EXPLORING READING GROWTH PROFILES FOR MIDDLE SCHOOL STUDENTS WITH SIGNIFICANT COGNITIVE DISABILITIES

by

DANIEL PATRICK FARLEY

A DISSERTATION

Presented to the Department of Educational Methodology, Policy, and Leadership
and the Graduate School of the University of Oregon
in partial fulfillment of the requirements
for the degree of
Doctor of Philosophy

March 2017

DISSERTATION APPROVAL PAGE

Student: Daniel Patrick Farley

Title: Exploring Reading Growth Profiles for Middle School Students with Significant Cognitive Disabilities

This dissertation has been accepted and approved in partial fulfillment of the requirements for the Doctor of Philosophy degree in the Department of Educational Methodology, Policy, and Leadership by:

Joseph Stevens	Chairperson
Gina Biancarosa	Core Member
Keith Hollenbeck	Core Member
Laura Lee McIntyre	Institutional Representative

and

Scott L. Pratt	Dean of the Graduate School

Original approval signatures are on file with the University of Oregon Graduate School.

Degree awarded March 2017

DISSERTATION ABSTRACT

Daniel Patrick Farley

Doctor of Philosophy

Department of Educational Methodology, Policy, and Leadership

March 2017

Title: Exploring Reading Growth Profiles for Middle School Students with Significant Cognitive Disabilities

Statewide accountability programs are incorporating academic growth estimates for general assessments. This transition focuses attention on modeling growth for students with significant cognitive disabilities (SWSCD) who take alternate assessments based on alternate achievement standards (AA-AAS), as most states attempt to structure their AA-AAS systems as similarly as possible to their general assessments (GA). Test scaling, group heterogeneity, small sample sizes, missing data, and the use of status-based assessments that were not necessarily designed to measure a developmental continuum complicate modeling growth for SWSCD. This study addressed these challenges by: (a) analyzing test results from a common scale, (b) modeling achievement and growth for students in multiple demographic and exceptionality categories, and (c) using multiple cohorts to increase sample sizes.

Latent growth curve modeling (LGCM) was used to define growth estimates based on exceptionality, sex, race, and economic disadvantage. Unconditional latent class growth analysis (LCGA) was used to determine the number of homogeneous subgroups that existed within the heterogeneous population of SWSCD for subsequent growth mixture modeling (GMM). Unconditional GMM was used to define the number of homogeneous subgroups of students with similar intercept and growth patterns within the overall population of SWSCD. Discriminant function analysis (DFA) including student exceptionality, sex, race, and economic disadvantage status was also used to analyze class membership post hoc.

SWSCD with different exceptionalities generally had significantly different average initial achievement but growth rates that did not differ significantly from each other. SWSCD classified as economically disadvantaged performed significantly lower than their peers in initial achievement, yet exhibited growth rates that were not statistically different than the reference group. This study also found evidence for two separate latent classes of students with exceptionalities on the Oregon AA-AAS. The first class had lower achievement and larger growth rates, while the second class had higher achievement and slower growth rates. Students identified as SLD and CD were generally higher-performing, while students identified as ID, ASD, and OI were lower performing across all analytic models.

CURRICULUM VITAE

NAME OF AUTHOR: Daniel Patrick Farley

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene, OR
University of New Mexico, Albuquerque, NM
St. John's College, Annapolis, MD

DEGREES AWARDED:

Doctor of Philosophy, Educational Methodology, Policy, and Leadership, 2017, University of Oregon
Master of Arts, Secondary Education, 1997, University of New Mexico
Bachelor of Arts, Liberal Arts, 1994, St. John's College

AREAS OF SPECIAL INTEREST:

Special Education Assessment
Respectful Accountability Systems

PROFESSIONAL EXPERIENCE

Courtesy Research Assistant, Behavioral Research and Teaching, Eugene, Oregon, 2011

Education Director, The Child Center, Springfield, OR, 2010-11

PUBLICATIONS:

Farley, D., Anderson, D., Irvin, P.S., & Tindal, G. (2016). Modeling reading growth in grades 3-5 with an alternate assessment. *Remedial and Special Education,* Advanced online publication. doi: 10.1177/0741932516678661

Tindal, G., Nese, J. F. T., Farley, D., Saven, J. L., & Elliot, S. (2016). Documenting reading growth for students with significant cognitive disabilities. *Exceptional Children*, 82, 1-16. doi: 10.1177/0014402915585492

ACKNOWLEDGMENTS

I decided to pursue my PhD over a decade ago, after having met Dr. Tindal. His level of energy and enthusiasm is infectious and awe-inspiring. His support over the years in my role at BRT has been invaluable and is one of the primary reasons that I was able to finish this ambitious project. The support I received from Dr. Anderson allowed me to venture well beyond my comfort zone and I am heavily indebted to him for teaching me how to use *R* as well as for his patient handling of all of my methods questions. Of course the project would not have been feasible without my advisor, Dr. Stevens, whose clear understanding of methods and writing helped make my dissertation better in all respects. I want to thank Dr. Irvin, Dr. Nese, Dr. Melton, and Dr. Rowley, who supported me as I completed this adventure. I would also like to thank each of my committee members, who helped me think more critically about my project, from the proposal to the final dissertation. Thanks to Brad Lenhardt at ODE for supporting my work and helping me address the consequences of the findings in such a collegial manner. The data were also collected as part of the National Center on Assessment and Accountability for Special Education (NCAASE), grant number R324C110004 from the Institute of Education Sciences, and I thank all the principal investigators for allowing me access to the data. I wouldn't have been motivated to take on this task without my children, Hannah, Dylan, and Samuel, who keep my motor running and keep me honest. To my parents and siblings, thank you for modeling an unerringly academic approach to existence and a love of thinking, reading, writing, and discussion. Finally, the largest heaping of kudos goes to my wife, Elizabeth, who put up with me for the last 4.5 years as I waxed and waned through periods of joy and frustration. You are my rock.

For Liz, Hannah, Dylan, and Samuel.

TABLE OF CONTENTS

Chapter Page

Chapter | Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

INTRODUCTION

Prior to the implementation of the Individuals with Disabilities Education Act (IDEA, 1997), students with severe disabilities were largely excluded from statewide assessment systems; federal accountability initiatives targeting more inclusive assessment approaches were developed to address this disparity (Kleinert, Kennedy, & Kearns, 1999). Students with severe disabilities also experienced limited access to academic content, and instruction and assessment focused instead upon functional curricula that taught students how to develop self-care, adaptive behaviors, such as toileting, hygiene, and cooking (Kleinert et al., 1999; Quenemoen, 2008). The realization that academic content can be effectively taught to and mastered by this group of students has been followed by efforts to define how this group of students demonstrates academic growth over time. This is a notable shift, yet there is much that remains unknown regarding how students within this population progress academically.

Initial growth modeling research treated this group of students as a homogeneous population, determining only one growth estimate that was assumed to apply to the entire population. Yet the literature demonstrates that students with severe cognitive disabilities exhibit more differences than similarities even though a single label often defines them. This study attempted to look at subgroup differences and build upon the work of researchers who have been able to explore or generate estimates of academic growth (Dunn, Roussos, Lonczak, & Sukin, 2012; Farley, Anderson, Irvin, & Tindal, 2016; Karvonen, Flowers, & Wakeman, 2013; Tindal, Nese, Farley, Saven, & Elliott, 2015).

Defining the prerequisites for modeling growth for SWSCD and developing a better understanding of the factors that affect how SWSCD grow should support the development of more appropriate assessment and instructional expectations. It is hoped that the results gathered from this project will also help guide policy discussions in the areas of determining how much growth is sufficient, or what factors may be important to consider when setting growth targets, within statewide accountability systems, particularly when considering the well-established heterogeneity of the population (Ysseldyke & Olsen, 1999).

This dissertation begins with a description of the development of AA-AAS, followed by what research says about the growth patterns of students with disabilities (SWD) and the factors that affect growth estimates. Available literature related to the academic reading growth of SWSCD is reviewed next. The methods used to model growth in this challenging context are subsequently elaborated, followed by a description of results and a discussion of the potential impact of these results upon the field. Study limitations are addressed and the dissertation concludes with an overall summary that contains recommendations for future research.

Alternate Assessment Background and Trends

Quenemoen's (2008) analysis of the social impacts of having 15 years of alternate assessments conveyed the danger in making assumptions about what students can learn, "We have learned that we have expected too little of students with significant cognitive disabilities in the past, but they still have much to teach us about what is possible" (p. 26). Initially, students with severe disabilities had limited access to academic content and were instructed primarily with functional curricula (Browder et al., 2002, April; Kleinert, Browder, & Towles-Reeves, 2009; Thompson & Thurlow, 2001). More recently, alternate assessments moved from targeting functional skills toward assessing academic skills, including reading, writing, and mathematics, that are tied to grade level general content standards (Kearns, Towles-Reeves, Kleinert, Kleinert, & Thomas, 2011; Kettler et al., 2010; Quenemoen, 2008).

Related to the transition of statewide accountability systems that include assessments that generate growth estimates, there is increased emphasis on modeling the academic growth of SWD in these statewide accountability systems (Buzick & Laitusis, 2010). SWSCD are included in the discussion, as reporting requirements for alternate assessments mirror those established for the general assessments in the No Child Left Behind Act (NCLB, 2002). Available literature and the efforts of states and national consortia also help identify this trend.

Alternate Assessments: The Result of Federal Mandates

The federal government mandated alternate assessments, though did not define them, in the Individuals with Disabilities Education Act (IDEA) for students whose disabilities precluded meaningful participation in statewide general assessments (1997). The No Child Left Behind Act defined, *alternate assessments based on alternate achievement standards* (AA-AAS), in which the group identifier of *students with significant cognitive disabilities* (SWSCD) was also established (NCLB, 2002). In the subsequent reauthorization of the IDEA, the federal government defined alternate assessments as assessments for the small number of students who could not participate in the regular assessments, even with appropriate accommodations (IDEA, 2004). The recently-adopted Every Student Succeeds Act (ESSA, 2015) continued the statewide assessment requirements for accountability purposes, including what were retitled *alternate assessments aligned with alternate academic achievement standards* (AA-AAAS). The ESSA also includes specific allowances for states to incorporate growth modeling into such assessment and accountability frameworks (ESSA, 2015)

In response to the initial IDEA (1997) requirement to have alternate assessments in place by July 1, 2000, states developed alternate assessments with little guidance, resulting in wide variations in practice. States implemented observational checklists, portfolio systems, performance-based assessments, or hybrids of multiple approaches (Quenemoen, 2008). Pursuant to NCLB, states implemented AA-AAS to meaningfully involve SWSCD in statewide academic assessment systems and public accountability determinations in a manner linked to grade-level expectations.

Out-of-level assessments, where students take an assessment at a grade level that is lower than their assigned grade level, was explicitly forbidden, as it could not demonstrate a sufficient connection to grade level expectations (USDE, 2003). Updated guidance allowed for out-of-level assessment in AA-AAS contexts, but only in situations where states met three criteria: (a) AA-AAS must be set through a validated standard setting process, (b) assessments must be aligned to the state's academic content standards, promote access to the general education curriculum, and reflect the highest achievement standards, and (c) scores from AA-AAS were subject to a 1.0 percent reporting cap established for calculating Adequate Yearly Progress (AYP; USDE, 2005).

The 2003 federal regulations also clarified the definition of AA-AAS, elaborating the purposes and technical requirements. "The requirements for high technical quality set forth in §§ 200.2(b) and 200.3(a)(1), including validity, reliability, accessibility, objectivity, and consistency with nationally recognized professional and technical standards, apply to alternate assessments as well as to regular State assessments" (USDE, 2003, p. 68699). Despite this further clarification, confusion remained regarding AA-AAS, and soon after the U.S. Department of Education published Non-Regulatory Guidance (2005) to provide additional guidance:

> An alternate assessment based on alternate achievement standards may cover a narrower range of content (e.g., cover fewer objectives under each content standard) and reflect a different set of expectations in the areas of reading/language arts, mathematics, and science than do regular assessments or alternate assessments based on grade-level achievement standards. The questions on an alternate assessment might be simpler than those on a regular assessment or the expectations for how well students know particular content standards may be less complex but still challenging for students with the most significant cognitive disabilities. (p. 15)

The Non-Regulatory Guidance (2005) underscored the commitment that AA-AAS must demonstrate a robust connection to grade-level content through strong alignment with state academic standards, promoting access to general curriculum, and reflecting high standards.

Though NCLB (2002) fully coined the term SWSCD*,* the comment section in the December 9, 2003, Federal Register elaborated that the U.S. Secretary of Education removed a precise definition from the proposed regulations, wherein intelligence quotient (IQ) and adaptive behavior scores three or more standard deviations below the mean were included. "Removing the definition while maintaining the one-percent cap gives states and LEAs [Local Education Agencies] more latitude in identifying the population that should appropriately be evaluated against alternate achievement standards, while ensuring that alternate achievement standards are not used as a loophole to evade accountability for unwarrantedly large numbers of students with disabilities" (USDE, 2003, p. 68706). Over time, state eligibility criteria stabilized and identified primarily students with Intellectual Disabilities (ID), Multiple Disabilities (MD), and severe

Autism Spectrum Disorder (ASD) as eligible to participate in statewide AA-AAS. Kearns et al. (2011) found that the overall percentages of these three combined groups ranged from 74.0% to 96.9% of the SWSCD taking AA-AAS in the seven states included in their survey. Kearns et al. (2011) elaborated that SWSCD typically exhibited significant limitations in terms of expressive and receptive communication, severely limited attention resources, and often had complicating orthopedic and medical factors that affected learning.

