Coordinated Paper Session, Summary and Overview (748 words excluding references)

*Research and Development on Assessment and Accountability for Special Education*

*Overview and Background*. This coordinated paper session provides information on a federally funded National Research Center and recent research completed by the center. The center is designed to develop and test various approaches for measuring achievement growth of students with and without disabilities. The purpose of the proposed NCME session is to provide an overview and description of the center, follow-up on a 2012 NCME session, and present five papers that focus on key recent research findings.

The new center is now beginning the second year of a 5-year program of research on reading and mathematics achievement growth based primarily on existing sets of state longitudinal achievement data for students from Arizona, North Carolina, Oregon, and Pennsylvania. We use formative, interim, and summative assessments to compare outcomes using different analytic procedures (comparing various multi-level growth models to frequently used models based on status/proficiency, transition matrix, residual gain, and value-added models), as well as studies of correlates of growth, relationships between interim and summative assessments, and use of alternate assessments.

The context for center research is the use of achievement data in grades K-8 to measure student learning progress, evaluate programs, and estimate school effectiveness within state accountability systems. While The No Child Left Behind Act (NCLB, 2001) requires reporting of school-level outcomes as well as disaggregation of achievement for subgroups, it originally precluded the use of growth models. More recently, federal flexibility and the Race To The Top program have encouraged the use of growth models for educational evaluation (Spellings, 2006; U. S. Department of Education, 2011). Nonetheless, we know little about the implications of choosing different methods for evaluating growth or the accountability inferences based on growth models (Author, 2005; Goldschmidt et al., 2012; Heck, 2006; Hess, 2005; Linn & Haug, 2002). Even less is known about growth of targeted subgroups (Kiplinger, 2008), especially for students with disabilities (SWD) whose performance has been a concern for decades (Author, 1998; Carlberg & Kavale, 1980). With many states reporting that over 70% of students with disabilities are below achievement expectations, there is critical need to provide accurate information about the effectiveness of practices with this subgroup.

A number of educational scientists have argued for shifting accountability metrics away from achievement status to achievement growth (e.g., Author, 2004; Betebenner, 2008; Hanushek & Raymond, 2005). Work being completed by the center supports the further development of growth models using a variety of assessment types for several measurement purposes. This work is unusual in that it simultaneously focuses on formative, interim, and summative measures and examines growth models applied for several purposes ranging from monitoring and enhancing student learning to estimation of school effects for accountability. In this session, we will share results of center research from the past year on important issues surrounding the use of longitudinal methods such as: What is the developmental nature of achievement growth in reading and mathematics for students with and without disabilities? What is the course of achievement growth during the school year? What is the functional form of

growth within the school year and across school years? What factors are related to student growth in achievement? How do students with disabilities enter and exit special education over time?

*Session Summary*. The session will include three components: (a) five separate papers presented by the grant's principal investigators and collaborators, (b) reactions from an invited discussant, and (c) structured and open audience participation. The five separate papers are (see details in the individual paper summaries below): (a) *Within-Year Achievement Growth Trajectories Using Progress Monitoring Measures*, (b) *Learning to Read: A synthesis of research on growth in reading skills*, (c) *Special Education Growth: Contrasting Stable and Variable Identification of Special Education Student Status across Grades*, (d) *Reading Achievement Growth at the Student and School Levels for Regular and Special Education Elementary Students*, and (e) *School Effects on the Middle School Reading Achievement of Students with Disabilities: A Multilevel, Longitudinal Analysis*. A national expert in statistics and psychometrics, Dr. H. Gary Cook, University of Wisconsin, is the session discussant and will provide commentary on the session.

*Audience participation*. At the beginning of the session we will provide audience members with index cards to submit questions they would like to see addressed in a discussion at the end of the session. The session organizers will collect and organize questions thematically during the discussant's presentation, and present themes to the panel for response. Following that, there will be opportunities for open discussion and questions at the end of the session.

## References


Betebenner, D. (2008). Toward a normative understanding of student growth. In K. Ryan & L. Shepard (Eds.), The future of test-based educational accountability (pp. 155-170). New York: Routledge.

Carlberg, C., & Kavale, K. (1980). The efficacy of special versus regular class placement for exceptional children: A meta-analysis. Journal of Special Education, 14, 295-309.

Goldschmidt, P., Choi, K., & Beaudoin, J. P. (2012). *Growth Model Comparison Study:Practical Implications of Alternative Models for Evaluating School Performance*. Council of Chief State School Officers: Washington, D. C.

