A Constellation of Findings
From NCAASE

Drs. Stevens and Schulte on Across Year Growth and Gaps
Drs. Elliott, Tindal, and Nese on Within Year Growth
Dr. Stevens on Comparison of Growth Models
Six Questions turned in Numerous Studies

1. What is the **natural developmental progress in achievement** for students with disabilities?

2. What **models best characterize achievement growth** for students with disabilities who are participating in general achievement tests, as well as those taking alternate assessments?

3. How do various **growth models represent school effects** for students with and without disabilities, and how do results compare to those derived from status models now in use?

4. What are the **reliability and validity of the estimates of school effectiveness** for students with disabilities produced by alternative growth models and how are these estimates influenced by contextual differences among schools and students?

5. How do **results from different types of interim assessments of students’ achievement** meaningfully contribute to a model of academic growth for students with disabilities?

6. How can **information about opportunity to learn and achievement growth** be used to enhance academic outcomes for students with disabilities?
Growth in Alternate Assessments with HLM and Transition Analyses

Across Year Growth (and Gaps) for Students with and without Disabilities

Within Year Growth to Monitor Growth and Gaps

Dissemination

Alternate Assessments and Growth for SWSCD

OTL to increase Growth and Close GAPS

Participation in Testing Programs

Multiple Testing Opportunities

Presentation

In Briefs

Publications

Students in and Out of Testing Programs

Students in and Out of Testing Programs

Within Year Growth to Monitor Growth and Gaps

Within Year Growth to Monitor Growth and Gaps

NCAASE National Center on Assessment and Accountability for Special Education

Advancing research on growth measures, models, and policies for improved practice
Achievement Growth for Students With and Without Disabilities

Ann Schulte
Arizona State University
Investigating Achievement Growth for SWD

- Widespread agreement that achievement growth is an important outcome to be considered in educational accountability models.

- Little information on achievement growth for SWD or how it compares to students without disabilities (SWOD) on which to base alternative models.
Investigating Achievement Growth for SWD

- NCLB treats SWD as a unitary subgroup, but specific disabilities are likely to have different achievement trajectories—how different are the exceptionality groups in terms of intercept, growth, and gaps?
- What are some implications for charting growth of the changing nature of the special education population?


Followed two entire cohorts (n > 95,000 each) in one state across grades 3-7, single editions of state reading and mathematics tests

Exceptionality based on classification at Grade 3 (students who entered system after Grade 3 not included in sample)

Two level HLMs (time and students), controlling for demographic characteristics at Level 2

Intercept at Grade 3, both linear and quadratic functions included
North Carolina End of Grade (EOG) Tests

- EOG-R (2\textsuperscript{nd} ed.) Reading passages followed by multiple-choice items testing comprehension of passage content
- EOG-M (2\textsuperscript{nd} ed.) Mathematics stressed application and problem solving
- Administered annually, grades 3-8
- Developmental scale across grades
Reading Comprehension Growth by Exceptionality

Academically gifted, reading
Academically gifted, other
General education
Speech-language impairment
Intellectual disability
Learning Disability in Reading vs. LD in Other Area

**Reading Developmental Scale Score**

- General education
- Learning disability, other
- Learning disability, reading
Reading Comprehension Growth by Exceptionality

- Academically gifted, reading
- Academically gifted, other
- General education
- Speech-language impairment
- Learning disability, other
- Autism
- Other health impairment
- Emotional disturbance
- Hearing impairment
- Learning disability, reading
- Intellectual disability

