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

Joseph J. Stevens

University of Oregon

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., precise disaggregation of groups.

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

- Because of our short time, we only present some highlights:

  - Reading achievement over time for LD vs. not-LD and for EL vs. not-EL (what you get with usual regression models).

  - Reading achievement over time for the LD-EL underline interaction effect.

  - Differences in LD-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

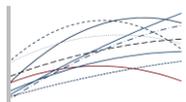# Interactions of Student Characteristics

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

- Usually these variables are examined as one of several predictors in a regression model (i.e., a partial regression effect).

- Even though two predictors (LD-sex) are in the same regression model, they do not estimate an 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. (2017). The interaction of learning disability status and student demographic characteristics on mathematics growth. *Journal of Learning Disabilities*. DOI: 10.1177/0022219415618496
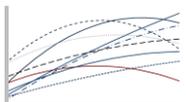
Interactions of Disability Status and Student Characteristics

- 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/

19

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

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

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

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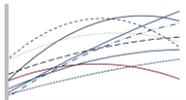
Method

- 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; 5.6% LD; 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 author.
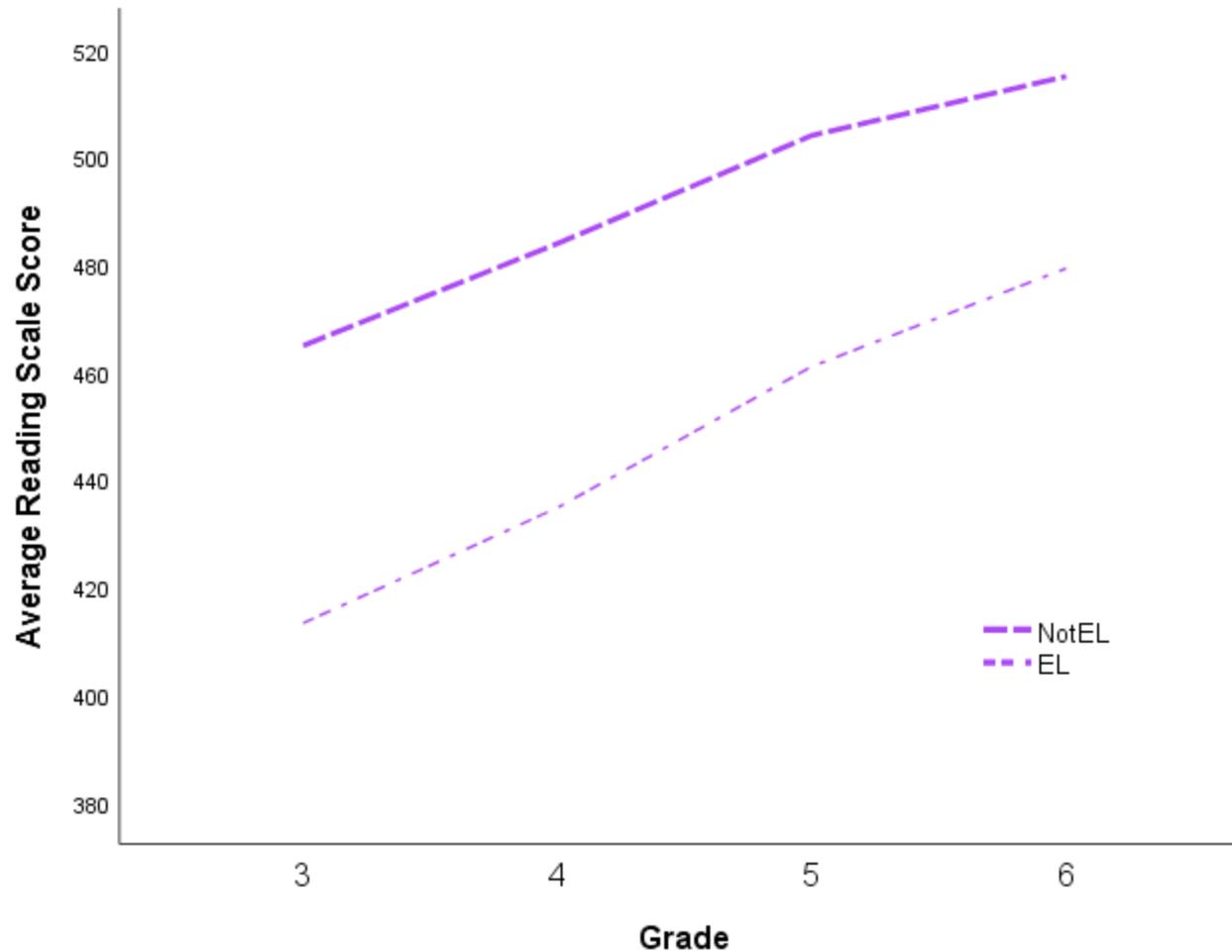
Results