Statewide Accountability Assessment Context

NCLB required states to report student demographic and performance data from AA-AAS as part of status-based statewide accountability models, in which achievement was measured at a single point in time, typically at the end of the school year. Results from AA-AAS were used to calculate whether or not schools, districts, and states met established performance benchmarks, AYP, in English language arts and mathematics. Science was assessed and reported, but was not subject to AYP requirements.

However, interest is increasing in assessing student growth over time in statewide accountability testing. Some argue that "including measures of student growth in accountability is important because it allows schools and teachers to be recognized for student learning not just for student performance at a fixed point in time" (Buzick & Laitusis, 2010, p. 537). Focusing on growth credits education systems that are increasing achievement in positive ways, but may not be meeting pre-established state standards for performance. Thus, similar to the manner in which state general education systems have evolved to incorporate growth modeling into their accountability practices since the inception of NCLB, so too have practices around AA-AAS moved away from status-based models, and toward measurement of individual student growth (Altman et al., 2010; Elliott, 2015; Rieke, Lazarus, Thurlow, & Dominguez, 2013; USDE, 2015).

Based on the available literature, states have yet to incorporate AA-AAS growth estimates into statewide accountability models despite this momentum and interest. One of the two national AA-AAS consortia, Dynamic Learning Maps (DLM; http://dynamiclearningmaps.org), whose membership included 17 states, claimed that their AA-AAS in English language arts and math provided information about student growth that was connected to complex learning maps. Validation of this claim has not been provided on their website, or published, at the time of this dissertation. The second national consortium, the National Center and State Collaborative (NCSC; http://www.ncscpartners.org), whose membership included 11 states and three other jurisdictions (Pacific Assessment Consortium, U. S. Virgin Islands, and the District of Columbia) did not include the words *grow* or *growth* in the description of its project or products. Thus it appeared that modeling of academic growth has not yet been included in new, consortium-based alternate assessments. The NCSC website listed 10 additional states that were participating in the consortia in an unidentified capacity. At least 17 states were involved in efforts that were ostensibly pursuing academic growth estimates for SWSCD. There were 12 states implementing independent AA-AAS programs outside of these two consortia, some of which might be in the process of implementing growth models. This is unknown, however.

Academic Growth Research for SWD

It is important to review the growth modeling research available for SWD, as available research helps us understand the factors that affect intercept and growth estimates. Comparisons between SWD and SWSCD might yield important similarities and differences to consider. The research documenting achievement growth for SWSCD is in a nascent stage. A number of studies focused on academic growth models based on large-scale assessments for SWD while accounting for exceptionality status in some manner (Blackorby et al., 2005; Judge & Bell, 2010; Judge & Watson, 2011; Morgan, Farkas, & Wu, 2011; Schulte, Stevens, Elliott, Tindal, & Nese, 2016; Stevens, Schulte, Elliott, Nese, & Tindal, 2015; Wei, Blackorby, & Schiller, 2011; Wei, Lenz, & Blackorby, 2012), yet only four explored academic growth for SWSCD who participated in AA-AAS (Dunn et al., 2012; Farley et al., 2016; Karvonen et al., 2013; Tindal et al., 2015). The Tindal et al. (2015) and Farley et al. (2016) studies explored the feasibility of growth models and are the only SWSCD studies to generate growth estimates based on discrete scaled scores tied to a common scale.

Researchers revealed patterns in studies addressing reading and mathematics growth on statewide accountability assessments for SWD, though general agreement is not always apparent. General education peers generally outperformed SWD academically in terms of average initial achievement in reading (Blackorby et al., 2005; Morgan et al., 2011; Schulte et al., 2016) and in mathematics (Judge & Watson, 2011; Morgan et al., 2011; Stevens et al., 2015). Growth rates, however, were larger for SWD in reading, yet insufficient to close the achievement gap between SWD and students in general education (Schulte et al., 2016). The relative performances between exceptionality subgroups showed that students identified with Speech-Language Impairment (SLI; termed Communication Disorder, or CD, in Oregon) and Visual Impairments (VI) exhibited the highest reading comprehension skills on the Woodcock-Johnson Tests of Achievement, Version 3 (WJ-III) Passage Comprehension subtest, while students identified as having an Intellectual Disability (ID), Autism Spectrum Disorder (ASD), or Multiple Disabilities (MD) were the lowest performing (Blackorby et al., 2005).

In addition, demographic characteristics of students identified as having a specific learning disability (SLD) or CD relate to academic outcomes. Exhibiting economic disadvantage (FRL; eligible for Free/Reduced Price Lunch), or being African American, Hispanic, and/or Female was associated with lower reading achievement and growth (Morgan et al., 2011) as well as lower mathematics achievement and growth (Morgan et al., 2011; Wei et al., 2012) compared to other SWD.

Schulte et al. (2016) found that achievement gaps between SWD exceptionality groups in reading comprehension based on the North Carolina End of Grade Reading Comprehension (EOG-RC) in Grades 3 to 7 remained relatively stable over time compared to general education peers, and that growth was not linear, as more growth occurring at lower grades than upper grades. The authors pointed to the importance of selecting similar intercepts, constructs, and scales when making comparisons across studies of academic growth.

Schulte and Stevens (2015) used statewide longitudinal data and hierarchical linear modeling (HLM) to demonstrate that SWD exhibited lower achievement and slower growth in math compared to non-disabled peers. However, they noted that Hispanic and African American SWD exhibited more rapid mathematics growth compared to non-disabled peers of other ethnicities.

Wei et al. (2011) used HLM to estimate student growth, while incorporating multiple exceptionality categories as covariates, and determined that average reading performance differed substantially between exceptionality groups for SWD aged 7 to 17. Students identified with CD or VI performed the highest, whereas students identified with MD and ID performed at the lowest levels in comparison to students identified with SLD. Quadratic growth estimates were comparable across exceptionality categories, but growth rates were slower for students with ASD and CD. Wei and colleagues recommended accountability approaches that account for specific student exceptionality categories rather than aggregating into a single SWD group. The authors were able to include the majority of the IDEA exceptionality categories in their analyses due to the Special Education Elementary Longitudinal Study (SEELS) dataset analyzed, which had sufficient *n*-sizes to support growth estimates for even low-incidence categories (e.g., Deaf-Hard of Hearing, Traumatic Brain Injury, Deaf-Blind). Nonetheless, Wei et al. (2011) did not include SWSCD in their study because the assessment administered, the Woodcock-Johnson Tests of Achievement, 3rd Edition (WJ-III), was generally inaccessible to SWSCD, even with appropriate accommodations.

In a follow-up study examining mathematics achievement, Wei et al. (2012) again used quadratic HLM growth models. They found that students with a CD or VI had the highest average math performance, while those with MD or ID demonstrated the lowest average math performance in comparison to students who were identified as SLD. Growth rates for students with ASD were significantly lower than those for students with SLD, whereas growth rates for students with CD were significantly higher than the SLD reference group. Wei and colleagues also found significant achievement differences based on sex and race, with males outperforming females, and White students outperforming African American students. The researchers did not find that the achievement gap widened over time among demographic groups, with the exception of the White-Hispanic achievement gap. Similar to the earlier reading study, SWSCD were not included in this mathematics study because the WJ-III was deemed inaccessible.

According to Schulte et al. (2016), longitudinal growth research findings in reading and mathematics for SWD exhibited several patterns, depending upon the selected scale, sample, reference group, intercept (e.g., grade level, season), and construct (e.g., letter naming, comprehension, numeracy). Initial achievement analyses showed that SWD generally performed at a lower level than general education peers, with considerable variance around intercept related to exceptionality membership (Schulte et al., 2016). Students identified with ID were generally the lowest-performing exceptionality group, while students identified with CD were generally the highest-performing group. Two exceptions were noted in the literature. Morgan et al. (2011) found that students identified with SLD performed higher in reading comprehension than students identified with CD at the end of 1st Grade. Schulte et al. (2016) found that students identified with ASD outperformed students identified with SLD in reading, though their study did not include students identified with severe ASD, who typically participated in the AA-AAS, and only included students in Grades 3 to 7.

Longitudinal growth estimate patterns were difficult to compare in the literature, as different approaches generated different patterns (i.e., stable, fan-spread, or fan-closed). In addition, most of the available studies involved only students identified with SLD or CD, and thus were not representative of the SWD spectrum. A consistent finding was that growth was curvilinear, with more growth occurring at lower grades than at upper grades. Schulte et al. (2016) found that reading achievement gaps for SWD, not merely students identified with SLD and CD, remained largely stable across Grades 3 to 7, though noted that students identified with SLD in reading closed the achievement gap over this timeframe. Economic disadvantage and Female status was consistently associated with lower intercept and slope estimates, as well. Hispanics and African Americans demonstrated lower intercepts and slopes in reading and mathematics, but there were conflicting findings regarding growth estimates in mathematics. Stevens et al. (2015) found that Hispanic and African American students grew at a significantly faster rate in mathematics, while Morgan et al. (2011) and Wei et al. (2012) found growth rates to be decelerating over time for these subgroups.

In addition, Stevens and Schulte (2016) analyzed interaction effects in mathematics achievement, with White, male, non-SWD, non-Economically Disadvantaged (non-FRL) students as the reference group in comparison to students identified with SLD representative of these categories. The authors found that differences between non-SWD students and students identified with SLD were consistently moderated by sex, race, and economic disadvantage status across Grades 3 to 7. These findings demonstrated that heterogeneity within subgroups was critical to model because the presence of significant subgroup achievement gaps might be masked if interaction effects are not accounted for.

Academic Growth of SWSCD

Within the AA-AAS context, states reported the performance of SWSCD who took the assessments in comparison to proficiency categories. Standard setting procedures were typically used to provide labels for each category, establish cut scores for making such determinations, and define achievement level descriptors (ALDs) that describe what performance at each level means (Perie, 2007). Federal requirements established that performance must be reported in reference to at least three levels, one at the level of proficiency, one above, and one below (NCLB, 2002). However, most states reported performance on AA-AAS in reference to four categories, such as Not Proficient, Nearing Proficiency, Proficient, and Advanced (Perie, 2007).

As accountability systems incorporate growth expectations for SWSCD into status-based approaches to determining proficiency challenges are presented. Status-based assessment systems were not developed to model change over time. The importance of alignment between instruction, curricula, and assessments that are tied to an empirically-founded developmental continuum, or learning progression, that reflects the range of expected academic growth should guide measurement systems designed to model the academic growth of SWSCD. The shortcomings present in the current AA-AAS systems and available literature was reviewed in reference to this context.

Dunn et al. (2012) used six different transition matrix (TM) models to examine how student performance category attainment changed across multiple years in reading and mathematics. The authors used a framework that included nine performance levels, four of which were considered below proficient. They varied cohorts, the number of years included in the models (two or three), the number of performance levels used below the proficient level, as well as the maintenance of proficiency status over the studied period (i.e., did students who were below proficient and became proficient remain proficient) to analyze growth. Dunn and colleagues also looked at growth within a performance level for the Florida SWSCD involved in the study. The authors found that combining a model that accounted for measurement error and within-performance level growth was the most comprehensive of the TMs studied. The researchers argued their results generally suggested that more discrete approaches, including TM approaches that include within-level growth, identified growth more effectively. However, these results were of limited utility in most statewide accountability contexts, as states were required to have only three performance categories, not nine or more (NCLB, 2002)

In a precursor to modeling growth for SWSCD, Karvonen et al. (2013) conducted an exploratory, descriptive study using extant AA-AAS data. The authors applied TM, growth to standards (GTS), and ordinary least squares (OLS) regression models to three unidentified states' data from spring 2009 to 2011 in Grades 3 to 8 in reading and mathematics. The authors described several challenges that impeded growth modeling, including sample size (despite the fact that they collapsed across grade-levels), test design (the measures were developed without a vertical scale and reflected status, not growth), low correlations among the outcomes (even within states), and missing data. Karvonen, et al. (2013) cautioned that growth estimates might not convey actual growth, as there were no guarantees that the content on each assessment was articulated on a vertical or developmental continuum. The same critique could have been levied against the measures employed and the measurement scales involved. The authors pointed to the national consortia that were developing new AA-AAS (i.e., DLM and NCSC) as potentially being able to pool data across many states and model growth for SWSCD.

Tindal et al. (2015) compared linear reading growth in Grades 3 to 5 in Oregon (OR) from spring 2009 to 2011 for SWSCD using both TM and HLM methods. The researchers found that a majority of SWSCD remained at the same performance level from one year to the next with the TM approach, whereas in the hierarchical model, students' scores revealed small but statistically significant growth from year to year. They found that students from economically disadvantaged backgrounds, students who participated in general education classrooms more than 40% of the time each day, and students at the proficient level had significantly higher achievement scores compared to the reference group. Interpreting the performance of students from economically disadvantaged backgrounds was not attempted, but the authors recommended a review of additional characteristics that might explain this unusual finding. SWSCD who participated in general education classrooms more than 40% of the school day also demonstrated significantly more growth, as did students who were rated proficient. The authors explained that general education participation and proficient performance likely accounted for this difference, as students who spend more time in general education contexts or earn proficient status were likely higher-functioning, which confounded results. There were no significant differences between achievement or growth trajectories for the remaining student covariates analyzed, including sex, race, and intellectual exceptionality status. These results suggested that other characteristics were more likely to explain variations around growth trajectories for SWSCD.

Farley, et al. (2016) modeled reading growth for SWSCD in Grades 3 to 5 from 2009 to 2011 with a nonlinear latent growth curve model (LGCM) that incorporated student exceptionality status (ID, CD, OHI, ASD, SLD, and Low Incidence – Emotional Disturbance, Orthopedic Impairment, Deafness, Deaf-blindness, Traumatic Brain Injury, and Visual Impairment) and missing data patterns into the structural equation model as covariates. The authors determined that exceptionality status significantly impacted intercept, yet had non-significant effect on slope estimates. Adding missing data pattern to the models improved model fit, suggesting that incorporating missing data patterns as control variables into modeling procedures was an important procedural step to include in growth modeling with AA-AAS. The authors also suggested that incorporating a known missing data mechanism called test switching, where student participation in the AA-AAS, general assessment (GA), and missingness is modeled as a control variable, might help to account for missing data and improve model fit.