Hanushek, E., & Raymond, M. (2005). Does school accountability lead to improved student performance? Journal of Policy Analysis and Management, 24(2), 297-327.

Heck, R. (2006). Assessing school achievement progress: Comparing alternative approaches. Educational Administration Quarterly, 42, 667-699.

Hess, F. M. (2005) Commentary: Accountability policy and scholarly research. *Educational Measurement: Issues and Practice*, *24*(4), 53-57.

Kiplinger, V. (2008). Reliability of large-scale assessments and accountability systems The future of test-based educational accountability (pp. 93-114). New York, NY: Routledge.

Ladd, H., & Lauen, D. (2009). Status versus growth: The Distributional Effects of School Accountability Policies (CALDER Working Paper 21). Washington, DC: Urban Institute.

Linn, R., & Haug, C. (2002). Stability of school-building accountability scores and gains. Educational Evaluation and Policy Analysis, 24, 29-36.

No Child Left Behind Act of 2001, Pub. L. No. 107-110 (2002).

Spellings, M. (2005). *Letter to Chief State School Officers Regarding a Pilot Project on Growth Models for Measuring School Accountability*. Retrieved from http://www2.ed.gov/policy/elsec/guid/secletter/051121.html

U. S. Department of Education (2009). *Race to the Top Program: Executive Summary*. Author: Washington, D.C.

Paper I: *Within-Year Achievement Growth Trajectories Using Progress Monitoring Measures*

(334 words excluding title and references)

Gerald Tindal & Joseph F. T. Nese

As part of a response to intervention (RTI) framework, students are administered curriculum-based measures (CBM); technically sound, brief measurements of achievement in reading and mathematics that can be used multiple times within a school year. CBM progress monitoring assessments are used by educators to evaluate students' risk for poor learning outcomes, as well as the effectiveness of instruction (i.e., students' response to intervention). Oral Reading Fluency (ORF) is defined as the number of words read correctly per minute (wcpm) by the student on a grade level reading passage. ORF is the academic construct most often assessed as part of an RTI model (Shapiro et al., 2006; Speece, Case, & Malloy, 2003; Wood, 2006). More knowledge of ORF growth patterns can increase the accuracy of teachers' instructional decisions and also inform researchers' evaluations of educational programs, but little inquiry has yet been made into the nature and patterns of ORF growth of students targeted for progress monitoring.

This presentation expands upon a previous study (Author, 2012) to analyze the within-year ORF growth of students in grades 3-5 who were progress monitored within an RTI framework. Using ORF measures from the 2010-2011 easyCBM system, this study provides a two-level, HLM analysis of growth, with time at level-1 (with up to 15 progress monitoring occasions), and student at level-2 (with student characteristics covariates including ELL, special education, sex, race-ethnicity, and economic disadvantage).

We begin by summarizing the findings of ORF growth, highlighting the consistencies among various studies over the past 20 years. We will note the population characteristics (particularly their selection for progress monitoring), as well as the count and proportion of students for each of the level-2 covariates. Particular attention will be given to the data rendering procedures (the number and percentage of students that are moved in each step of preparing the data for analysis). Emphasis will be given to the findings from one state (Oregon) as opposed to other states participating in center work (for which full student demographic information is not yet available).

## References

Shapiro, E. S., Keller, M. A., Lutz, J. G., Santoro, L. E., & Hintze, J. M. (2006). Curriculum-based measures and performance on state assessment and standardized tests: Reading and math performance in Pennsylvania. *Journal of Psychoeducational Assessment, 24*, 19-35.

Speece, D. L., Case, L. P., & Molloy, D. E. (2003). Responsiveness to general education instruction as the first gate to learning disabilities identification. *Learning Disabilities Research & Practice, 18*, 147-156.

Wood, D. E. (2006). Modeling the relationship between oral reading fluency and performance on a statewide reading test. *Educational Assessment, 11,* 85-104.