- In all the analyses we conducted, the interaction effects were 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 summaries of the size of achievement gaps expressed as effect sizes (ES).
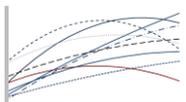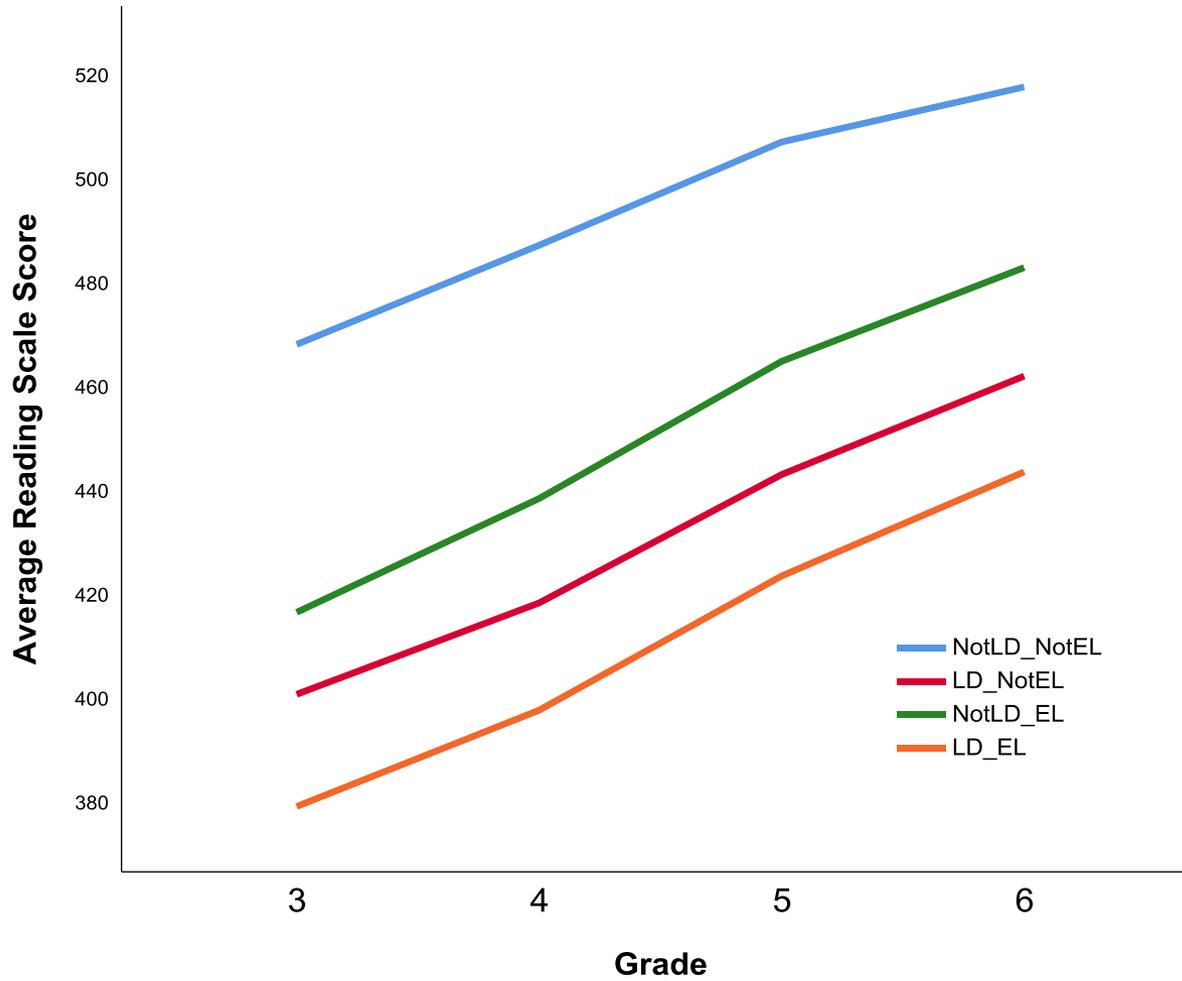
Single Predictor Comparison of LD Status