Grade

Reading Developmental Scale Score
### Reading Comprehension Two-level HLM (Time and Students)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Intercept</th>
<th>Linear</th>
<th>Quadratic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grand Mean</td>
<td>250.82 (.04)</td>
<td>5.03 (.03)</td>
<td>-.42 (.01)</td>
</tr>
<tr>
<td>Sex</td>
<td>1.02(.05)</td>
<td>-0.17 (.03)</td>
<td>.07 (.01)</td>
</tr>
<tr>
<td>Free lunch</td>
<td>-3.61 (.06)</td>
<td>0.14 (.03)</td>
<td>-.05 (.01)</td>
</tr>
<tr>
<td>Limited English</td>
<td>-5.57(.16)</td>
<td>0.70 (.10)</td>
<td>-.05 (.02)</td>
</tr>
<tr>
<td>Asian</td>
<td>.31† (.17)</td>
<td>0.18† (.10)</td>
<td>.05 (.02)</td>
</tr>
<tr>
<td>American Indian</td>
<td>-2.74 (.21)</td>
<td>-0.62 (.13)</td>
<td>.12 (.03)</td>
</tr>
<tr>
<td>Black</td>
<td>-4.34 (.06)</td>
<td>0.25 (.04)</td>
<td>-.05 (.01)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-1.54 (.13)</td>
<td>0.40 (.08)</td>
<td>-.05 (.02)</td>
</tr>
<tr>
<td>Multiracial</td>
<td>-1.05 (.16)</td>
<td>0.19† (.10)</td>
<td>-.03† (.02)</td>
</tr>
</tbody>
</table>

†Not significant, \( p > .05 \)
## Reading Two-level HLM (con’t.)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Intercept</th>
<th>Linear</th>
<th>Quadratic</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gifted-Rdg</td>
<td>7.82 (.08)</td>
<td>-0.59 (.06)</td>
<td>.09 (.01)</td>
</tr>
<tr>
<td>Gifted-Other</td>
<td>6.84 (0.15)</td>
<td>-0.54 (.12)</td>
<td>.08 (.03)</td>
</tr>
<tr>
<td>Autism</td>
<td>-7.80 (.74)</td>
<td>.38† (.41)</td>
<td>-.01† (.10)</td>
</tr>
<tr>
<td>Emotional Disturbance</td>
<td>-7.17 (.35)</td>
<td>1.00 (.23)</td>
<td>-.25 (.06)</td>
</tr>
<tr>
<td>Hearing Impairment</td>
<td>-8.89 (.65)</td>
<td>0.77† (.40)</td>
<td>-.15† (.10)</td>
</tr>
<tr>
<td>Intellectual Disability</td>
<td>-14.82 (.22)</td>
<td>1.38 (.18)</td>
<td>-.25 (.04)</td>
</tr>
<tr>
<td>Other Health Impairment</td>
<td>-7.86 (.22)</td>
<td>0.62 (.13)</td>
<td>-.13 (.03)</td>
</tr>
<tr>
<td>Specific Learning Disability-Rdg</td>
<td>-9.78 (.13)</td>
<td>1.72 (.08)</td>
<td>-.27 (.02)</td>
</tr>
<tr>
<td>Specific Learning Disability-Other</td>
<td>-4.96 (.25)</td>
<td>0.85 (.15)</td>
<td>-.18 (.03)</td>
</tr>
<tr>
<td>Speech-language Impairment</td>
<td>-2.64 (.15)</td>
<td>0.38 (.09)</td>
<td>-.06 (.02)</td>
</tr>
<tr>
<td>Pseudo R² (as %)</td>
<td>39.44</td>
<td>8.45</td>
<td>8.26</td>
</tr>
</tbody>
</table>

†Not significant, p > .05
Special Education Subgroup
Membership Changes Across Time

- Students exit and enter special education and changes are related to student achievement status.
- What are the implications of the changing membership?

# Special Education Membership

**Grades 3-7**

<table>
<thead>
<tr>
<th>SWD Subgroup Identification Method</th>
<th>Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Current Year</td>
<td>11.1 to 12.4</td>
</tr>
<tr>
<td>Wave 1</td>
<td>11.8</td>
</tr>
<tr>
<td>Ever in Special Education</td>
<td>16.1</td>
</tr>
<tr>
<td>Always in Special Education</td>
<td>6.0</td>
</tr>
</tbody>
</table>
Observed Means by SWD Identification Method

![Graph showing Mathematics Scale Score by Grade for Wave 1 NonSWD and Wave 1 SWD, with lines indicating Always in Special Education Non-SWD and Always in Special Education SWD.]
Implications

- SWD subgroup is comprised of heterogeneous group of students, who vary greatly in achievement in grade 3.