Challenges in Modeling Growth for AA-AAS

Test scaling and design features presented the first challenge to consider in modeling growth for AA-AAS. Observational checklists and portfolio-based assessments were generally designed to provide a snapshot of students' knowledge and abilities at a specific time, but scores were not always tied to a measurement scale or comparable over time. Performance-based assessments provided potential for monitoring students' growth, but proper scaling was required to link the assessments to a common scale over time (which was unusual for AA-AAS). Test design features, particularly scaling, must be part of the growth modeling approach if reliable growth estimates over time are expected.

A common scale is generally a prerequisite for modeling longitudinal growth, which typically requires at least three years of data. Using LGCM and Item Response Theory (IRT) -based, continuous interval scales allow for estimates of the magnitude of growth, modeled over specific time periods. However, most states reported the results of AA-AAS on categorical proficiency scales, greatly complicating attempts to estimate growth. The AA-AAS included in this study was unusual in that it was tied to a common, interval scale in Grades 3 to 5 and 6 to 8, respectively. In essence, students in these grade bands took the same test; they simply had increasing performance expectations across the three-year period. This is an optimal situation for modeling longitudinal growth and expanding theory, but was not necessarily optimal for inclusion within accountability contexts due at least in part to annual reporting requirements.

A substantial challenge to modeling the academic growth of SWSCD, in addition to the lack of a common, continuous scale and the modeling approaches undertaken, is the issue of missing student data, which is pervasive in longitudinal AA-AAS datasets. For example, Saven, Farley, and Tindal (2013) found that of the 1,182 students who took the OR AA-AAS in Grade 3, only 293 students remained by Grade 8, having participated in all years and made typical matriculation patterns. Students may move out of or into the AA-AAS for many reasons. Saven, Anderson, Nese, Farley and Tindal (2015) investigated the extent to which students switched between the alternate and general accountability assessments between years. Perhaps unsurprisingly, the authors found higher test-switching likelihood rates than would be expected by chance, with students who performed in the top performance level of the AA-AAS or the bottom performance level of the general assessment more likely to switch test types between years. Such missing data patterns are inherently non-random. Because of the degree of missingness and the non-random nature of the missing data patterns, excluding missing data from modeling may introduce bias into growth modeling parameter estimates.

Group heterogeneity and small sample sizes are also issues that must be addressed in SWSCD growth modeling contexts. Heterogeneity must be addressed in order to generate accurate expectations for heterogeneous subgroups within a population that was treated as homogeneous. Grand means hide the unique performances of student subgroups that deviate from the norm. Small sample sizes can be addressed by gathering additional data. One way to do this is by including multiple cohorts.

Summary

There are only a few studies that successfully modeled academic growth for SWSCD. The research suggests that discrete approaches, founded in continuous scaled scores and not merely in proficiency categories, are better able to capture and model growth at a finer level of detail (Dunn et al., 2012; Farley et al., 2016; Tindal et al., 2015). In addition, the group of SWSCD is heterogeneous, which requires modeling strategies that can account for inter-individual differences and avoid making homogeneous group performance assumptions (Saven et al., 2015; Wei et al., 2011; Wei et al., 2012; Ysseldyke & Olsen, 1999). The reliability of growth estimates across time is also a significant concern (Tindal et al., 2015). Using a modeling approach that is able to account for error should increase the accuracy and consistency of growth estimates and allow for more appropriately interpretable results (Kline, 2016).

There are many challenges to modeling students' growth, even with assessments that are designed for that purpose (Raudenbush, 2001; Raudenbush & Bryk, 2002; Rogosa, Brandt, & Zimowski, 1982). For example, a vertical scale is generally a prerequisite to meaningful interpretation of score changes. Without a vertical scale, observed score changes are difficult to interpret, as changes in students' ability are confounded with changes in the measurement scale. Similarly, if the functional form of the growth trajectory is not adequately modeled, estimates will over- or under-represent the true trajectory at specific points in time (Rogosa & Willett, 1985). Challenges to modeling growth are amplified when the student group of interest is SWSCD (Dunn et al., 2012; Farley et al., 2016; Karvonen et al., 2013). Substantial heterogeneity exists among SWSCD in terms of skills (Ysseldyke & Olsen, 1999). Small sample sizes exist for student subgroups, inconsistent exceptionality classifications occur, and student impairments are often confounded with constructs assessed (Koretz & Hamilton, 2006; Tindal, Schulte, Elliott, & Stevens, 2011).

Study Context and Purpose

Though other approaches, such as TMs that include within-level growth estimates, may be necessary for states who do not have the requisite systems in place to conduct latent growth evaluations, and latent growth evaluations may not be directly useful in accountability contexts, results in the relevant literature suggest that TM, GTS, and OLS approaches may not be sufficiently sensitive to detect the academic growth of SWSCD. Using a more sophisticated statistical model, such as HLM or LGCM, provides several advantages over TM, GTS, or OLS methods, including increased flexibility in modeling change across multiple years, more sophisticated methods for handling missing data, and methods for examining different functional forms of growth (e.g., linear, curvilinear, quadratic).

The purpose of this study was to estimate the reading growth trajectories of middle school SWSCD in OR and to parse out homogeneous subgroup profiles within a heterogeneous population of SWSCD. These homogeneous subgroups, or classes, were analyzed post hoc to determine the demographic characteristics of the classes. The study context was unique, in that a common assessment vertically linked over grades was administered in three consecutive years from Grades 6 to 8.

Research Questions

Prior research generally relies on average estimates of slope, although some studies have included relevant covariates, such as sex, race, economic disadvantage, education setting, prior performance level, and IDEA exceptionality (see Tindal et al., 2015, for example). Given the gaps in the literature and the known heterogeneity of this population, this dissertation addressed five research questions:

1. How much average reading growth do SWSCD demonstrate in Grades 6 to 8 from 2009 to 2012?
2. Do academic reading growth trajectories vary by IDEA exceptionality classification?
3. How many homogeneous subgroups, or latent classes, of students' reading growth are present in AA-AAS data in Grades 6 to 8?
4. If there are multiple classes, what intercept and slope differences define the performance of students in each latent class?
5. Are there student demographic patterns that are related to class membership, including differences based on exceptionality, sex, race, or economic disadvantage?

CHAPTER II

METHODS

All data preparation for this dissertation was conducted in the statistical computing environment *R* (R Core Team, 2016). Data visualization was conducted with base *R*, as well as the *ggplot2* package (Wickam, 2009) and *esvis* package (Anderson, 2017). Multiple methods were used to address the research questions. First, Latent Growth Curve Modeling (LGCM), which fits growth trajectories within an SEM framework (Bollen & Curran, 2006; Kline, 2016), was employed to address research questions (a) and (b). All SEM analyses were conducted with *Mplus 7.3* software (Muthén & Muthén, 1998-2015). All models used IRT-based scale scores that were expressed on a common scale across grades. All LGCM models were run using full information maximum likelihood (FIML) estimation. Effect size (ES) based on observed means were calculated and reported (Bloom, Hill, Black, & Lipsey, 2008; Farley et al., 2016). The ES calculations were based on the difference between the two means of interest divided by the pooled standard deviation, or Cohen's *d* (Cohen, 1988). For purposes of ES interpretation, I interpreted the magnitude of an ES using Cohen's rules of thumb (Cohen, 1992), small (.20), medium (.50), and large (.80). Latent class growth analysis (LCGA) and growth mixture modeling (GMM; Bilir, Binici, & Kamata, 2008; Jung & Wickrama, 2008; Nesselroade & Ram, 2004) were used to address research questions (c), (d), and (e). Following the cutoff criteria suggested by Hu and Bentler (1999), I used the following rules of thumb to define good model fit: a non-significant chi-square test, a comparative fit index (CFI) greater than .95, a standardized root mean square residual (SRMR) less than .08, and a root mean squared error of approximation (RMSEA) less than .06 (Hu & Bentler, 1999).

Sample

The analytic sample for this study included the reading portion of the AA-AAS for two cohorts of middle school students in OR. Cohort 1 included extant data from the operational administrations of the AA-AAS in Grades 6 to 8 in the spring of 2009, 2010, and 2011, respectively, while Cohort 2 included data from Grades 6 to 8 in the spring of 2010, 2011, and 2012, respectively. Students were included in the sample irrespective of whether they made typical grade level progressions. The grade level progressions of the sample were quite stable, however, as only 40 students did not make the typical Grade 6 to 7 to 8 progression (2.5%). Of this group, 23 students were retained at least once over that time period, while 17 advanced two grade levels in one school year. Cohort 1 included 798 students and Cohort 2 included 814 students. The total analytic sample thus included 1,612 students.

Student demographic variables for the study included sex (Male/Female), race (White/Non-White), economic disadvantage (FRL/non-FRL), and student IDEA exceptionality subgroup (elaborated below). Demographic variables were assigned based upon the most frequent subgroup identification across the three years. When the frequency of identification was equal, students were assigned a subgroup identification using random assignment (e.g., ID in Grade 6, missing in Grade 7, OHI in Grade 8). The percent of students who were randomly assigned to a demographic category was small, ranging from 0-3% of the student sample, demonstrating that student demographic classifications for the sample were stable over the time period studied. The dataset thus included complete data for all student demographics. The sample was composed of students who were 66.3% male, 60.7% White, and 30.3% economically disadvantaged. These demographic characteristics were consistent with population expectations in the state at that time for students participating in the AA-AAS. Correlations among demographic variables, exceptionality status, and outcomes are provided in Appendix A.

Power analyses were conducted a priori with *G*Power*, 3.1.9.2 (Faul, Ehrdfelder, Lang, & Buchner , 2009) to determine the minimum sample sizes needed to be able to detect effect sizes in the .20 to .25 range 80% of the time. Results demonstrated that samples sizes from 34 to 42 students were sufficient to detect effect sizes in this range. This minimum also happened to coincide with the minimum *n*-size for subgroup reporting of 42 defined by ODE (ODE, 2011A). I retained student exceptionality subgroups with *n*-sizes greater than 42 for analysis. Students with no exceptionality status at the time of the first administration were also excluded because they likely should not have participated in the alternate assessment ($n = 28$). The *n*-size exclusion rule reduced the sample by a total of 58 students (3.5%), who were all students identified with low-incidence disabilities (i.e., hearing impairment, visual impairment, and traumatic brain injury; no students who were deaf-blind participated in the AA-AAS). The overall sample means were virtually unchanged by the deletion of these records. The analytic sample of SWSCD students was composed of approximately 32.5% who were identified with intellectual disabilities (ID), 6.8% who were identified with communication disorder (CD), 2.9% were identified with an emotional disturbance (ED), 3.9% were identified with an orthopedic impairment (OI), 10.7% were identified with an other health impairment (OHI), 19.2% were identified with autism spectrum disorder (ASD), and 24.0% were identified with a specific learning disability (SLD). IDEA exceptionality was relatively stable over time for the sample of SWSCD. Longitudinal stability for all student exceptionality categories in the sample is displayed in Figure 1.

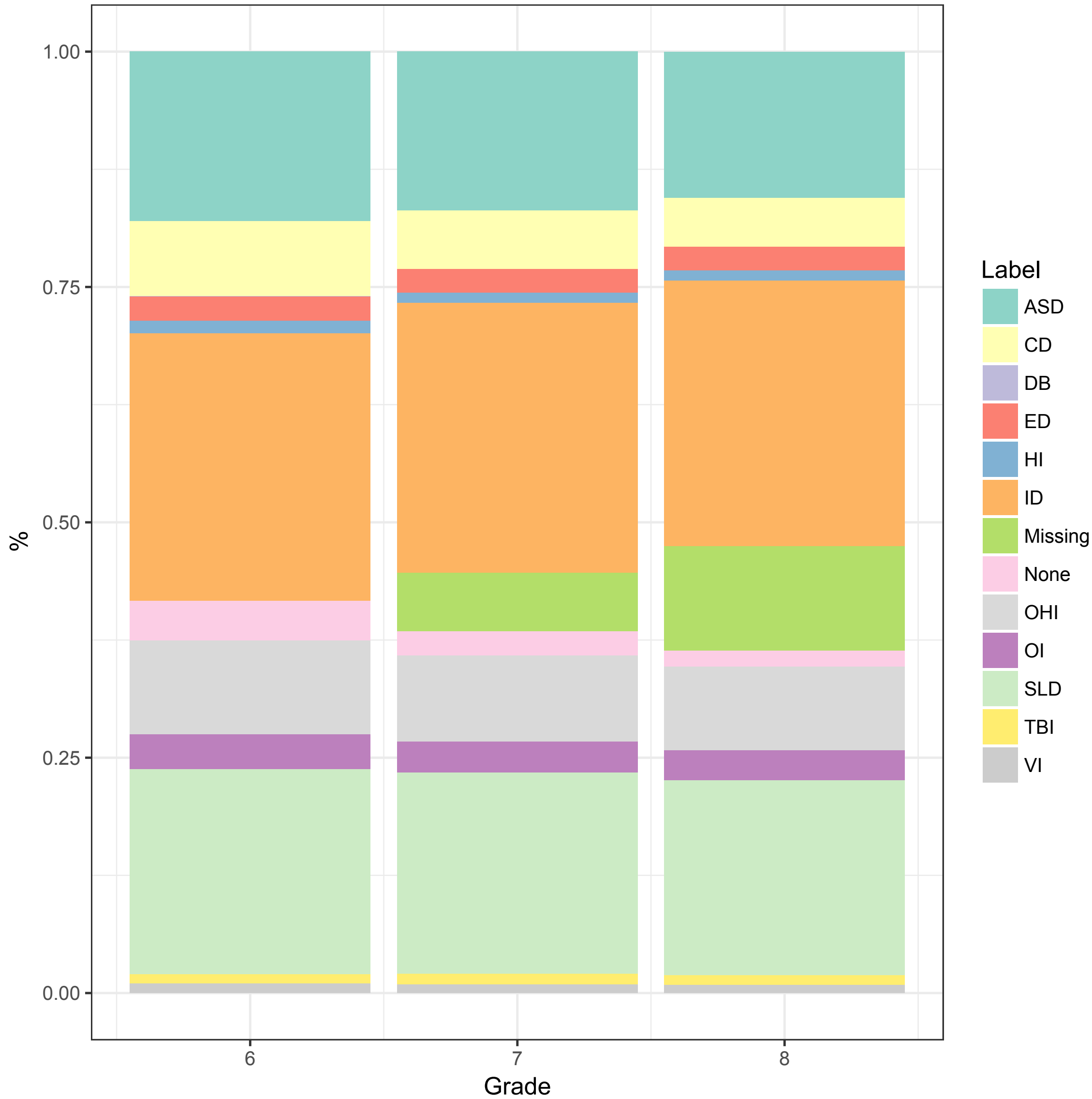

Figure 1. IDEA exceptionality percentages from Grade 6 to 7 to 8. ASD = Autism Spectrum Disorder, CD = Communication Disorder, DB = Deaf-Blindness, ED = Emotional Disturbance, HI = Hearing Impairment, ID = Intellectual Disability, OHI = Other Health Impairment, OI = Orthopedic Impairment, SLD = Specific Learning Disability, TBI = Traumatic Brain Injury, VI = Visual Impairment.