Paper II: *Learning to Read: A Review of Research on Growth in Reading Skills*

(347 words excluding title and references)

Shawn Irvin, Joe Nese, & Gerald Tindal

Much of the research on curriculum-based measurement (CBM) in reading has focused on oral reading fluency (ORF). Wayman et al. (2007) concluded that *reading aloud* (the primary CBM in reading) is reliable, related to important measures of achievement, and a sufficiently sensitive metric to reflect growth. However, ORF is but one of five critical skill areas in the wider construct of reading: phonological awareness, phonics, fluency, vocabulary, and comprehension (National Reading Panel, 2000). Yet, early reading skills have come into focus only recently and comprehension is often overlooked within CBM research. As Fuchs, Fuchs, and Compton (2004) note, "additional research is needed to examine the tenability of reading tasks that address an earlier phase of reading" (p.7). In the late 1990s, the research on curriculum-based measurement expanded to include measures of early reading literacy. Probably the most ubiquitous measure to appear is the *Dynamic Indicators of Early Literacy Skills (DIBELS*; Good, Simmons, & Kame'enui, 2001). Other research on early literacy skills explored the importance of phonological awareness and phonics as predictors of reading difficulty (Good, Gruba, & Kaminski, 2002; Good & Kaminski, 1996).

In this presentation, growth in literacy skills is reported using a full range of reading measures, including letter names, letter sounds, phoneme segmentation, word reading fluency, passage reading, and comprehension. These measures are represented in easyCBM, an online system for use in response to intervention (RTI). A synthesis of findings from the published research is first presented using systematic research conducted in the development of the measures. Subsequently, descriptive statistics are presented from teachers' application of easyCBM in the field, including benchmark performance cut scores used to define risk as well as progress monitoring (number of measures administered over time, grade level nature of measurement administration, and density of measurement).

Our findings reflect an emerging research base supporting the use of early literacy measures and their developmental reflection of learning to read. At the same time, we also report on the rather unsystematic use of these measures as teachers craft interventions to teach students to read. We conclude with the need for more research.

## *References*

Fuchs, L. S., Fuchs, D., & Compton, D. L. (2004). Monitoring early reading development in first grade: word identification fluency versus nonsense world fluency. *Exceptional Children, 71*(1), 7-21.

Good, R., Simmons, D., & Kame'enui, E. (2001). The importance and decision-making utility of a continuum of fluency-based indicators of foundational reading skills for third-grade high-stakes outcomes. *Scientific Reading Studies, 5*(3), 257-288.

Good, R. H., Gruba, J., & Kaminski, R. A. (2002). Best practices in Using Dynamic Indicators of Basic Early Literacy Skills (DIBELS) in an Outcomes-Driven Model. In A. Thomas and J. Grimes (Eds). *Best Practices in School Psychology IV* (pp.679-700). Washington, DC: National Association of School Psychologists.

Good, R. H., & Kaminski, R. A. (1996). Assessment for instructional decisions: Toward a proactive/prevention model of decision-making for early literacy skills. *School Psychology Quarterly, 11,* 326-336.

National Institute of Child Health and Human Development. (2000). *Report of the National Reading Panel. Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction:* Reports of the subgroups (NIH Publication No. 00-4754). Washington, DC: U.S.

Wayman, M. M., Wallace, T., Wiley, H. I., Tich'a, R., & Espin, C. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education*, *41,* 85-120.

Paper III: *Special Education Growth: Contrasting Stable and Variable Identification of Special Education Student Status across Grades*

(349 words excluding title and references)

Ann Schulte & Joseph Stevens

The marked achievement gap between students with and without disabilities has been a longstanding concern in education (Hanushek, Kain, & Rivkin, 2002). Most longitudinal studies of the achievement of students with disabilities have used a student's special education status at a single point in time as a basis for classifying students as disabled or nondisabled (e.g., Wei, Blackorby, & Schiller, 2011). However, this analytic strategy misrepresents the disability status of any student who enters or exits special education after the initial classification period. This identification strategy appears to be based on the widely-held assumption that the special education population is sufficiently stable (Finn, Rotherham, & Hokanson, 2001) that single point classification strategies will not affect characterizations of achievement outcomes or trajectories for this population. Contrary to this assumption, longitudinal studies of classification patterns suggest that there is substantial turnover in the special education population. For example, Ysseldyke and Bielinski (2002) found that approximately 20% of the special education population entered or exited services on an annual basis. Similarly, Schulte (2012) reported that only 76 percent of students who were in special education in third grade remained in special education in the fifth grade.

This presentation will use large-scale achievement data from two North Carolina cohorts of students (approximately 100,000 students per cohort) to examine achievement growth from grades three to seven for students in special education. Growth trajectories for students with stable special education status from third to seventh grade (always in special education) will be compared to trajectories for students with unstable special education status (entering or exiting special education some time from third to seventh grade). Achievement at each grade and growth across grades will be contrasted for the multiple methods of defining special education status (always in special education, ever in special education, time varying special education status). In addition, the special/general education achievement gap and its changes over grades will be characterized as a function of the different disability identification strategies. The methodological and policy barriers to employing more complex definitions of disability status in characterizing and monitoring special education outcomes will also be discussed.