- Most exceptionality groups made somewhat greater growth than general education students in reading and somewhat less growth in mathematics—although overall characterization is one of stable gaps.

- “One size may fit all” for growth, but only if differing starting points for SWD are recognized. Growth-to-standard expectations require much greater growth for most SWD groups than is typically observed.
Individual Differences and Gaps for Students With Disabilities

Joseph Stevens
University Oregon
Individual Differences and Achievement Gaps in Math and Reading for SWD

- Summarize a number of our study results that focus on individual differences in academic performance.
- Central NCLB and RTTT goal is universal proficiency and the reduction of achievement gaps between SWoD students and protected subgroups including SWD.
- Previous research on achievement gaps has limitations:
  - Often gaps are not evaluated empirically; visual inspection rather than statistical testing; no common, empirical metric (effect size) to describe differences.
  - Interactions not tested.
SWD Growth Achievement Gaps

- What is the size of the achievement gap in mathematics and reading for students in specific exceptionality categories?
- Does the gap increase, decrease or stay the same over time?

Mathematics Achievement Gaps, see:

Reading Achievement Gaps, see:
Figure. Reading achievement gap effect sizes based on differences in empirical Bayes estimated means across grades for students in different exceptionality categories compared to students in general education (from Schulte et al., in press).
Achievement Gaps as Differences in Proficiency Rates

- We also examined SWoD-SWD achievement gaps in other ways

- Difference in percent proficient \((P-P)\)
  
  - Most common method in public dissemination (e.g., report cards); district, state, and federal reports
  
  - Easy to interpret?
  
  - \(P-P\) and Cohen’s \(h\) reported below for North Carolina

- Problems with these metrics, however:
  
  - Size of gap depends on test used, score scale, and location of cutscore
  
  - Size of gap depends on shape of score distributions for the two groups
  
  - Proportions are ordinal, units may be different at different locations on the scale (i.e., not an interval scale)
<table>
<thead>
<tr>
<th>North Carolina</th>
<th>Mathematics</th>
<th>Reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grade</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>Student Group</td>
<td></td>
<td></td>
</tr>
<tr>
<td>General Education</td>
<td>78.3</td>
<td>79.3</td>
</tr>
<tr>
<td></td>
<td>(97680)</td>
<td>(94162)</td>
</tr>
<tr>
<td>All SWD</td>
<td>53.2</td>
<td>49.9</td>
</tr>
<tr>
<td></td>
<td>(11208)</td>
<td>(11046)</td>
</tr>
<tr>
<td></td>
<td>.54</td>
<td>.63</td>
</tr>
<tr>
<td>Autism</td>
<td>62.4</td>
<td>64.9</td>
</tr>
<tr>
<td></td>
<td>(351)</td>
<td>(365)</td>
</tr>
<tr>
<td></td>
<td>.35</td>
<td>.32</td>
</tr>
<tr>
<td>Communication</td>
<td>67.3</td>
<td>66.3</td>
</tr>
<tr>
<td></td>
<td>(3842)</td>
<td>(2501)</td>
</tr>
<tr>
<td></td>
<td>.25</td>
<td>.29</td>
</tr>
<tr>
<td>Emotional</td>
<td>44.4</td>
<td>43.9</td>
</tr>
<tr>
<td></td>
<td>(331)</td>
<td>(394)</td>
</tr>
<tr>
<td></td>
<td>.71</td>
<td>.75</td>
</tr>
<tr>
<td>Hearing</td>
<td>49.0</td>
<td>50.0</td>
</tr>
<tr>
<td></td>
<td>(143)</td>
<td>(130)</td>
</tr>
<tr>
<td></td>
<td>.62</td>
<td>.63</td>
</tr>
<tr>
<td>Grade</td>
<td>Mathematics</td>
<td>Reading</td>
</tr>
<tr>
<td>-------------</td>
<td>-------------</td>
<td>---------</td>
</tr>
<tr>
<td></td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>General Education</td>
<td>78.3 (97680)</td>
<td>79.3 (94162)</td>
</tr>
<tr>
<td>Intellectual</td>
<td>7.9 (252)</td>
<td>4.8 (229)</td>
</tr>
<tr>
<td>Orthopedic</td>
<td>55.6 (36)</td>
<td>40.5 (37)</td>
</tr>
<tr>
<td>Other</td>
<td>44.1 (1663)</td>
<td>41.7 (2085)</td>
</tr>
<tr>
<td>Language Disability</td>
<td>47.1 (4524)</td>
<td>46.4 (5246)</td>
</tr>
<tr>
<td>TBI</td>
<td>30.0 (10)</td>
<td>- (8)</td>
</tr>
<tr>
<td>Visual Impairment</td>
<td>62.8 (43)</td>
<td>75.0 (48)</td>
</tr>
<tr>
<td></td>
<td>.34 .10</td>
<td>.43 .29</td>
</tr>
</tbody>
</table>
Achievement Gaps as Areas Between Score Distributions