Measures

The reading measure was composed of two versions: *Standard* and *Scaffold*. The Scaffold version included the same item prompts as the Standard version, but with supports such as additional directive statements and visual supports. Each version included a *Prerequisite Skills* task and eight *Content* tasks. The Prerequisite Skills task included 10 items and was used to determine the level of support students required during the administration of the Content tasks in order to effectively demonstrate their knowledge and skills (Anderson, Farley, & Tindal, 2015; ODE, 2009, 2010, 2011B, 2012). The Content tasks assessed students' academic knowledge and skills linked to the OR reading content standards. The standards included vocabulary, locate and synthesize information, main idea, supporting details, themes, characterization, making predictions, fact versus opinion, and author's purpose. Typical test items asked the students to generate synonyms for words presented, translate idiomatic phrases, interpret schedules, read directions, identify the main idea and supporting details in a story, make predictions about what would happen next in a story, and identify the fact in a sentence (ODE, 2009, 2010, 2011B, 2012). In addition to the 10 referenced prerequisite skills items, there were 40 content items composed of eight tasks of five items each. Only the content item scores were used for state accountability analyses and reporting.

The assessment used a paper/pencil administration distributed via a secure state website. Trained district staff members served as qualified assessors and individually administered the reading portion of the state's AA-AAS. Students were asked to select the appropriate answer from three answer choices with each item scored on a three-point scale (0 = incorrect, 1 = partially correct, 2 = correct). Alternate forms of the same test were administered over the three consecutive years, with scores calibrated on a common, item response theory (IRT) scale. RIT scale scores were used for all analyses, estimated via a partial credit Rasch model (Masters, 1982), where RIT = Rasch Unit = $(\theta * 10) + 100$, with θ representing students' estimated latent ability level from the partial credit model (ODE, 2009, 2010, 2011B, 2012). The internal consistency of the measures was high, with total test coefficient alphas reported at .97, .97, .94, and .97 for 2009-2012, respectively (ODE, 2009, 2010, 2011B, 2012). The correlation of Grade 6 RIT scores with Grade 7 RIT scores was 0.86, the correlation of Grade 7 RIT scores with Grade 8 RIT scores was 0.86, and the correlation of Grade 6 RIT scores with Grade 8 RIT scores was 0.83 (ODE, 2009, 2010, 2011B, 2012).

Missing Data

Data were analyzed to determine the extent of missing data and any discernible patterns of missingness. Little's Missing Completely at Random (MCAR) test calculated using the MissMech *R* software package (Jamshidian, Jalal, & Jansen, 2014) was computed and was statistically significant ($p < .05$), indicating data were probably not MCAR, a condition which could result in estimation bias (Little & Rubin, 2002). Because one of the most likely reasons for missing data in this sample was the switching of students from the alternate to the general assessment or vice versa, a dummy-coded test-switching variable was analyzed, where students who switched test type from one year to the next were defined (Saven, et. al, 2015). Appendix B shows the different test taking patterns, and the number and percentage of students with each pattern. As can be seen in Appendix B, approximately 54% of students in the sample participated in the AA-AAS for all three years (n = 874), 27% participated in the AA-AAS for two of the three years (n = 423), and 20% participated in the AA-AAS for one of the three years (n = 315).

Approximately 14% of students had missing data at the first time point, while about 24% and 33% of students had missing data at the second and third time points, respectively. All LGCM used in the study applied full information maximum likelihood (FIML), to address missingness. The default estimator for LCGA and GMM was full information maximum likelihood estimation with robust standard errors (MLR).

Analytic Models

Three analytic models and one post hoc evaluation were used to address the five study research questions. Model 1 addressed Research Questions 1 and 2 regarding reading average reading growth and growth based on student demographic and IDEA exceptionality subgroups, respectively. Model 2 addressed Research Question 3 examining the number of latent reading growth classes, Model 3 addressed Research Question 4 estimating the variance of the intercept and slope estimates in the identified latent classes, and post hoc discriminant function analysis addressed Research Question 5. Each of these models is described below.

Competing models were compared using Akaike's information criterion (AIC) and the Bayesian information criterion (BIC). AIC and BIC are transformations of the log-likelihood, balancing model fit with parsimony (i.e., both fit indices include penalties for the number of estimated parameters), in which lower values indicate better fitting models. Akaike weights for competing models were calculated (using both AIC and BIC; see Burnham & Anderson, 2004), which transformed the raw AIC and BIC values into conditional probabilities for each model. The weight for model *i* was defined as:

$$w_i \ = \frac{\exp\{-\frac{1}{2}\Delta_i \ (AIC/BIC)\}}{\sum_{k=1}^{K}\{-\frac{1}{2}\Delta_k \ (AIC/BIC\}} \qquad (1)$$

where Δ represents the difference between the value for model *i* and the value for the model with the lowest criterion value. This formula transforms the AIC/BIC estimates to a probability scale. The weights can be interpreted as the relative evidence in favor of one model over the best fitting model in the set of models.

The functional form of the growth trajectories was empirically evaluated four ways. First, descriptive and visual methods were used to examine the shape of observed growth over the three grades. Then a linear model was analyzed, followed by a model freely estimating the third time point for linearity (Kamata, Nese, Patarapichayatham, & Lai, 2013). Last, a fixed, quadratic effect model was fit to determine whether there was any improvement in variance explained over and above the linear model.

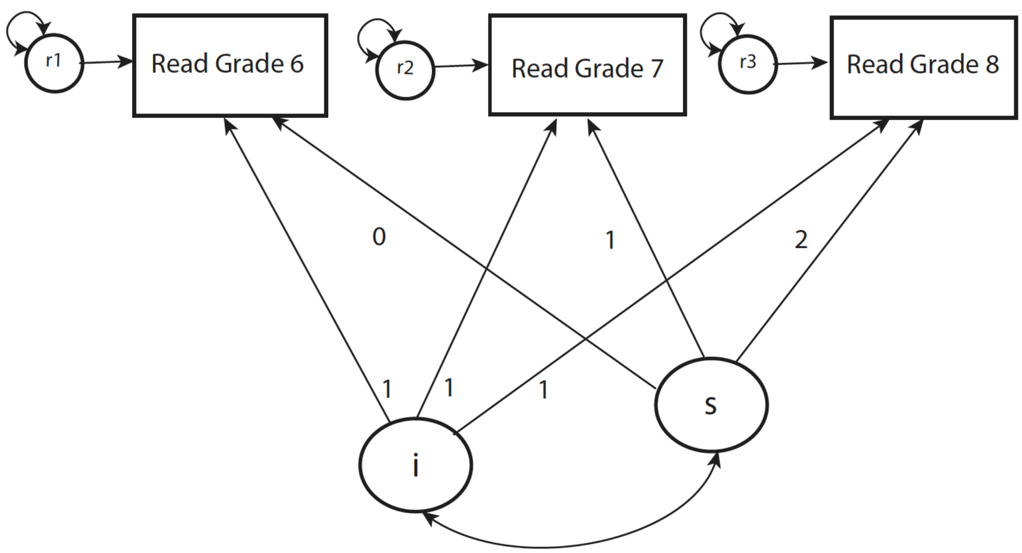

Figure 2. Unconditional linear latent growth curve model 1A of reading achievement across Grades 6 to 8.

Model 1. An unconditional LGCM was fit across the three time points (Model 1A). Factor relations with each measurement occasion were fixed at 1 for the latent intercept variable, and for the latent slope variable were fixed at 0, 1, and 2, for the three time points, respectively. The latent intercept and slope factors were also allowed to freely intercorrelate.A second model variation (1B) was estimated to allow the Grade 8 slope coefficient to be freely estimated (i.e., 0, 1, λ). A third model variation (1C) used a fixed, quadratic term to evaluate model fit (i.e., 0, 1, 4). Model fit and differences in explained variance were used to determine functional form for growth estimates (Kamata et al., 2013). After testing functional form, the influence of student demographics was estimated (Model 1D). Dummy-coded student demographic variables included: (a) sex (0 = male, 1 = female); (b) race (0 = White, 1 = non-White); and, (c) economic disadvantage (0 = non-FRL eligible, 1 = FRL eligible). Finally, a model was tested that added student exceptionality predictors (Model 1E) . In this model, dummy-coded student exceptionality categories (ID, CD, ED, OI, OHI, ASD, and SLD) were added as predictors with students identified with ID defined as the reference group. Reference group selections across all covariates were made for both theoretical and statistical reasons. Theoretically, ID students represent the targeted population for the AA-AAS. In addition, students with ID were also the largest exceptionality subgroup. The reference group for all other variables was also selected based on having the largest *n*-size.

Student demographic and exceptionality variables were included to determine if they were significant predictors of intercept (initial achievement) or slope (rate of growth) that improved model fit and thus should be carried forward in the model building process. Non-significant predictors of intercept or slope that did not improve model fit were dropped from subsequent models.

Finally, as shown in Appendix B, test-switching variables that represented each test switching pattern were evaluated. The test-switching variables represented participation in the general assessment (GA), the AA-AAS (AA), or that the data were missing (NA). There were 14 total test-switching patterns. Figure 3 displays Model 3, the fully conditional LGCM that incorporated all covariates, including student demographic predictors and exceptionality status.

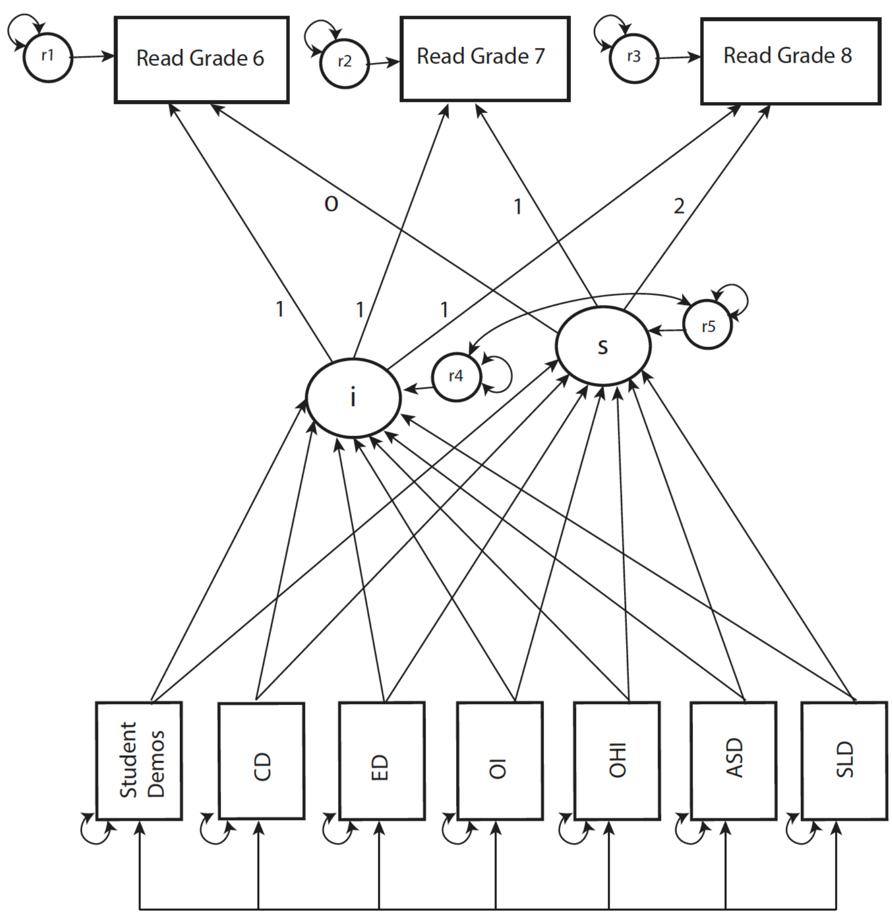

Figure 3. Linear latent growth curve model of reading achievement conditional on demographic and exceptionality status predictors.