References

Finn, C.E., Rotherham, A. J., & Hokanson, C. R. (Eds.) (2001). *Rethinking special education for a new century*. Washington DC: Thomas B. Fordham Foundation and the Progressive Policy Institute.

Hanushek, E. A., Kain, J. F., & Rivkin, S. G. (2002). Inferring program effects for special populations: Does special education raise achievement for students with disabilities? *Review of Economics and Statistics, 84*(4), 584-599.

Schulte, A. C. (2012). *Critical issues for examining special education outcomes in status and growth accountability models*. Paper presented at annual meeting of the National Council on Measurement in Education, Vancouver, BC.

Wei, Blackorby, & Schiller, E. (2011). Growth in reading achievement of students with disabilities, ages 7 to 17. *Exceptional Children, 78*(1), 89-106.

Ysseldyke, J., & Bielinski, J. (2002). Effect of different methods of reporting and reclassification on trends in test scores for students with disabilities. *Exceptional Children, 68*(2), 189-200.

Paper IV: *Reading Achievement Growth at the Student and School Levels for Regular and Special Education Elementary Students*

(347 words excluding title, references, and figure)

Joseph Stevens & Ann Schulte

Despite the promise of growth models, there is a paucity of research that applies these methods to students with disabilities (SWD). Published growth analyses that include special education students have shown that, although there are large differences in intercept, there are often no statistically significant differences in slope for students with disabilities or students receiving a modified test administration (Author, 2005; Wei et al., 2011). Other challenges in estimating SWD growth include scaling issues, representing test accommodations, attrition, grade retention, and changes in special education status over time (Author, 2012a). Growth is often nonlinear both within and between school years (Author, 2005; 2012b; Raudenbush & Bryk, 2002). Accurate estimation of the shape of growth functions is considered to be a critical step in evaluating learning and development (Rogosa, 1979; Willett, et al, 1998) and it is also important to test interactions between functional form and key student characteristics or demographics.

The proposed paper analyzes multi-year, large scale assessment data in mathematics from North Carolina to estimate student growth during elementary school grades 3-5. The cohort analyzed is composed of all students present in 2001 in grade three ($N = 92{,}028$) and tracks students through grades four and five. We apply two-level, HLM growth models to examine student growth over time and include eight demographic variables as well as student exceptionality category as predictors. We also use three-level HLM models to examine growth across schools and include several school level characteristics (e.g., school size, percent free/reduced lunch, percent special education, etc.) as predictors of level 2 parameters. The outcome measure used is the IRT mathematics scale score on the North Carolina state achievement test.

Results show statistically significant relationships between all demographic covariates and initial mathematics performance (intercept) but significant relationships with growth slope only for gender, parental education, and Title I participation. Results also show significant differences between each exceptionality category in comparison to regular education students for intercept but differences in slope only for gifted, other exceptionality, and learning disabled students (see Figure 1). Additional details and results will be presented in the final conference paper.

## References

Raudenbush, S. W., & Bryk, A. S. (2002). *Hierarchical linear models: Applications and data analysis methods* (2nd ed.). Thousand Oaks, CA: Sage.

Rogosa, D. (1979). Causal models in longitudinal research: Rationale, formulation, and interpretation. In J. R. Nesselroade & P. B. Baltes (Eds.). *Longitudinal research in the study of behavior and development* (pp. 263-302). New York, NY: Academic Press.

Wei, X., Blackorby, J., & Schiller, E. (2011). Growth in reading achievement of students with disabilities, ages 7 to 17. *Exceptional Children, 78(1)*, 89-106.

Willett, J. B., Singer, J. D., & Martin, N. C. (1998). The design and analysis of longitudinal studies of development and psychopathology in context: Statistical models and methodological recommendations, *Development and Psychopathology, 10*, 395-426.