- A limitation of traditional ES measures is they only compare groups at the mean or at the proficiency cutpoint, possibly overlooking important group differences lower or higher on the score scale.

- Alternatives are effect size measures based on nonparametric methods that examine group differences for all proficiency categories (see Ho & Reardon, 2012):
  - Area under the curve in Receiver Operating Curve (ROC) analysis
  - $V'$ statistic

- Because of time constraints, we only report a few examples of ROC analysis.
Whole distribution Achievement Gaps

- ROC analysis (and $V'$) uses nonparametric methods to address problems associated with characteristics of score distributions.
- Advantage is estimation of gap across all proficiency levels.
- ROC curve diagonal line represents no difference between reference group (SWoD) and focal group (SWD).
- Size of area between SWD group curve and diagonal is the area under the curve or the size of the difference between the two groups.
- In following examples, note differences: 
  - at different proficiency levels
  - for math vs. reading
  - by exceptionality subgroup
Achievement Gap for SWD vs. SWoD in Oregon Reading in Grade 3 (on left) and Grade 5 (on right)
Achievement Gap for SWD vs. SWoD in NC Math and Reading Grades 3-5
Achievement Gap for SWoD vs. Speech-language Impairment (on left) or Mild Intellectual Disability (on right) on NC Math and Reading Grades 3-5
Interactions of SWD status and Other Student Characteristics

- Many studies do not directly test the interaction of SWD status and factors thought to be related to student performance (e.g., LD status and sex of student)
- When these factors are included in statistical models (especially regression and HLM models), only partial regression effects not the actual interactions are analyzed
- This can be very misleading and result in incorrect interpretations as well as incomplete understanding of achievement gaps

Figure. Partial regression of FRL compared to the reference group on left; three way interaction effect of LD x FRL x grade interaction on right.
Figure. Partial regression of LD compared to the reference group on left; three way interaction effect of LD x Black race/ethnicity x grade interaction on right.
Mathematics Achievement Gaps for Elementary and Secondary Students:
The Influence of Opportunity to Learn and Special Education Status

Stephen N. Elliott
Arizona State University
Research Questions

Specific research questions motivating the study were:

1. Do students with and without disabilities who received instruction in the same general education classrooms have an equal opportunity to learn mathematics?

2. What is the relationship among five instructional variables (characterized as OTL) and within year academic growth on an interim assessments?

3. What is the predictive relationship among five instructional OTL variables and students’ end-of-year mathematics achievement?
Opportunity to Learn (OTL) the Intended Curriculum

Definition: Opportunity to Learn

The degree to which a teacher dedicates instructional time and content coverage to the intended curriculum objectives emphasizing higher-order cognitive processes, evidence-based instructional practices, and alternative grouping formats.

(Kurz, 2011)

A unified conceptualization of OTL based on 50+ years of empirical research.
Multiple Measures Study Design*

Teachers (N = 78; AZ 49, OR 29) and students (N = 327; 162 SWD + 165 SWoD) from AZ & OR schools grades 4th-8th.