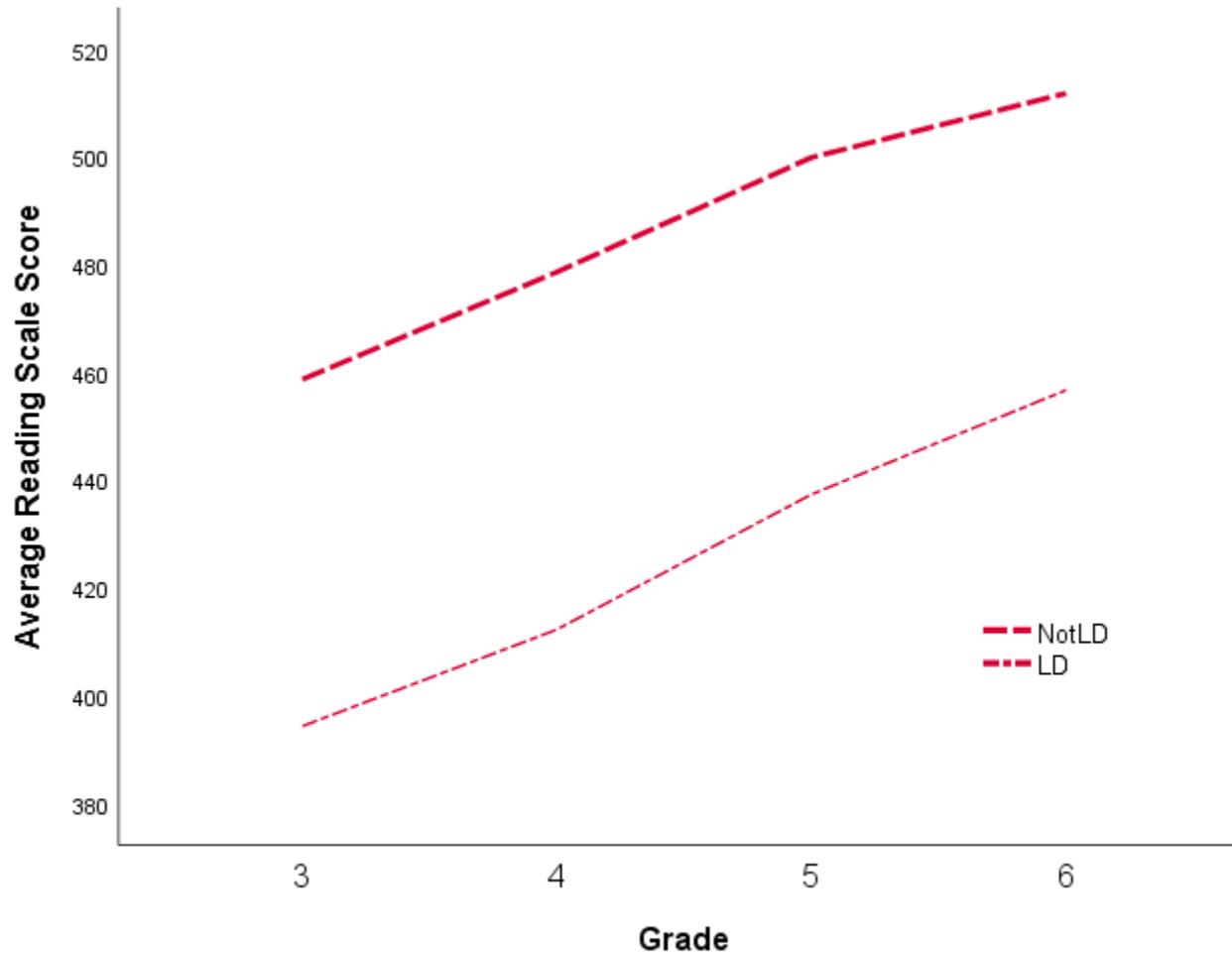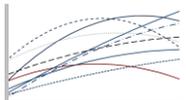
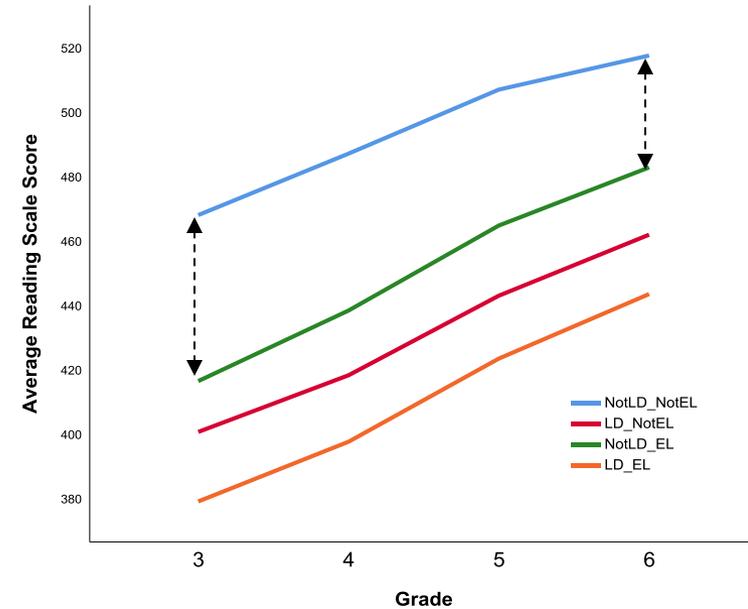# Single Predictor Comparison of EL status

