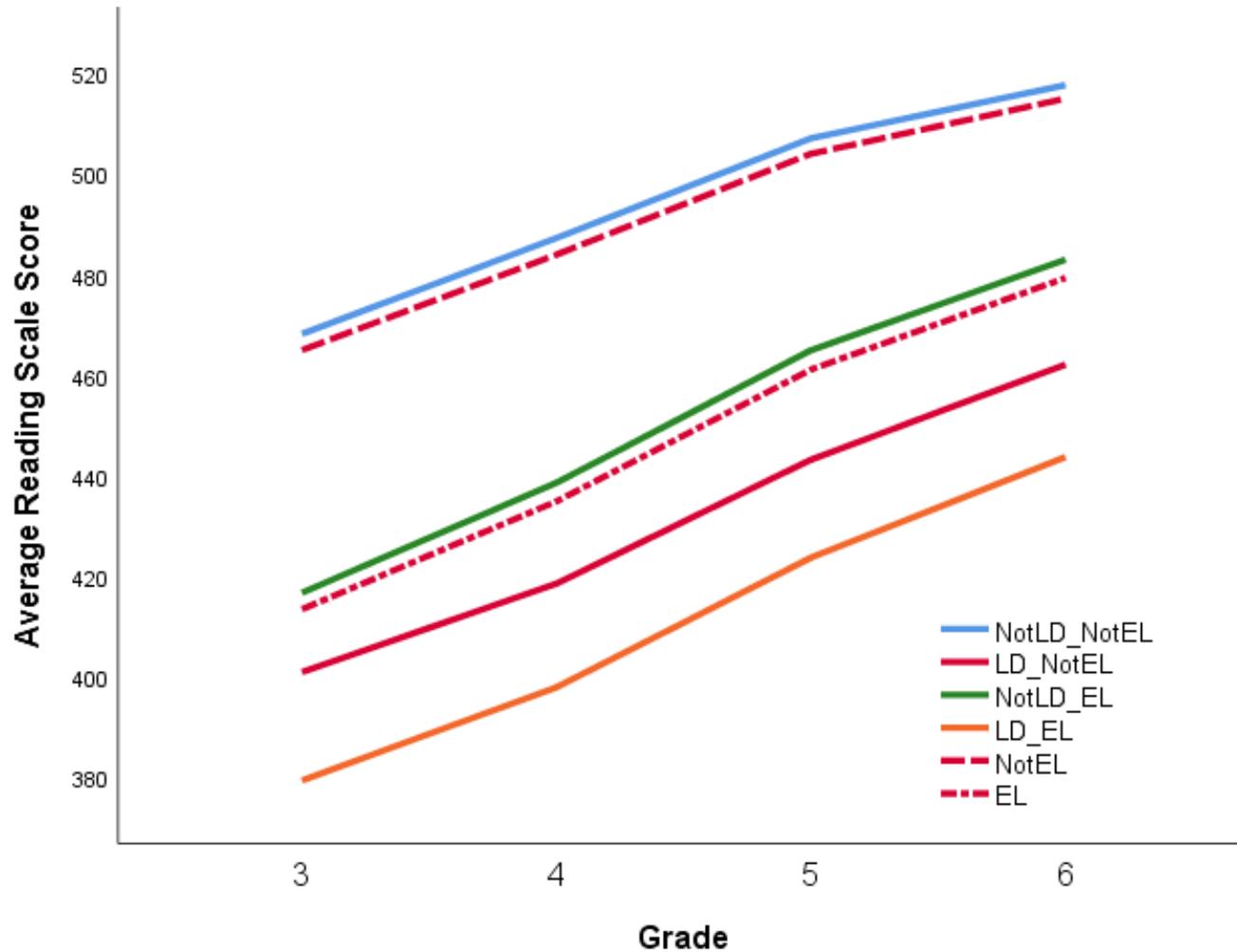
Results

- In all the analyses we conducted, the interaction effect was statistically significant.

- There are many follow-up analyses of interest, but here we only present some examples of achievement gaps expressed as Effect Sizes (ES).

- We have produced a series of brief, one-page summaries of our NCAASE research results called "Did You Know"; several describe the use of ES information to report achievement gaps (see DYK's 1, 2, 4, 13).

- There is also a research brief on ES: http://www.ncaase.com/publications/view?id=138

NCAASE National Center on Assessment and Accountability for Special Education
Advancing research on growth measures, models, and policies for improved practice

http://www.ncaase.com/

14

Results: Achievement Gap ES

- Here are a few of the findings illustrated graphically above with group differences now expressed as ES. Cohen's rules of thumb for interpreting ES are:
  - zero is equivalent to no difference between groups,
  - about 0.20 is considered "small,"
  - about 0.50 is "medium,"
  - 0.80 or more is "large."

NCAASE National Center on Assessment and Accountability for Special Education

Advancing research on growth measures, models, and policies for improved practice

http://www.ncaase.com/

Results: Achievement Gap ES

- The achievement gap for the "singular" ES were:
    - For students who are LD versus not LD (slide 9), -1.30 in Grade 4 and -1.29 in Grade 6; no appreciable change in the ES achievement gap.
    - For students who are EL versus not EL (slide 10), -0.97 in Grade 4 and -0.83 in Grade 6, a narrowing of the achievement gap.

- For the <u>interaction effects</u> of LD and EL, students who were not LD and not EL were the comparison group and achievement gaps were:

	Grade 3	Grade 6
LD-NotEL	-1.32	-1.30
NotLD-EL	-1.01	-0.81
LD-EL	-1.75	-1.73

http://www.ncaase.com/
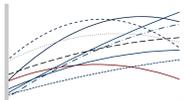
Conclusion

- Importance of investigating achievement gaps more carefully than usual methods (i.e., description of differences in percent proficient):

  - Longitudinal not cross-sectional so change in gaps can be evaluated.

  - Use objective measures of achievement gap size (e.g., ES) rather than "eyeballing" differences in percent proficient.

  - Test true interaction effects to correctly evaluate combinations of student characteristics and to further disaggregate results.

- Our results demonstrate that using these methods, previously unexamined student subgroups may emerge with substantially larger achievement gaps.

- These analyses can increase attention to the need for intervention for certain student subgroups who may be at greater risk academically.

Thank You

Contact information:

Joseph Stevens, Ph.D.

College of Education

5267 University of Oregon

Eugene, OR 97403

(541) 346-2445

stevensj@uoregon.edu

Presentation available on NCAASE web site: http://www.ncaase.com/

NCAASE
National Center on Assessment and
Accountability for Special Education
Advancing research on growth measures, models, and policies for improved practice

http://www.ncaase.com/