* A 3-year study with longitudinal student cohorts.
Summary of Year 1: Key Findings

- We observed very similar instructional processes for students with and without disabilities learning mathematics in the same elementary or secondary classrooms in AZ and OR schools. Significant achievement gaps between these groups of students, however, existed on the four interim CBM assessments and the end-of-year achievement state test.

- We found that the collection of five MyiLOGS scores, along with grade level and special education status, accounted for a substantial amount (i.e., 43% to 44%) of the variance in student’s end-of-year mathematics scores. A subset of OTL indices explained a statistically significant, although relatively small portion of unique variance in the end-of-year mathematics scores. The particular OTL scores found to be significant contributors varied across AZ and OR.
Year 2 Findings

- AZ teachers reported an average of 164 days and OR teachers reported 158 days of instruction; 25% of these days were Detail Days where instructional information on cognitive processes, practices, and grouping for SWD and SWOD was documented. Based on these Detail Days, we observed very similar mathematics instructional processes for students with and without disabilities in the same elementary or secondary classrooms in AZ and OR schools. Yet, there were significant achievement gaps between these groups of students on the four interim CBM assessments and the end-of-year achievement state test.

- We also found that Grade Level and Special Education Status, along with the collection of five MyiLOGS scores, accounted for a substantial amount (i.e., 30% OR, 39% AZ) of the variance in student’s end-of-year mathematics scores. OTL indices explained a relatively small portion of unique variance in the end-of-year mathematics scores.

- ICCs (Teacher-Observer) for Observations on 6 random Detail Days each Year: InstrTime = .80; CogProcess = .28; InstrPractice = .39; GroupFormat = .45
Comparison of Interim & End-of-Year Test Results for AZ Students

Arizona Elementary SWOD vs. SWD
Comparison of EasyCBM & State Test

Arizona Secondary SWOD vs. SWD
Comparison of EasyCBM & State Test

SWOD Black  SWD  Gray
Within Year Standardized Mathematics CBM Growth
Conclusion

Offering students with disabilities the same amount of instruction on the same content standards in the same general education classrooms was found to offer the same historic results—large and persistent gaps in achievement -- in comparison to students without disabilities.

The findings in Year 2 replicated those from Year 1. Thus, it indicates that students with disabilities will need more instructional time on the intended curriculum, and perhaps more differentiated instruction to increase their rate of achievement enough to close gaps that currently exist between them and students without disabilities.
Within Year Growth

Drs. Tindal and Nese
University of Oregon
Measurement Sufficiency

- RTI within classrooms using CBMs to screen with tiers of support
- Grades 3-5 with oral reading fluency
- Benchmark and progress monitoring
- Teacher decision making on grade level of measures
- Multi-level model with
  - Time
  - Student characteristics
  - Measurement characteristics

Major Findings
Measurement Sufficiency

Figure 1. Plot of a random sample (.05%) of students’ progress monitoring unconditional growth.
# Measurement Sufficiency