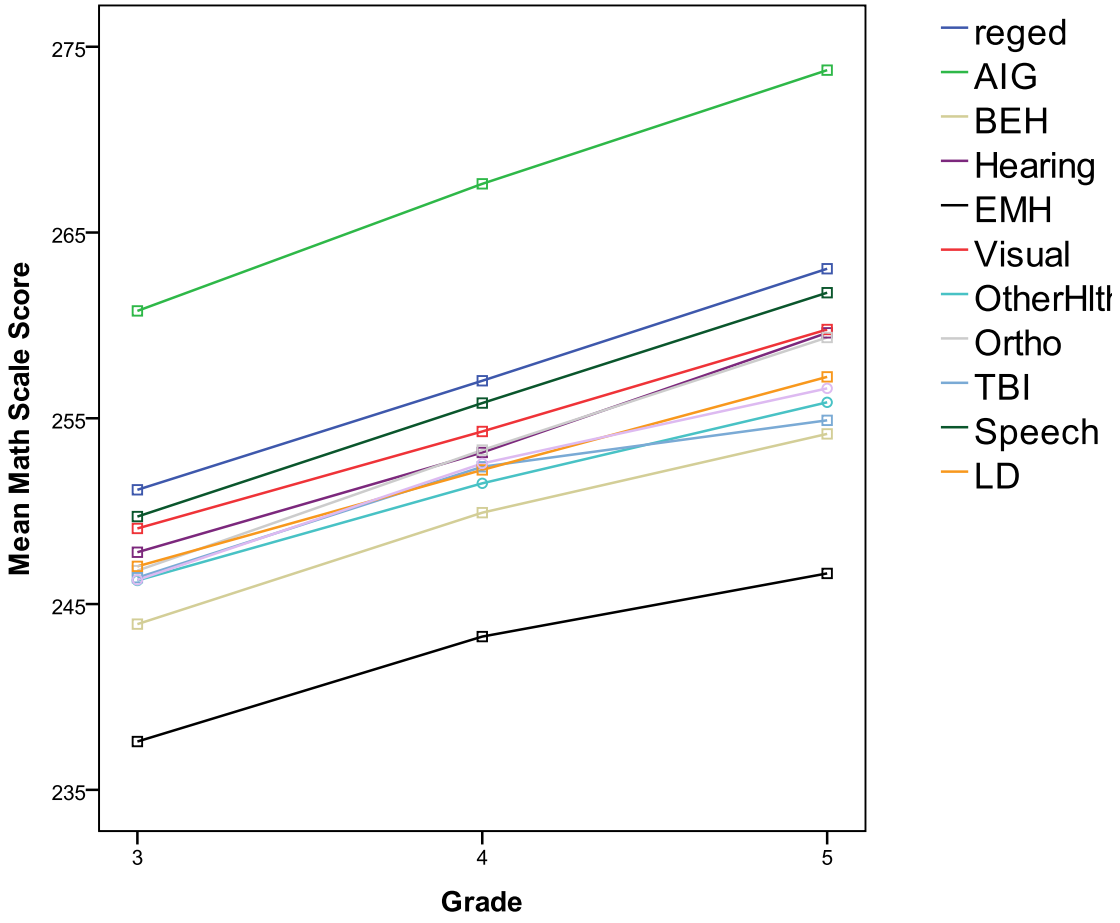


*Figure 1*. Elementary mathematics growth as a function of disability category and grade.

Paper V: *School Effects on the Middle School Reading Achievement of Students with Disabilities: A Multilevel, Longitudinal Analysis*

(339 words excluding title, references, and figure)

Biancarosa & Zvoch

The measurement and assessment of student achievement growth is a central issue in educational research (Raudenbush, 2004; Seltzer et al., 2003; Willett, 1988). Longitudinal study of student achievement enables educational researchers to estimate the rate at which individuals from different population groups acquire certain academic competencies and to determine the extent to which school context and practice impact student growth trajectories (Raudenbush, 2004). However, the differential exclusion of students with disabilities from accountability testing and reporting has left a somewhat incomplete and distorted picture of student and school achievement outcomes (Wei et al., 2011).

The purpose of the present study was to address the need for representative studies of the achievement progress of students with mild to moderate disabilities. In the following, state reading achievement data from a longitudinally matched student cohort were analyzed to investigate the pattern of reading growth among students with different disability classifications ($N$ = 90,593). Use of three-level longitudinal growth models to estimate the growth trajectories of middle school students revealed the presence of relatively large status score differences between different disability categories and with respect to the performance of students in the general education population. On average, academically gifted students demonstrated the highest levels of reading performance whereas students with an intellectual disability had the lowest levels of performance. Rates of growth were more similar. Regardless of classification status, students tended to acquire reading skills at a similar rate with each passing grade (see Figure 1). However, a partition of outcome variance into within and between-school components revealed that while more of the variation in middle school reading achievement and growth occurred within schools, the between school variance in reading growth was large and considerably greater than the amount of between school variance in mean reading achievement. Together these results highlight the heterogeneity manifested within and between different disability categories and suggest that conclusions drawn about student and school performance will vary on the basis of which students are included in a state accountability report. Implications for more inclusive assessment practices are discussed.