Model 2. Following the Model 1 LGCM analyses, a LCGA was conducted to determine the number of classes representing groups of students with similar growth trajectories (Jung & Wickrama, 2008; Muthén & Muthén, 1998-2015). In the LCGA, the focus shifted away from a variable-centered approach to modeling growth toward a person-centered approach (Jung & Wickrama, 2008). Within-class intercept and slope variances were fixed at zero. The best LCGA model to use was determined by searching for the number of classes that had the lowest model BIC, and a significant Lo, Mendell, and Rubin Likelihood (LMR) *P*-value, beginning with a two class model and adding classes until model fit was not sufficiently improved (Bilir et al., 2008; Jung & Wickrama, 2008). A bootstrap likelihood ratio test (BLRT) was also conducted on the final model as a confirmatory step. However, it is important to note that all statistical evidence was moderated by consideration of the class *n*-size and substantive interpretability of the classes. AIC, BIC, and negative loglikelihood (LL) results were plotted and visual inspection was also used to help determine when an optimal number of classes had been reached (Jung & Wickrama, 2008).

Model 3. Following Model 2, an unconditional GMM was estimated. It should be noted that serious challenges in parameter estimation are often associated with mixture models (Bilir et al., 2008). Unlike the LCGA, where intercept and slope variance was fixed at zero, the unconditional GMM allowed growth parameters to vary across individuals within each class. Because the models were nested, with the LCGA model being a restricted form of the GMM, the unconditional LCGA was compared to the unconditional GMM using AIC and BIC. During the GMM phase, the model used for the LCGA was modified such that the intercept and slope random effects were estimated. GMM modeling commenced with the number of classes identified by the LCGA.

Fit statistics, including the AIC and BIC, as well as the deviance statistic, were used to identify the point where improvements in model fit were not occurring, namely where visual inspection of the relationships between AIC, BIC, and LL across the number of classes becomes flat or shifts to positive slope. The lowest achieved AIC/BIC and a significant LMR *P*-value were used to identify the optimal number of classes, with BIC privileged in situations where the criteria were not in agreement (Bilir et al., 2008; Jung & Wickrama, 2008). In addition, entropy (closest to 1.0), latent class counts and proportions (no less than 10% of the total), and average latent class probabilities (close to 1.0) were considered in the determination of the appropriate number of classes. A BLRT was conducted on the final model as a confirmatory step (Jung & Wickrama, 2008). The GMM equation model was defined as

$$y_{ti} \mid (C_i = c) = (\alpha_{c0} + \xi_{c0i}) + (\alpha_{c1} + \xi_{c1i}), \alpha_t + \varepsilon_{ic} \quad (2)$$

where the observed score of student *i* at time *t*, y_{ti}, given membership in class *c*, is modeled by their location on the latent intercept within each class, α_{c0}, and slope, α_{c1}, with corresponding error terms for student *i* in class *c*, ξ_{c0i} and ξ_{c1i}, respectively. The α_t term represents *time*, coded in years (0, 1, 2), and ε_{ic} represents unmodeled residual error for student *i* in class *c* (Bilir et al., 2008) as shown in Figure 4.

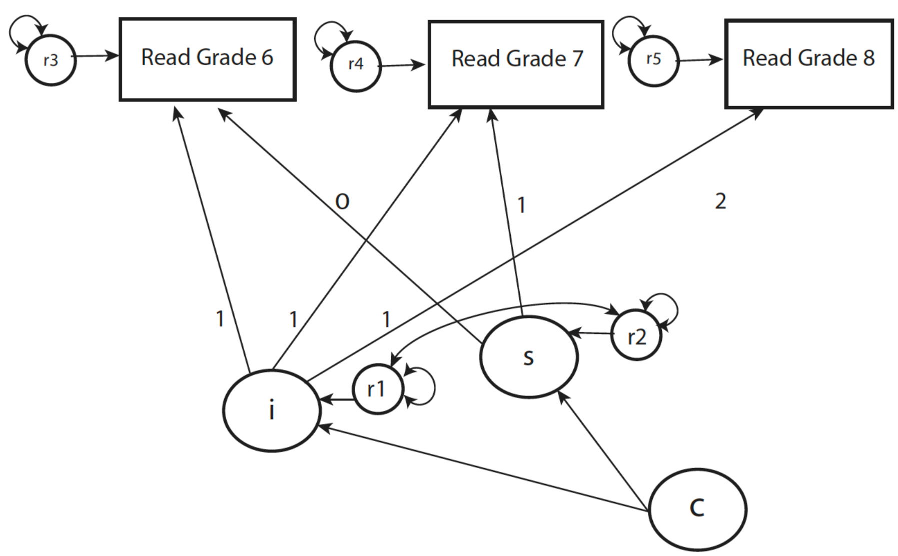

Figure 4. Unconditional growth mixture model of linear reading achievement.

Post Hoc Discriminant Function Analysis. In order to more fully interpret classes that might arise from LCGA or GMM analyses, an additional post hoc analysis was planned. Using the most probable, predicted class membership generated by Mplus, descriptive analyses of each class were conducted to better understand the composition of each class. Discriminant Function Analysis (DFA; Stevens, 2002) was used where class membership was predicted by student demographic variables. SPSS version 24.0 (IBM, 2016) was used for the DFA analyses. The same categorical variables were employed as described earlier: sex, race, and economic disadvantage.

CHAPTER III

RESULTS

The mean RIT scores, standard deviations, and *n*-sizes for the three time points were 101.32 (*SD* = 23.16; *n* = 1,387), 103.42 (*SD* = 23.24; *n* = 1,220), and 104.85 (*SD* = 23.04; *n* = 1,088), respectively. Observed score distributions and state proficiency cutoffs for accountability reporting are displayed in Figure 5. The distributions were bimodal,

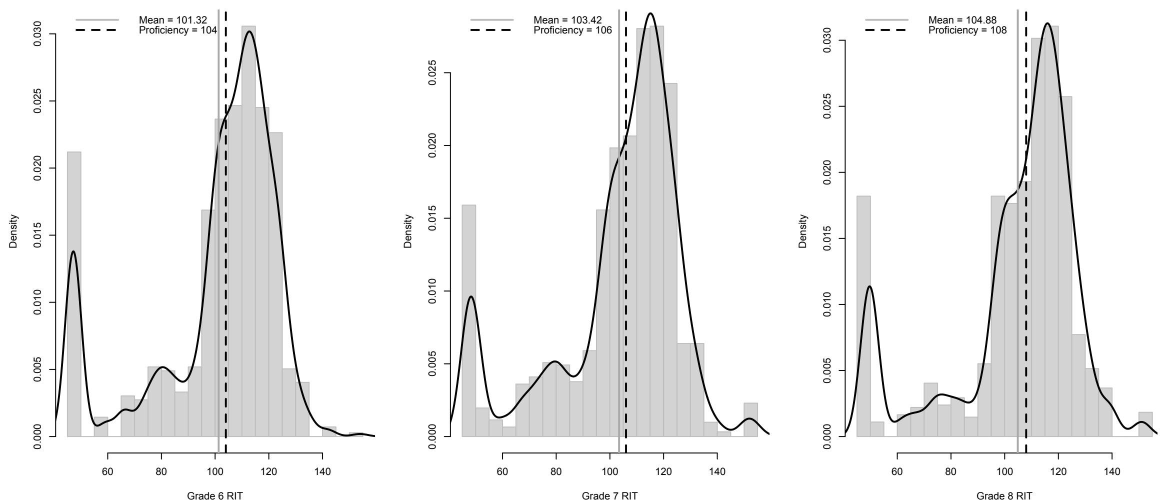

Figure 5. Grade 6 to 8 observed score means and score distributions, along with grade-level proficiency requirements.

and thus, exhibited substantial, negative skew (Grade 6 = -1.24; Grade 7 = -1.07; Grade 8 = -1.2). The Grade 6-8 distributions were leptokurtic, but leptokurtosis values were < 2.0 (Grade 6 = 0.68; Grade 7 = 0.62; Grade 8 = 0.85), and thus somewhat representative of a normal distribution, given the guidelines recommended by George and Mallery (2010). Full information maximum likelihood (FIML) estimation was used, as implemented in the Mplus software, version 7.3 (Muthén & Muthén, 1998-2015). The similarity of the distributions shown in Figure 5 indicates consistency in student performance with minimal differences across grades, and therefore, marginal growth.

Floor effect and ceiling effect subgroups were defined and evaluated descriptively. Floor effect descriptive statistics are presented in Table 1 and ceiling effect descriptive statistics are presented in Table 2. The floor effect subgroup included students whose RIT scores were two SD or more below the mean (RIT = 56.91) in Grade 6, 7, or 8. The ceiling effect subgroup was created similarly, with students who scored two *SD* or more above the mean (RIT = 149.5) in Grade 6, 7, or 8 included. Floor effect subgroup *n*-sizes and percentages, as well as overall sample group percentages, are provided above in Table 1. Percentages for the floor effect subgroup were compared to overall analytic sample percentages to determine whether subgroups deviated from population averages. Students who were economically disadvantaged (+ 7.23%), students with ID (+ 11.54%), OI (+ 9.10%), or ASD (+ 13.93%) were relatively over-represented in the floor effect subgroup. Students with SLD (- 22.97%) and CD (-6.82%) were under-represented in the floor effect subgroup. Ceiling effect subgroup *n*-sizes and percentages, as well as overall sample percentages, are provided below in Table 2. Students who were White (+ 26.77%) and those who were eligible within the SLD category (+ 17.66%) were over-represented in this highest-performing subgroup compared to overall analytic sample percentages. Students with ASD (- 15.06%) were under-represented in the ceiling effect group.

Model 1

The functional form of the data was evaluated to determine the best representation of reading growth across Grades 6 to 8 and determine whether reading growth trajectories varied by IDEA exceptionality. An unconditional linear growth curve model (LGCM) was fit across the three grades. The functional form of the data was evaluated in the first three models, including a linear model, a nonlinear model, and a fixed effects, quadratic model (Models 1A, 1B, and 1C, respectively).

Table 1

Floor Effect Descriptive Statistics

	Low Performers		Analytic Sample
Variable	*n*	%	%
Male	126	65.28	66.32
Female	67	34.72	33.68
White	133	68.91	60.73
Non-White	60	31.09	39.27
EconDis	99	51.30	30.27
Non-EconDis	94	48.70	69.73
ID	85	44.04	32.50
CD	0	0.00	6.82
ED	3	1.55	2.92
OI	25	12.95	3.85
OHI	14	7.25	10.67
ASD	64	33.16	19.23
SLD	2	1.04	24.01
Total	193	100.00	11.97

Note. ID = Intellectual Disability (reference group), CD = Communication Disorder, ED = Emotional Disturbance OI = Orthopedic Impairment, OHI = Other Health Impairment, ASD = Autism Spectrum Disorder, SLD = Specific Learning Disability, and EconDis = Economic Disadvantage.

Table 2

Ceiling Effect Descriptive Statistics

	High Performers		Analytic Sample
Variable	*n*	%	%
Male	17	70.83	66.32
Female	7	29.17	33.68
White	21	87.50	60.73
Non-White	3	12.50	39.27
EconDis	9	37.50	30.27
Non-EconDis	15	62.50	69.73
ID	8	33.33	32.50
CD	2	8.33	6.82
ED	0	0.00	2.92
OI	1	4.17	3.85
OHI	3	12.50	10.67
ASD	1	4.17	19.23
SLD	10	41.67	24.01
Total	24	100.00	11.97

Note. ID = Intellectual Disability (reference group), CD = Communication Disorder, ED = Emotional Disturbance OI = Orthopedic Impairment, OHI = Other Health Impairment, ASD = Autism Spectrum Disorder, SLD = Specific Learning Disability, and EconDis = Economic Disadvantage.

The quadratic model was restricted such that the quadratic parameter was not allowed to vary by student, as there were only three time points. The linear, nonlinear, and quadratic models had comparable AIC values (Model 1A = 30,773.940; Models 1B & 1C = 30,773.627), but BIC was lowest for the linear model (Model 1A = 30,817.022;

Models 1B & 1C = 30,822.094). The linear model also explained somewhat more intercept and slope variance compared to the nonlinear and quadratic models (Intercept = Model 1A: 438.16 σ^2; Model 1B: 445.93 σ^2; Model 1C: 441.25 σ^2; Slope = Model 1A: 438.16 σ^2; Model 1B: 445.93 σ^2; Model 1C: 441.25 σ^2). Deviance (*D*) testing between Models 1A and 1B, and between Models 1A and 1C did not result in statistically significant χ^2 difference tests for either comparison of models (*D* = 2.312; $p < .05$). It was therefore determined that the linear model, Model 1A, was most parsimonious and provided the closest approximation to the functional form of the data. The unconditional intercept for Model 1A was estimated at 102.99 RIT scale score points (*SD* = 0.57) with an annual growth estimate (slope) of 2.79 points (*SD* = .22). All subsequent modeling was conducted with the linear model. Complete fit statistics, intercepts, slopes, and variances for all variations of Model 1 are presented in Appendix C.

Model 1D added the student demographic predictors sex, race, and economic disadvantage. The reference group in Model 1D, White males who were not economically disadvantaged, had an estimated average intercept of 105.38 and an average slope of 3.32 RIT scale score units per grade. There were no statistically significant differences in initial achievement as a function of sex or race. However, students classified as economically disadvantaged had average initial achievement that was 4.28 scale score points lower than the reference group ($p < .05$). There were no statistically significant differences in average slope estimates between the reference group and students who were female, non-White, or economically disadvantaged ($p >$ .05). Fit statistics for Model 1D were within acceptable ranges (AIC = 30,722.47; BIC = 30,807.86; CFI = 0.999; SRMR = 0.005; RMSEA = 0.020). The chi-square model goodness-of-fit (GOF) test was not significant, χ^2 (4, N = 1,612) = 6.54, $p > .05$, indicating no significant discrepancy from a well-fitting model.