## Table 4. Final Conditional Model With Special Education Status and Measurement Condition.

<table>
<thead>
<tr>
<th>Fixed effect</th>
<th>Coefficient</th>
<th>SE</th>
<th>t ratio</th>
<th>df</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Grade 3</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\beta_{00}$</td>
<td>88.23</td>
<td>1.286</td>
<td>68.61</td>
<td>1278</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Special education, $\beta_{01}$</td>
<td>-19.70</td>
<td>3.041</td>
<td>-6.48</td>
<td>1278</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sufficient measurement, $\beta_{02}$</td>
<td>-21.72</td>
<td>2.23</td>
<td>-9.74</td>
<td>1278</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Slope, $\beta_{10}$</td>
<td>0.67</td>
<td>0.03</td>
<td>20.87</td>
<td>1278</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Special education, $\beta_{11}$</td>
<td>0.08</td>
<td>0.12</td>
<td>0.67</td>
<td>1278</td>
<td>.505</td>
</tr>
<tr>
<td>Sufficient measurement, $\beta_{12}$</td>
<td>0.16</td>
<td>0.05</td>
<td>2.99</td>
<td>1278</td>
<td>.003</td>
</tr>
<tr>
<td><strong>Grade 4</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\beta_{00}$</td>
<td>107.56</td>
<td>1.09</td>
<td>98.69</td>
<td>1235</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Special education, $\beta_{01}$</td>
<td>-23.71</td>
<td>2.70</td>
<td>-8.79</td>
<td>1235</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sufficient measurement, $\beta_{02}$</td>
<td>-19.89</td>
<td>2.14</td>
<td>-9.31</td>
<td>1235</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Slope, $\beta_{10}$</td>
<td>0.62</td>
<td>0.03</td>
<td>24.31</td>
<td>1235</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Special education, $\beta_{11}$</td>
<td>0.01</td>
<td>0.05</td>
<td>0.16</td>
<td>1235</td>
<td>.872</td>
</tr>
<tr>
<td>Sufficient measurement, $\beta_{12}$</td>
<td>0.12</td>
<td>0.04</td>
<td>2.71</td>
<td>1235</td>
<td>.007</td>
</tr>
<tr>
<td><strong>Grade 5</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Intercept, $\beta_{00}$</td>
<td>133.24</td>
<td>1.30</td>
<td>102.37</td>
<td>1094</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Special education, $\beta_{01}$</td>
<td>-28.42</td>
<td>2.70</td>
<td>-10.53</td>
<td>1094</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Sufficient measurement, $\beta_{02}$</td>
<td>-18.27</td>
<td>2.53</td>
<td>-7.22</td>
<td>1094</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Slope, $\beta_{10}$</td>
<td>0.70</td>
<td>0.03</td>
<td>21.93</td>
<td>1094</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Special education, $\beta_{11}$</td>
<td>-0.08</td>
<td>0.05</td>
<td>-1.53</td>
<td>1094</td>
<td>.126</td>
</tr>
<tr>
<td>Sufficient measurement, $\beta_{12}$</td>
<td>-0.10</td>
<td>0.05</td>
<td>-2.03</td>
<td>1094</td>
<td>.043</td>
</tr>
</tbody>
</table>
30 Years of Research on ORF

- Initial findings in 1984 – Pine County Norms
- Fuchs and Deno estimates of growth
- Deno and Shin estimation with a 'national sample'
- Recent studies with large data sets
  - Hasbrouck and Tindal publications
  - Florida studies
  - easyCBM
### Table 1

**Comparison of Studies on Research Variables Referenced in Studying Growth of Reading Fluency**

<table>
<thead>
<tr>
<th>Authors (date)</th>
<th>Grades (n):</th>
<th>Students</th>
<th>Measures</th>
<th>N-Measures</th>
<th>Slope Calculation</th>
<th>Averages and Growth – WCPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Marston, Lowry, Demo, &amp; Mirkin (1981)</td>
<td>1 (13), 2 (9), 3 (10), 4 (7), 5 (7), 6 (9)</td>
<td>Students from a small Midwest city</td>
<td>Third grade basal reading series: Allyn-Bacon, Ginn 720, Houghton Mifflin</td>
<td>Three administrations: Fall, winter, and spring</td>
<td>Plot of raw scores and percentage of change; significance tests of differences</td>
<td>Grades 18.1 31.1 45.7 Fall 73.2 101.1 127.8 Winter 108.3 123.6 136.2 Spring 125.4 131.7 155.3</td>
</tr>
<tr>
<td>Tindal, Germann, Marston, &amp; Demo (1983)</td>
<td>1 (5), 2 (13), 3 (17), 4 (22), 5 (18), 6 (21)</td>
<td>Students from six districts referred, assessed, and eligible for special education district in use in the school</td>
<td>Three administrations: Fall, winter, and spring</td>
<td>Change in raw score and in discrepancy from general education</td>
<td>Grades 6.8 14.3 9.6 Fall 5.5 16.2 23.6 Winter 20.5 36.6 41.3 Spring 31.0 50.3 52.9</td>
<td></td>
</tr>
<tr>
<td>Tindal, Germann, &amp; Demo (1983)</td>
<td>1 (276), 2 (284), 3 (302), 4 (294), 5 (315), 6 (328)</td>
<td>Students randomly sampled from six districts in Pine County</td>
<td>Two passages sampled from basal reading curriculum</td>
<td>Three administrations: Fall, winter, and spring</td>
<td>Change in raw scores</td>
<td>Grades 5 63 75 Fall 35 83 93 Winter 67 89 108 Spring 98 111 128</td>
</tr>
</tbody>
</table>
| Fuchs, Demo, & Mirkin (1984) | All students were ‘handicapped’ | 3rd grade passage reading test from Ginn 720 | Pre-post (unknown: sometime between Nov. and May) | Pre-post difference @ 28 weeks | Condition* Pre Post Experimental* 41.6 70.2 Contrast 51.5 51.3