Interaction of LD Status with EL Status

Results

- Note there is generally parallel growth over grades.

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

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

- <u>Bad news</u>: on the whole the gap is not closing.

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

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/

26

Interaction Subgroups (solid lines) vs.
Partial Regression Effects **LD** Subgroups (dashed lines)
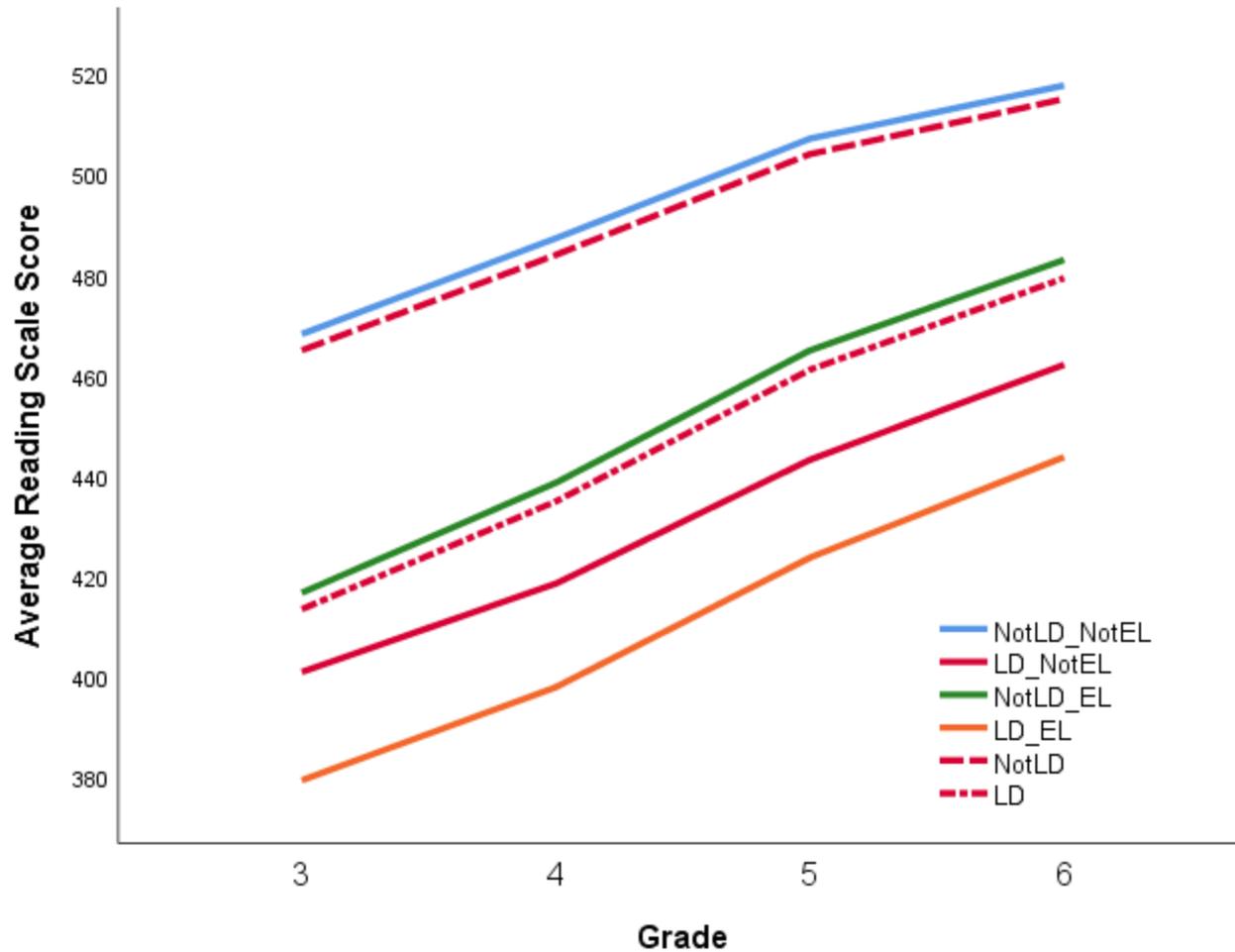
Interaction Subgroups (solid lines) vs. Partial Regression Effects **EL** Subgroups (dashed lines)
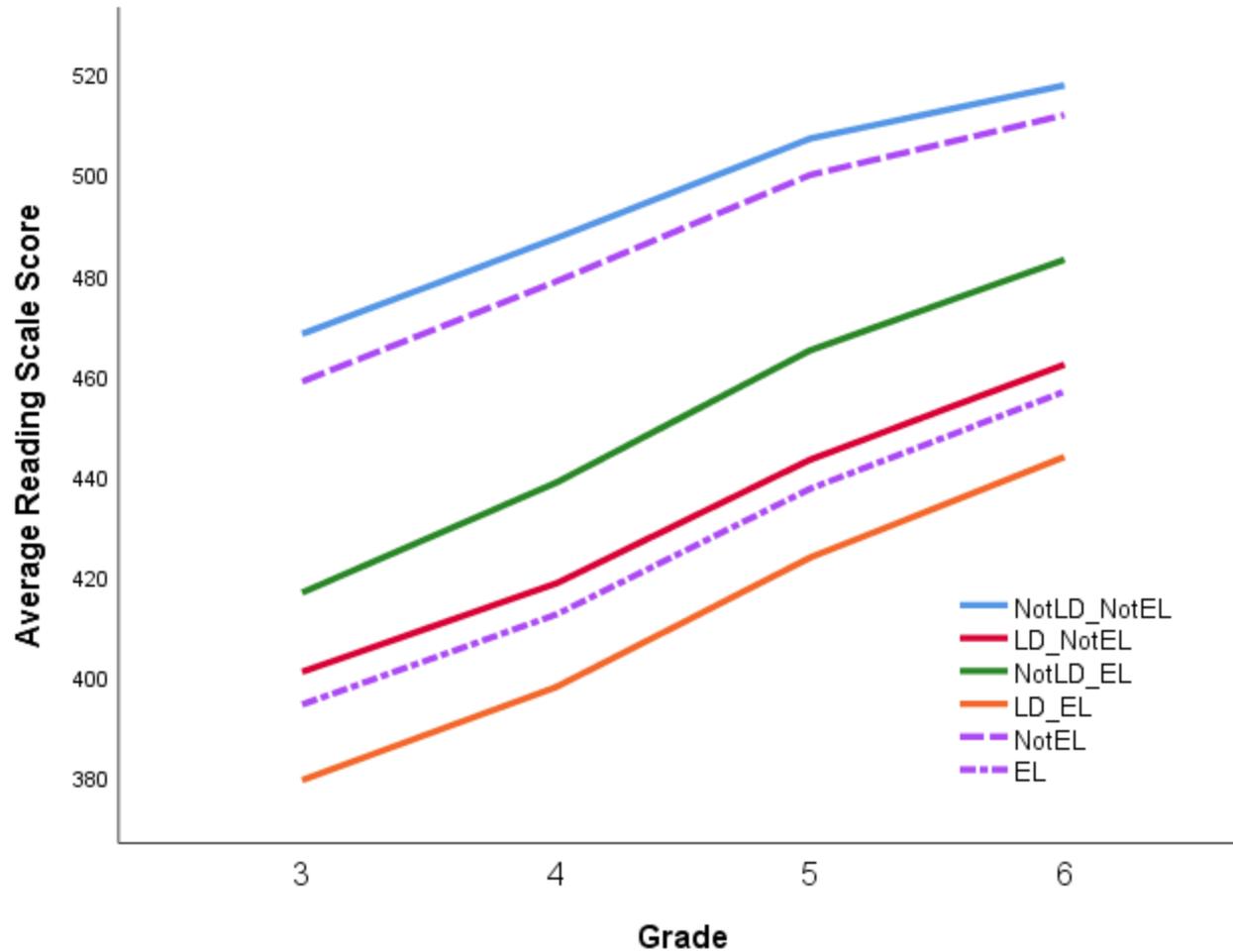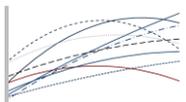
Results

- 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/

29

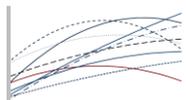# Results: Achievement Gap ES

- Using the findings illustrated graphically above, we now examine the size of achievement gaps using estimates of ES.

- Cohen's rules of thumb for interpreting ES are:
  - zero is equivalent to no difference between groups,
  - about 0.20 is considered a "small" effect,
  - about 0.50 is a "medium" effect,
  - 0.80 or higher is a "large" effect.

http://www.ncaase.com/

Results: Achievement Gap ES

- The ES for the partial regression comparisons 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 for the remaining subgroups were:

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

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/

Conclusion

- Importance of investigating achievement gaps more carefully than usual methods (i.e., description of differences in percent proficient):
  - Longitudinal not cross-sectional, so <u>change</u> 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.

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/

32

Conclusion

- These results exactly parallel Stevens & Schulte (2017).
    - Replication with a different state sample of students and a different state testing system (AZ vs. NC).
    - Extension of previous analyses to an important student subgroup (EL students).

- Through the use of interaction effects, important student subgroups are more clearly identified and evaluated.

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

http://www.ncaase.com/