Model 1E added the student exceptionality categories (CD, ED, OI, OHI, ASD, and SLD). Because economic disadvantage was the only statistically significant predictor in Model 1D, it was the only demographic predictor retained in Model 1E. Thus, in Model 1E, the intercept represented the average Grade 6 achievement for White males with ID who were not economically disadvantaged. Intercept and slope parameter estimates from Model 1E are displayed in Table 3 below.

The average reading achievement in Grade 6 for the reference group was 97.40 scale score points. Average linear growth was 2.72 RIT scale score points per grade (t = 7.53, SE = 0.36, $p < .001$), on average. Model fit statistics showed that the LGCM Model 1E was the best fitting model tested (AIC = 30,382.15; BIC = 30,500.62, CFI = 0.999; SRMR = 0.004; RMSEA = 0.016). The chi-square model GOF test was not significant, χ^2 (8, N = 1,612) = 11.27, $p > .05$, indicating no significant discrepancy from a well-fitting model. All model fit statistics met the a priori fit criteria proposed by Hu and Bentler (1999). In addition, Model 1E had an AIC weight of 1.0 and a BIC weight of 1.0, demonstrating the best relative fit to the data among the variations of Model 1 that were tested (Burnham & Anderson, 2004). Model 1E accounted for 32.4% of the intercept variance and 3.3% of the slope variance relative to Model 1A.

Table 3

LGCM Model 1E Intercept and Slope Estimates with Standard Errors

Variable	Coefficient	*SE*	Slope	*SE*
Intercept	97.40*	0.36	2.72*	0.36
CD	17.02*	2.14	0.14	0.94
ED	18.56*	1.60	-0.85	1.60
OI	-13.54*	2.70	1.72	1.02
OHI	12.56*	1.81	-0.12	0.75
ASD	-0.98	1.45	0.10	0.55
SLD	19.10*	1.39	1.19	0.64
EconDis	-4.28*	0.46	-0.33	0.46

Note. CD = Communication Disorder, ED = Emotional Disturbance OI = Orthopedic Impairment, OHI = Other Health Impairment, ASD = Autism Spectrum Disorder, SLD = Specific Learning Disability, and EconDis = Economic Disadvantage.
* $p < .001$

Statistically significant differences in Grade 6 initial achievement were found between the reference group (White, non-FRL males identified with ID) and all other groups except for students with ASD. This included students from economically disadvantaged backgrounds who performed significantly lower than the reference group in initial achievement (-4.28, $p < .001$). However, no statistically significant slope differences were found between the reference group and the other predictors. Students with SLD had the highest initial achievement (+19.10, $p < .001$), while students with OI had the lowest initial achievement (-13.54, $p < .001$). Students with OI demonstrated the largest average annual growth (+1.72) and students with ED demonstrated the lowest average annual growth (-0.85). Figure 6 displays the observed means and estimated growth trajectories from Model 1E. The unconditional mean is provided for reference.

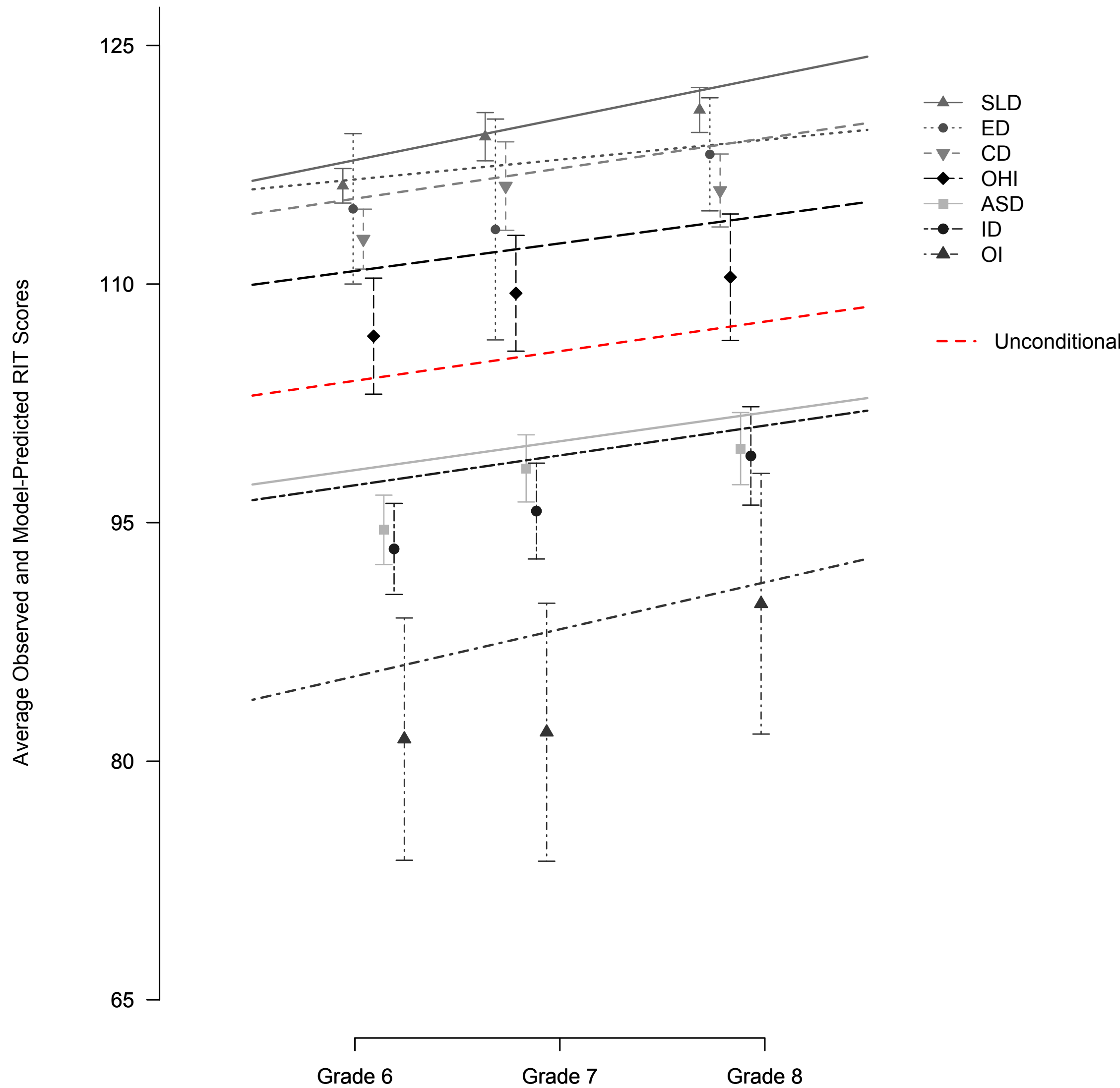

Figure 6. Average observed means and average Model 1E estimated means. Means given by grade for students in each exceptionality category, with an unconditional results reference line. ID = Intellectual Disability (reference group), CD = Communication Disorder, ED = Emotional Disturbance OI = Orthopedic Impairment, OHI = Other Health Impairment, ASD = Autism Spectrum Disorder, SLD = Specific Learning Disability, and EconDis = Economic Disadvantage

A 95% confidence interval is shown in Figure 6 for each observed mean. Though a linear model fit the data best, the observed means show that there was some curvilinearity in the data, which varied among subgroups. For example, students identified as OHI grew more from Grade 6 to 7 than they did from Grade 7 to Grade 8. Effect sizes for all student subgroups for each adjacent grade transition are provided in Table 4.

Table 4

Observed Effect Sizes for Student Demographic and Exceptionality Groups for Each Grade Transition

Variable	Grade 6 to 7	Grade 7 to 8	Grade 6 to 8
Male	0.11	0.06	0.17
Female	0.05	0.08	0.12
White	0.12	0.06	0.18
Non-White	0.04	0.08	0.12
Non-EconDis	0.11	0.02	0.13
EconDis	0.09	0.08	0.17
Cohort 1	0.07	0.11	0.18
Cohort 2	0.11	0.02	0.13
ID	0.17	0.05	0.22
CD	0.32	- 0.03	0.34
ED	- 0.08	0.35	0.31
OI	0.02	0.28	0.29
OHI	0.13	0.05	0.17
ASD	0.10	0.14	0.24
SLD	0.29	0.16	0.51
Total	0.09	0.06	0.15

Note. ID = Intellectual Disability, CD = Communication Disorder, ED = Emotional Disturbance OI = Orthopedic Impairment, OHI = Other Health Impairment, ASD = Autism Spectrum Disorder, SLD = Specific Learning Disability, and EconDis = Economic Disadvantage.

The magnitude of observed growth was calculated as the mean difference in achievement between each adjacent grade divided by the pooled standard deviation across the two adjacent grades (Bloom et al., 2008). As shown in Table 4, students with SLD had the largest growth compared to all other subgroups, growing a little more than one-half of a standard deviation unit (0.51). Students with OHI grew the least across the three-year period, growing approximately one-sixth of a standard deviation unit (0.17). Inconsistent growth patterns were present. Students with ID, CD, OHI, and SLD grew more from Grades 6 to 7 compared to growth between Grades 7 to 8. Students with OI, CD, and ED exhibited the opposite growth pattern, with less growth between Grades 6 and 7, and greater growth from Grades 7 to 8. However, modeling observed curvilinearity did not improve model fit. Figure 7 conveys effect size gaps that exist based on student exceptionality, with ID as the reference group depicted as the diagonal reference line.

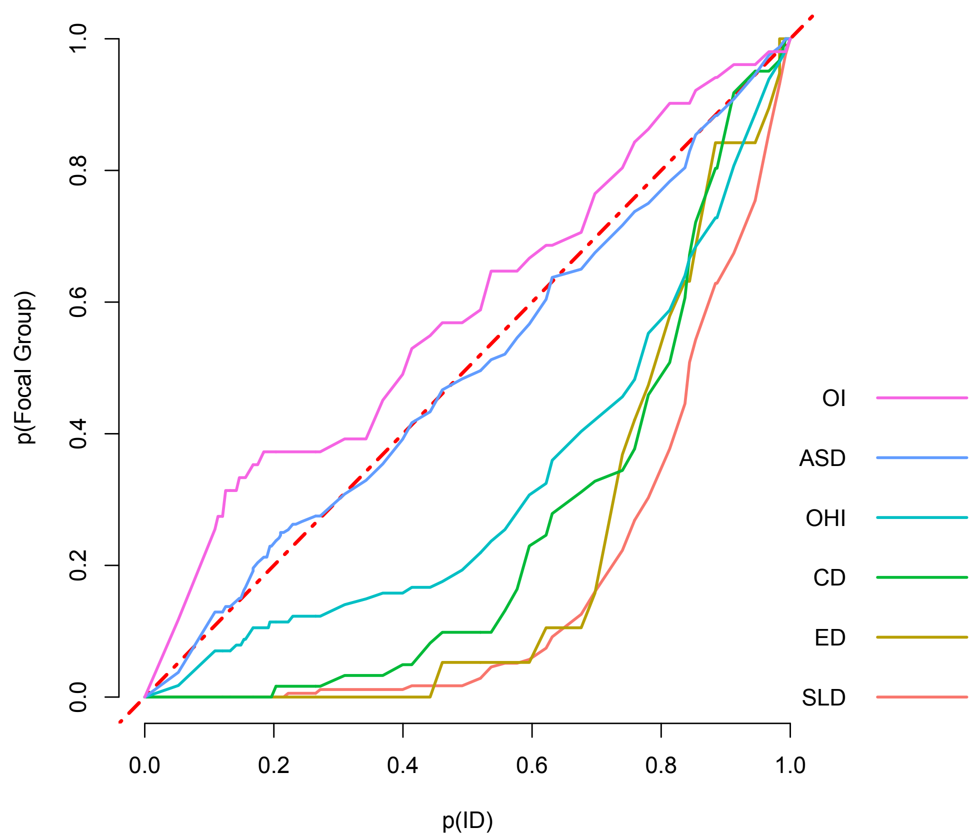

Figure 7. Effect size gaps in Grade 8. Gaps shown between student exceptionality groups with students identified as ID as the reference group. ID = Intellectual Disability, CD = Communication Disorder, ED = Emotional Disturbance OI = Orthopedic Impairment, OHI = Other Health Impairment, ASD = Autism Spectrum Disorder, SLD = Specific Learning Disability.

Figure 7 shows how the effect sizes between student exeptionality groups change across the scale for Grade 8. Results for Grades 6 and 7 were quite similar (see Appendix D). Effect size gaps across grades and scale were rather stable, with consistent differences from the ID reference diagonal. Students identified as ID or ASD performed similarly; students identified as OI performed at lower levels. Students identified as SLD generally outperformed all others, with students identified with ED, CD, or OHI ranked lower.

Models 2 and 3

Next, a series of unconditional LCGA analyses were conducted. LCGA modeling added a latent class variable to the unconditional LGCM, with within-class intercept and slope variances fixed at zero. The first LCGA analysis included one class and subsequent analyses added an additional latent class one at a time, up to seven classes in this analysis. The models were evaluated using the a priori decision rules recommended by Jung and Wickrama (2008), including visual analysis of the score distributions (see Figure 2), visual analysis of AIC and BIC (see Figure 8 below) and LL (see Figure 9). Figures 7 and 8 display the AIC and BIC results, and the LL GOF values, respectively, for each LCGA model with from one to seven classes. LCGA classes were also evaluated with theoretical and practical justifications, which included a review of score distributions, effect size differences, and a practical understanding of the context.