*DBPM versus none*
### 30 Years of Research on ORF

<table>
<thead>
<tr>
<th>Authors (date)</th>
<th>Grades</th>
<th>Students</th>
<th>Measures</th>
<th>N-Measures</th>
<th>Slope Calculation</th>
<th>Averages and Growth – WCPM</th>
</tr>
</thead>
<tbody>
<tr>
<td>Al Otaiba, Petscher, Pappamichiel, Williams, Dyrlund, &amp; Connor (2009)</td>
<td>Grades 2-3: 5,004 Latino students</td>
<td>(a) Proficient in English, (b) Not proficient and receiving English as a second language (ESL) services, and (c) Proficient enough to have exited from ESL</td>
<td>DIBELS (Good &amp; Kaminski, 1996)</td>
<td>Four administrations: first 20–30 days of school (Sep); between the 65th and 75th days of school (Nov); between the 110th and 120th days of school (Feb); and between the 155th and 165th days of school (Apr).</td>
<td>Two level HLM with growth centered on the first testing time in third grade (September) and student characteristics (i.e., language group and special education subgroup) were entered at Level 2 designed to model both second- and third-grade growth trajectories</td>
<td>Grade 2 Weekly Growth:</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Proficient ESL ESL-exit</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GE</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LD</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SL</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Grade 3 Weekly Growth:</td>
<td>Proficient ESL ESL-exit</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>GE</td>
<td>1.2</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>LD</td>
<td>1.3</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>SL</td>
<td>1.2</td>
</tr>
</tbody>
</table>

| Ardoin & Christ (2009) | Grades 2-3: 28 and 40 respectively | Race and free or reduced-price lunch reported for students in each of two schools | Three passage sets: (a) FAIP-R, (b) AIMSweb, and (c) DIBELS | 12 weeks (with approximately 2 administrations per week) with time of year unknown | Ordinary least square (OLS) regression used to calculate an intercept and slope for each passage set | Intercept Slope* |
| | | | | | FAIP-R | 84.8 | 1.4 |
| | | | | | AIMSweb | 94.3 | 1.8 |
| | | | | | DIBELS | 100.4 | .5 |
| *Slope times 7 to estimate weekly gain |

| Crowe, Connor, & Petscher (2009) | Grade 1 (n=9,993), Grade 2 (n=9,869), Grade 3 (n=10,141) | Lower SES (eligible for free or reduced price lunch) and non lower SES students | DIBELS (Good & Kaminski, 2002) | Two administrations in September and April | Hierarchical Linear Modeling (HLM) to estimate mean growth trajectories for curriculum interacting with SES over the school year (7 months) | Sept. April |
| | | | | | Grade 1 | 18 | 50 |
| | | | | | • Higher SES | 25 | 63 |
| | | | | | • Lower SES | 16 | 46 |
| | | | | | Grade 2 | 53 | 89 |
| | | | | | • Higher SES | 63 | 77 |
| | | | | | • Lower SES | 50 | 62 |
| | | | | | Grade 3 | 73 | 101 |
| | | | | | • Higher SES | 85 | 113 |
| | | | | | • Lower SES | 69 | 97 |
| *Curriculum differences reported that interacted with students SES. |
PRF WITHIN all Grades

- Rates of growth in research and aim lines in practice are used to characterize student growth; in either case, growth is generally defined as linear, increasing at a constant rate over time.
- Linearity assumption may be inaccurate.
- We examined ORF growth within-year for students in Grades 1-8.
  - Other research limited by using only 3 testing occasions.
  - Our sample included Grade 1 to 8 students, drawn from the full range of abilities within each grade level and assessed up to 8 times per year.
Comparing the trajectories across grades, we found that a decelerating growth curve best described ORF data.