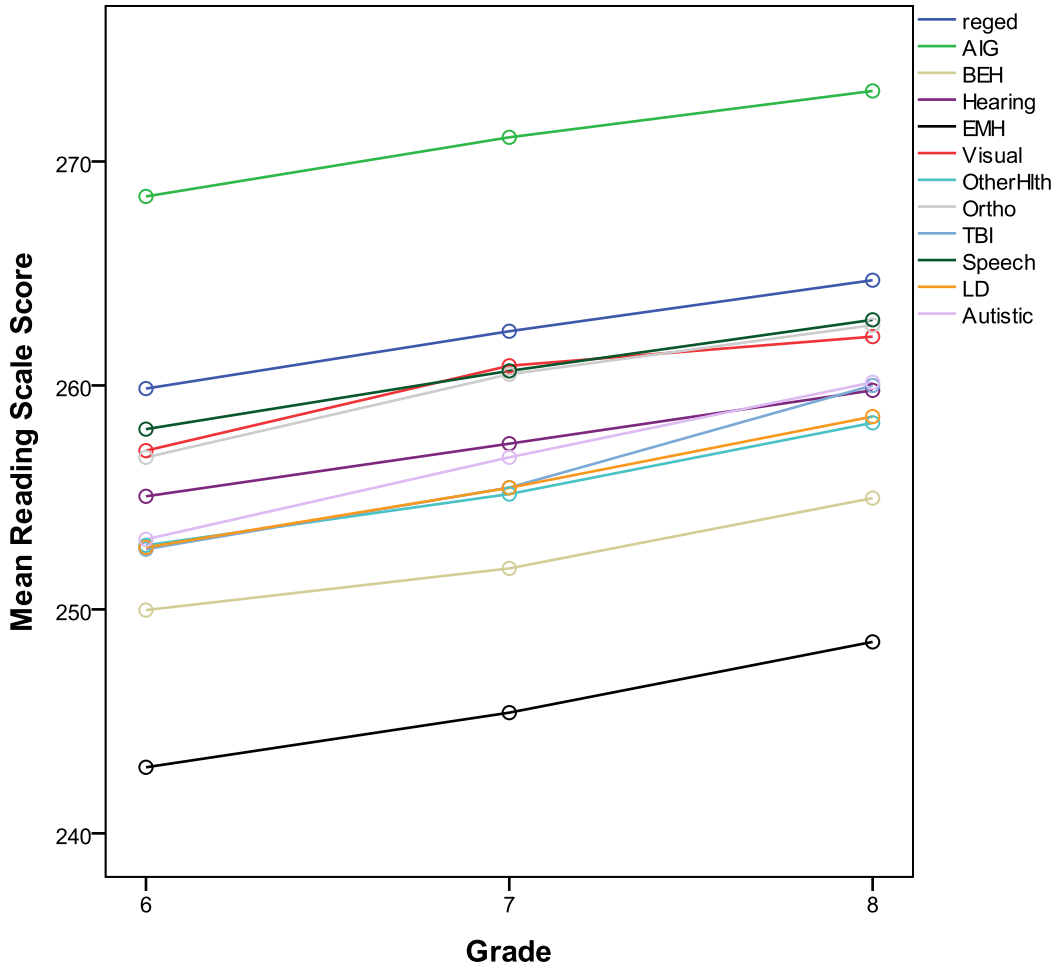


*Figure 1*. Reading growth as a function of disability category and grade.

References

Raudenbush, S. W. (2004). *Schooling, statistics, and poverty: Can we measure school improvement?* Paper presented at the William H. Angoff Memorial Lecture Series, Princeton, NJ. Retrieved from www.ets.org/research/.

Seltzer, M., Choi, K., & Thum, Y. M. (2003). Examining relationships between where students start and how rapidly they progress: Using new developments in growth modeling to gain insight into the distribution of achievement within schools. *Educational Evaluation and Policy Analysis, 25*, 263-286.

Wei, X., Blackorby, J., & Schiller, E. (2011). Growth in reading achievement of students with disabilities, ages 7 to 17. *Exceptional Children, 78(1)*, 89-106.

Willett, J. B. (1988). Questions and answers in the measurement of change. In E. Rothkopf (Ed.),

*Review of research in education 1988-89* (pp. 345-422). Washington: American Educational Research Association.