The two-class solution resulted in a relatively steep reduction in AIC (30,799.30) and BIC (30,842.38), statistically significant LMR-LRT ($p < .001$) and BLRT values ($p < .001$), and exhibited entropy results very close to 1.0 (0.979). In addition, subgroups from the two-class solution were each larger than 10% of the sample and had high average class membership probabilities (0.982 for Class 1; 0.997 for Class 2). The two-class LCGA solution also resulted in substantial reductions in unexplained variance compared to the three-class solution, explaining about 72% of the residual variance at Grade 6, 67% of the residual variance at Grade 7, and 66% of the residual variance at Grade 8. Though the three-class solution met the a priori statistical rules of thumb provided by Jung and Wickrama (2008), I decided to pursue a two-class solution based on a review of the entire body of evidence. The two-class intercept and slope results were more consistent with score distributions. The observed patterns in the effect size gaps, where one higher-performing group of students identified with SLD, ED, and CD were consistently contrasted with a lower-performing group of students identified with ID, ASD, and OI, supported this determination. The planned GMM would also not be limited to the number of classes determined by the LCGA, so I determined to err on the side of parsimony and pursue a two-class solution. Additional LCGA results are presented in Appendix E.

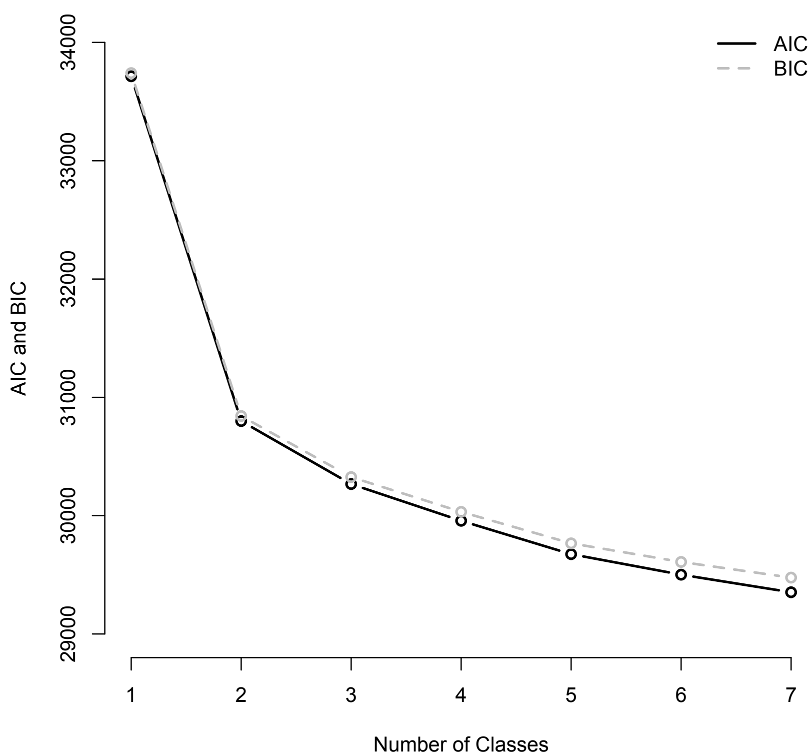

Figure 8. AIC and BIC values as a function of the number of latent classes.

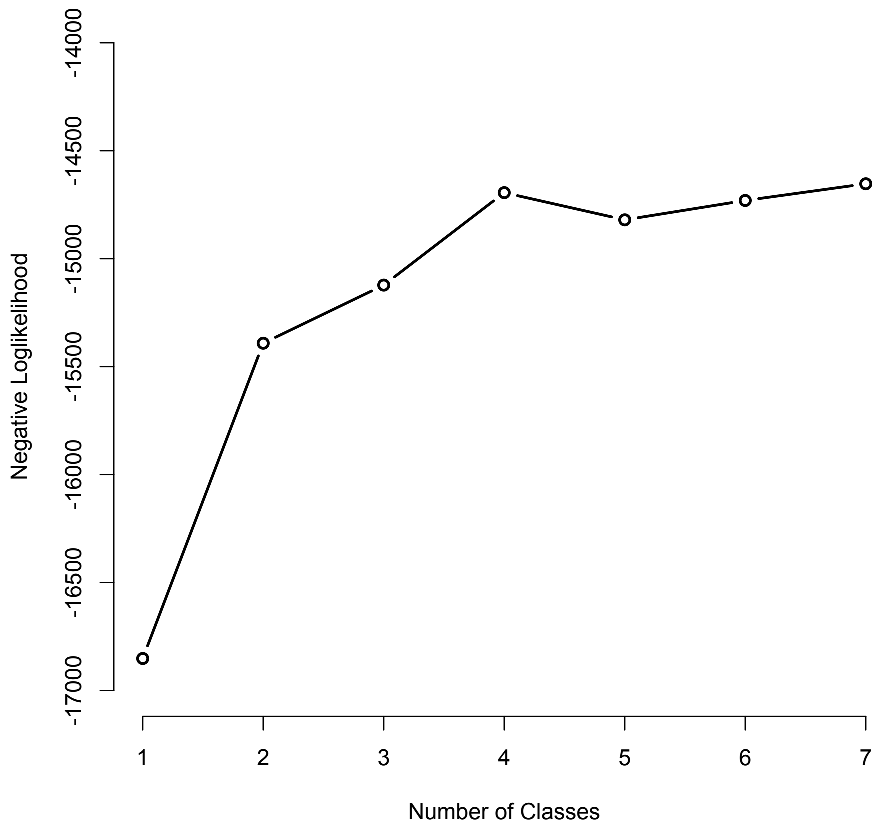

Figure 9. Loglikelihood values as a function of the number of latent classes.

The two-class GMM showed that Class 1 (n = 340) had an initial achievement of 75.33 (t = 31.73, SE = 2.37, p < .001) and grew 4.42 points per year on average (t = 6.72, SE = 0.66, p < .001). Class 2 (n = 1,272) had an initial achievement of 112.04 (t = 360.82, SE = 0.31, p < .001) and grew 2.22 points per year on average (t = 13.68, SE = 0.16, p < .001). The Class 1 versus Class 2 intercept estimates were significantly different (p < .01), as were the slope estimates (p < .01).

A three-class GMM did not converge. However, because the slope variance estimates for Class 2 were so small, an effort was made to conduct a three-class GMM with the slope variance for Class 2 set to zero—this model did not converge. In addition, the lowest-performing students (> 2.0 *SD* below the mean) were removed from the dataset to determine whether their results were impacting the GMM convergence.

However, the three-class GMM with the lowest-performing students removed also did not converge. No attempts were made to fit GMMs with additional latent classes.

The two-class GMM solution had lower AIC (29,400.40) and BIC (29,470.41) values, compared to the LCGA two-class model, indicating that allowing individual growth parameters to vary within each class resulted in better model fit. Model fit statistics for the two-class GMM are presented in Table 5. The two-class GMM estimates explained approximately 85% of the intercept variance and 72% of the slope variance in comparison to the unconditional one-class model.

Table 5

LCGA Two-Class Model Fit and GMM Class Sizes and Model Fit Results

Model	*n*	%	*Int* σ^2	*S* σ^2	*AIC*	*BIC*	*LMR-LRT*	*BLRT*	*Entr*
LCGA	-	-	-	-	30,799.30	30,842.38	<.001	<.001	0.98
One-Class	1,612	100.0	438.21	14.21	30,773.94	30,817.02	-	-	-
Two-Class	-	-	-	-	29,400.40	29,470.41	<.001	<.001	0.74
C(1)	340	21.1	673.49	132.85	-	-	-	-	-
C(2)	1,272	78.9	66.42	3.96	-	-	-	-	-

Post Hoc Discriminant Function Analysis

Finally, discriminant function analysis (DFA) was used to provide further interpretation of the composition of the classes estimated in Model 3, the two-class GMM. The DFA used student's sex, race, economic disadvantage, and exceptionality group to predict students' class membership as predicted by the two-class GMM. Box's M (1,252.53, $p < .001$) indicated that the assumption of equality of covariance matrices was violated; however, this problem is likely influenced by the large sample size and would not have any substantial impact on the estimation of the discriminant coefficients, the primary focus of the analysis. With only two classes, only one discriminant function was estimated. DFA structure coefficients are provided in Table 6.

Table 6

DFA Structure Coefficients

Variable	*Structure Coefficients*
SLD	0.669
ASD	- 0.453
OI	- 0.358
CD	0.342
ED	0.142
OHI	0.135
EconDis	0.133
Sex	- 0.115
Race	0.070

Note. ID = Intellectual Disability, CD = Communication Disorder, ED = Emotional Disturbance OI = Orthopedic Impairment, OHI = Other Health Impairment, ASD = Autism Spectrum Disorder, SLD = Specific Learning Disability, and EconDis = Economic Disavantage

There was a statistically significant association between classes and exceptionality predictors, accounting for 12.5% of between class variability. Inspection of the structure matrix (see Table 6) revealed one large, three moderate, four small, and one small relation between student characteristics and the discriminant function based on Cohen's correlation coefficient rules of thumb (1988). Students identified with SLD had a large, positive relation with the discriminant function (.669). Students with ASD (-.453) and OI (-.358) had moderate, negative relations. Students with CD had a moderate, positive relation (.342). Students identified with ED, OHI, and economically disadvantaged and male students had small relations, with a negative male relation. The relation between the discriminant function and White was small and positive.

Table 7

Descriptive Statistics for GMM Class Membership

	Class 1		Class 2	
Variable	*n*	*% Within Class*	*n*	*% Within Class*
Male	212	62.35	857	67.37
Female	128	37.65	415	32.63
White	215	63.24	764	60.06
Non-White	125	36.76	508	39.94
Non-EconDis	222	65.29	902	70.91
EconDis	118	34.71	370	29.09
ID	156	45.88	368	28.93
CD	2	0.59	108	8.49
ED	4	1.18	43	3.38
OI	30	8.82	32	2.52
OHI	26	7.65	146	11.48
ASD	109	32.06	201	15.80
SLD	13	3.82	374	29.40
Total	340	21.09	1,272	78.91

Note. ID = Intellectual Disability, CD = Communication Disorder, ED = Emotional Disturbance OI = Orthopedic Impairment, OHI = Other Health Impairment, ASD = Autism Spectrum Disorder, SLD = Specific Learning Disability, and EconDis = Economic Disadvantage

Inspection of relative class *n*-sizes shows that students identified with CD and SLD were highly likely to be members of the higher-performing Class 2. Students with ED were ten times as likely to be in Class 2. Students with OHI were seven times as likely to be in Class 2. Students with ID and ASD were twice as likely to be assigned to Class 2. Students with OI had balanced membership in Class 1 and Class 2.

Comparing actual class membership with class membership as predicted by the DFA analysis showed that class membership was correctly classified for 78.8% of the students. Descriptive results matched the DFA analysis as well, supporting the finding that there were substantive class differences based on SLD, ASD, OI, and CD group membership, while differences between the other student subgroups by class were small. Descriptive statistics for GMM class membership are provided in Table 7 for comparison. There were higher relative percentages of students identified as ID (+ 17%), ASD (+ 16%), and OI (+ 6%) in Class 1 compared to Class 2, but lower relative percentages of students identified with SLD (- 26%) and CD (- 8%) when comparing Class 1 membership to Class 2 membership.

Modeling Summary

A series of LCGA models showed that the most parsimonious description of SWSCD academic growth in reading was a linear model over Grades 6 to 8. There were statistically significant differences in initial achievement between students with economic disadvantage, as well as students in each exceptionality subgroups except ASD in comparison to the reference group (White male students who were not economically disadvantaged and were identified as ID). The LGCM results showed no statistically significant differences between subgroups in growth rate over grades. In Model 2, an

LCGA was conducted to help discover whether there were homogeneous latent growth classes within the data. Two classes were identified, the first with higher average intercept and slower average growth and the second class with lower average intercept and faster average growth. In Model 3, a two-class GMM that allowed individual growth trajectories to vary within class resulted in similar results. Post hoc analyses were then conducted using DFA. The DFA and descriptive analyses showed that, in comparison to Class 2, Class 1 was composed of substantially more students identified as ID, OI, or ASD, and substantially fewer students identified as CD, OHI, or SLD.

CHAPTER IV

DISCUSSION

There are few published studies on the academic growth of SWSCD in reading, likely due in part to challenges related to test scaling, group heterogeneity, small sample sizes, missing data, and the use of status-based assessments that were not designed to measure a developmental continuum. The results of my study add to a small but emerging body of literature that focuses upon modeling academic growth for SWSCD (Dunn et al., 2012; Farley et al., 2016; Karvonen et al., 2013; Tindal et al., 2015).

Growth Modeling Challenges for SWSCD

Test scaling issues in this study were less of a challenge than growth research on most other alternate assessments, including Dunn et al. (2012) and Karvonen et al. (2013), because the OR AA-AAS was created using IRT scaling that expressed scores on a common interval scale. Students were administered assessments linked to a common scale from Grades 6 to 8, which provided a unique opportunity to review growth compared to most AA-AAS assessment results. Test scaling nonetheless remains an area of concern when modeling growth in general, and for SWSCD in particular (Dunn et al., 2012; Farley et al., 2016; Karvonen et al., 2013). It is unclear whether the items included on the OR AA-AAS, or any AA-AAS, are sufficiently sensitive to measure growth for SWD or that items are tied to a scale that increases incrementally in a manner that matches natural development of reading skills and knowledge.

Group heterogeneity in the current study was addressed in several ways. First, LGCM modeling included three student demographic categories (sex, race, and economic disadvantage) and seven exceptionality categories as covariates (ID, CD, ED, OI, OHI, ASD, and SLD), which allowed for comparisons of subgroups within the SWD population. Heterogeneity was also addressed through the use of LCGA and GMM models that explored latent classes of students with similar growth trajectories.