On average, across grades, students exhibit a decrease in growth across the year.

PRF within GRADE 4

- The purpose of this study was to demonstrate ways to model nonlinear growth using three testing occasions: fall, winter, and spring passage reading fluency benchmark assessments.
- 2,100 Grade 4 students.
- Passage reading fluency (PRF)
Figure 3. Estimated growth patterns for 1-class PGM and 2-class and 3-class piecewise growth mixture models (PGMM).
Purpose

Introduce and apply a two-step growth mixture model (GMM) approach for modeling repeated measures with distributions changing over time.
Results

Letter Sound Fluency (spcm)

Kindergarten

Zero Class, Class 1
Zero & Above Class, Class 1
Zero Class, Class 2
Zero & Above Class, Class 2
Zero Class, Class 3
Zero & Above Class, Class 3
Growth Models for School Effects

Joseph Stevens
University Oregon
Models Evaluating School Level Growth

- Comparison of different models of estimating school performance using OR, AZ, NC, and PA state data
- Models include:
  - status; gain and residual scores
  - transition matrix models
  - value-added models
  - Student Growth Percentiles (SGP)
  - Hierarchical linear growth models (HLM)
- Model variations include: (a) three cohorts; (b) two grade levels, elementary (Grades 3-5) and middle (Grades 6-8), and (c) unconditional vs. conditional models (school size, student composition of school)
Some Preliminary Results

- Estimates of school effects across models sometimes very consistent, but other times very discrepant:
  - Most transition matrix models moderately to highly correlated with each other
  - SGPs almost perfectly correlated with conditional regression
  - Low to moderate correlation of SGPs with HLM growth
  - Low correlation of status models with HLM Growth

- Substantial cohort instability in the first state studied (Oregon)

- Several examples follow
Student Growth Percentiles (SGP)

- Described as a Relative Growth Model
  - Current year performance conditioned on prior year(s) of performance using quantile regression
  - Relative rank in a distribution of those who had similar scores in previous years

- Oregon sample composed of all those who had a math or reading-language score in 2011 and at least one prior year score in years 2008-2010

- Currently popular approach in state accountability models used in dozens of states

- Result illustrated here is cohort instability using SGPs
Cohort Differences in SGP Models
Hierarchical Linear Models

- Another alternative representation of student growth rests on the statistical modeling of change over time.
- These models are absolute growth models in that they relate change to a time function and maintain the metric of the score scale.
- Therefore a vertically linked score scale is necessary.
- HLM models produce two primary outcomes, intercept (very similar to other status measures like school mean or PP) and growth estimate (slope).
- HLM results differ substantially from some other models especially status models (e.g., percent proficient).
- Illustration of cohort instability with HLM follows.
Cohort Differences in HLM Growth Models
Covariate Relations with School Estimates

R² Linear = 0.347
Relations of Model Estimates to School Proportion SWD

- School proportion SWD very positively skewed; variable was categorized for comparisons: Group 1, schools with 10% SWD or less; Group 2, 11-14% SWD; and Group 3: more than 14% SWD
- Significant differences by School SWD group for PP and HLM intercept; not statistically significant for SGP, HLM Slope
Relation of School Percent Proficient with School Proportion SWD ($p = .018$)
Relation of School SGP with School Proportion SWD (ns, $p = .654$)
Relation of HLM EB Intercept with School Proportion SWD ($p < .001$)
Relation of HLM EB Slope with School Proportion SWD (ns, $p = .796$)
Q and A

http://ncaase.com

Joe Stevens – stevensj@uoregon.edu
Ann Schulte – Ann.Schulte@asu.edu
Steve Elliott – Steve_Elliott@asu.edu
Gerald Tindal – geraldt@uoregon.edu
Joe Nese – jnese@uoregon.edu