Small sample sizes limited the inclusion of students with low incidence disabilities in this study (VI, HI, DB, and TBI), as *n*-sizes within these exceptionality groups were not sufficient to provide reliable estimates. Although joining these students into a low-incidence group so that they contributes to grand mean estimation, similar to Farley et al. (2016), was feasible, such inclusion did not support the within-group disaggregation and comparisons that were a goal of this study. I was not able to evaluate the performance of students in these exceptionality groups even though two cohorts were used, and future studies might profit from the inclusion of additional cohorts.

SWD Reading Achievement and Growth

Previous research examining the relation of student demographics with the reading achievement of SWD has found a number of significant relations. For example, Morgan et al. (2011) found that kindergarten students identified with SLD or CD from economically disadvantaged backgrounds, African American and Hispanic subgroups, and females exhibited significantly lower initial achievement on the ECLS-K reading test compared to their respective reference groups. Schulte et al. (2016) found significant race/ethnicity initial achievement differences in reading comprehension based on a state achievement test between specific SWD exceptionality groups, and between race/ethnicity groups with Asian students performing significantly higher than White students, and students from American Indian, Black, Hispanic, and multiracial backgrounds performing lower than the reference group.

In studies that compare SWD to general education (GE) peers, SWD performed significantly lower than GE peers in initial achievement (Blackorby et al., 2015; Morgan et al., 2011; Schulte et al., 2016). SWD in one study demonstrated more rapid growth compared to GE students, yet this was likely attributed to the skills assessed at their level of function, and growth decelerated over time (Schulte et al., 2016). The SWD growth rates were not sufficient to close achievement gaps between SWD and GE students, however (Schulte et al., 2011) and one study found the achievement gap between students identified with SLD and GE students to be widening (Judge & Bell, 2010).

A small number of studies on the achievement of SWD have also demonstrated that there are important differences between specific exceptionality subgroups within the general SWD classification. In these studies, students with SLD, CD, and VI achieved at the highest levels in reading relative to other SWD peers and students with MD and ID achieve at the lowest levels relative to their SWD peers (Blackorby et al., 2005; Morgan et al., 2011; Schulte et al., 2016; Wei et al., 2011). Growth also decelerated across grades (Blackorby et al., 2005; Schulte et al., 2016; Wei et al., 2011).

In my study, I found significant within-group achievement differences based upon exceptionality status, consistent with other findings in the research on SWD reading achievement (Blackorby et al., 2005; Morgan et al., 2011; Schulte et al., 2016; Wei et al., 2011). The relative rank of reading achievement performance based on student exceptionality classification showed that students with SLD and CD exhibited higher levels of reading achievement, while students with ID had the lowest levels of reading achievement. However, OR had several atypical approaches to IDEA eligibility determination that affected eligibility classifications that must be considered. First, OR did not have a Multiple Disabilities category. Instead, IEP teams list a student's primary, secondary, and tertiary eligibilities, but these additional classifications are not included in the OR database. This practice also complicates researcher understanding of other exceptionality categories, as the primary eligibility may or may not tell a full story of the student's needs. For example, membership in the OI category as a primary eligibility does not explain why students performed well below their SWSCD peers, as the OI exceptionality does not include a cognitive disability. It is likely that students with OI have serious concomitant secondary and tertiary disabilities that may impact reading achievement. This area is worthy of further study.

Research demonstrates that growth is usually curvilinear with deceleration occurring in late elementary or middle school (Blackorby et al., 2005; Schulte et al., 2016; Wei et al., 2011). Because only three years of data and a linear growth model were used here, deceleration of growth could not be evaluated effectively. However, I observed curvilinearity among some exceptionality subgroups in the observed means, as shown in Table 4 and Figure 6. Students with ID, CD, OHI, and SLD grew more from Grades 6 to 7 compared to growth rates between Grades 7 to 8. Students with OI, CD, and ED exhibited the opposite growth pattern, with less growth between Grades 6 and 7, and greater growth from Grades 7 to 8. Although the majority of students in the dataset exhibited curvilinear growth patterns, the conflict of accelerating and decelerating growth for different groups may have resulted in linear models being the best fit to the data overall and attention to the functional form of growth is a worthy object of future study.

The initial achievement differences found in my study of SWSCD reading achievement appeared to be consistent with other literature examining the growth of

SWD on general assessments. My results, though using a different comparison group, are consistent with those found by Schulte et al. (2016), who found significant differences in initial achievement but generally similar growth rates across grades comparing SWD to GE. However, my findings diverge from Judge and Bell (2010) who found that students identified with SLD fell further and further behind GE peers over time.

SWSCD Reading Achievement and Growth

Current research regarding the functional form of growth for SWSCD is inconclusive, though the majority of the SWD research suggests curvilinearity (Schulte et al., 2016; Wei et al., 2011). Tindal et al. (2015) modeled reading growth in Grades 3 to 5 with a linear functional form using OR AA-AAS data from 2008 to 2010. The functional form of growth was discussed as an advantage of HLM, but was not explicitly evaluated in the Tindal et al. (2015) study. The Farley et al. (2016) study evaluated the functional form of growth as a first step and determined that a nonlinear functional form best fit the data in reading over Grades 3 to 5. Slope coefficients in the Farley et al. (2016) study were determined to be 0-1-1.35 for the final LGCM, demonstrating that substantially more growth occurred between Grades 3 to 4 than from Grade 4 to 5. Addressing the functional form of growth is important for accurate modeling, but is also related to accountability considerations. For example, more growth is expected at earlier grades than at later grades given these results.

Because effect sizes are standardized, they are not scale dependent. Effect sizes thus provide a basis for comparison of results across studies. Farley et al. (2016) found average observed reading gain effect sizes were 0.19 between Grades 3 and 5. In comparison, the current study found that SWSCD reading growth effect sizes were 0.15

from Grades 6 to 8. It is likely that a linear functional form was a sufficient approximation of growth for Grades 6 to 8 in this study because there were smaller magnitudes of growth over the Grade 6 to 8 period compared to Grades 3 to 5. Though not an ideal comparison due to the fact that the researchers used nationally-standardized assessments and the results were gathered from general education peers, comparisons to Bloom et al. (2008) indicate that SWSCD are growing less than their GE peers during the Grade 6 to 8 time span.

In addition, Farley et al. (2016) found that all student exceptionality groups included in the study (SLD, CD, ASD, OHI, SLD, OI, ED, and Low Incidence) exhibited larger effect size gains between Grades 3 and 4 than they did between Grades 4 and 5. My dissertation study found that students identified with ID, CD, OHI, and SLD exhibited larger effect size gains between Grades 6 to 7, but students identified with ED, OI, and ASD grew more from Grade 7 to 8. While the reasons for these performance differences are not possible to explain with direct evidence, several factors may have affected the achievement of the atypical group that grew more from Grade 7 to 8, including teacher quality and switching schools/programs between those grades.

Final LGCM results from Model 1 in my study showed that students identified with SLD, ED, CD, and OHI performed significantly higher than the reference group, White students identified with ID from non-economically disadvantaged backgrounds. Students identified with OI and economic disadvantage performed significantly lower than the reference group. Students identified with ASD were statistically indistinguishable from the reference group. This pattern of performance matches the results from Farley et al. (2016), indicating the same relative ranking of exceptionality groups. These results suggest that any high stakes accountability use of these test results should consider IDEA exceptionality status, consistent with the recommendations of Schulte et al. (2016) and Wei et al. (2011).

There is no published literature that addresses latent classes of SWSCD reading achievement growth. This is a new area of research for this population. Two latent classes were identified in the LCGA, one composed of lower-performing students with moderate growth rates, and one composed of higher-performing students with slower growth rates. Subsequent GMM analyses yielded a two-class solution, as well. In the final GMM model, there were two classes, with Class 1 composed of the lower-performing group that exhibited higher growth and Class 2 composed of the higher-performing group that exhibited less growth. However, the GMM intercept and slope variances and visual inspection of the growth trajectories demonstrated that within Class 1, members exhibited a high degree of variability around both intercept and slope estimates. One subgroup within Class 1 exhibited flatter trajectories, whereas others exhibited extremely accelerated or decelerated growth trajectories. To ensure that the subgroup of students in Class 1 with extreme growth trajectories was not impeding GMM estimation with a three-class solution, students in Class 1 with slopes beyond 3.0 SD above or below the mean were removed from the dataset and a three-class solution was attempted. The three-class solution still did not converge.

Finally, class membership was strongly related to student exceptionality, but not to student demographic characteristics. DFA and descriptive analyses of predicted GMM class membership showed that student exceptionality had a substantial relation with class membership, specifically for students identified as SLD, CD, and ED, who were higher performing, and students identified as ASD and OI, who were lower performing. Student sex, race, and economic disadvantage were not significantly associated with class membership. These results provide some evidence that the severity of the student's cognitive disability is associated with reading performance, whereas student demographics did not have strong associations.

Limitations

Some important limitations in this study were lack of information about whether students had the opportunity to learn the tested content, a study design that cannot support causal claims, and a status-based reading assessment in Grades 6 to 8 that may be insufficient to model growth. My results also rest on my model specifications, the constitution of the student population included in the study, small sample sizes, and the manner in which missing data were addressed.

Opportunity to learn. I have no evidence that the SWSCD involved in this study had a sufficient opportunity to learn the tested reading content (Elliott, 2015). I also lack evidence regarding the quality and quantity of instruction that SWSCD involved in this study received. Though the field is shifting to incorporate more academic content into the curriculum for SWSCD, particularly in the area of reading (Pennington & Courtade, 2015), opportunity to learn academic content cannot be assumed (Kleinert, et al., 2015).

Non-experimental design. This study used a non-experimental design and analysis of extant, longitudinal reading achievement data. The study was descriptive and exploratory in nature. There are thus no causal claims to be made from the study and the study only provides a description of the relation between predictor variables and the reading growth of SWSCD.

Status-based operational assessment. Results of the current study also depended on the design and operational administration of the OR reading AA-AAS. The OR AA-AAS differs substantially compared to many other state alternative assessment systems, particularly in its use of a common IRT scale and an interval level score scale. Most state alternate assessments also report student performance only using a limited number of proficiency categories in contrast to the OR AA-AAS use of an interval scale. An interval scale allows the kinds of analyses used in this study. An IRT, vertical linked scale of itself is insufficient to adequately model growth, however, as the scale also requires connection and representation of the developmental continuum of reading achievement the scale is intended to measure.

Model specification. The results of this study are limited to the way that I specified the statistical models. The results depend on the use of linear models, and may have differed if different predictor variables were included in the models. For example, one omitted variable in this study, attendance data, might have provided a proxy for students' opportunity to learn. The final GMM results may not be reliable, as well. The GMM may not have been sufficiently sensitive to variability in student performance and thus additional classes may have been present that were not identified by the GMM. The relatively small amount of growth for SWSCD in reading in my study may also present challenges for implementing GMMs.

Analytic sample. There are also important considerations related to the analytic sample. I had no control over the student population that participated in the OR AA-AAS in the study or whether determination of IDEA exceptionality status was consistent with legal requirements and best practice. Though students' most frequent exceptionality across Grades 6 to 8 was used in 97% or more of the cases in my study, eligibility classifications can change from year to year and evidence suggests that different methods for accounting for this variability may impact study results (Schulte & Stevens, 2013). However, analyses demonstrated that IDEA exceptionality classifications in this study remained quite stable for SWSCD and were likely not as concerning an issue compared to the special education population as a whole. This study had access only to primary exceptionality information, though it is common for SWSCD to have additional secondary exceptionality classifications documented in student Individualized Education Programs (IEPs). Related to this concern is that OR did not have a multiple disabilities (MD) category, which could have accounted for some of the unknown disability severity.

Sample exclusions. The study excluded student exceptionality subgroups with fewer than 42 students. As a result, inclusion of students with low-incidence exceptionalities (students who are Deaf/Hard of Hearing or Deaf-Blind, students with Visual Impairments or Traumatic Brain Injuries) was not possible and results here may not generalize to other states with different subgroups or additional students. The addition of a second cohort allowed for modeling two additional disability subgroups (ED and OI) compared to Farley at al., (2016), however, and the use of multiple cohorts should be considered, where feasible, as a tactic for increasing subgroup size.

Missingness. Finally, data were evaluated with Little's MCAR test and it was determined that the data were not MCAR (Little & Rubin, 2002) Missingness rates ranged from 14% in Grade 6 to 33% in Grade 8, suggesting that modeling missingness may be an important procedural step (Farley et al., 2016). However, including test-switching patterns that included known and unknown missingness mechanisms did not improve model fit in the LGCM. Pursuant to this finding, missing data were handled with FIML, which may not have resulted in unbiased estimates.

Conclusion and Future Direction

In the LGCM portion of this study, SWSCD with different exceptionalities generally had significantly different average initial achievement but growth rates that did not differ significantly from each other. Based on the LGCM, SWSCD classified as economically disadvantaged in my study performed significantly lower than their peers in initial achievement, yet exhibited growth rates that were not statistically different than the reference group. Using LCGA and GMM analyses, this study also found evidence for two separate latent classes of students with exceptionalities on the AA-AAS. The first class had lower achievement and larger growth rates, while the second class had higher achievement and slower growth rates. Students identified as SLD and CD were generally higher-performing, while students identified as ID, ASD, and OI were lower performing across all analytic models. These patterns were generally consistent with prior research, but would benefit from replication on other state alternative assessments.

Two latent classes, most notably the class composed of higher-achieving students, accounted for the GMM results showed that variance in achievement and growth estimates for SWSCD. A substantial degree of achievement and growth variance existed within the lower-performing class. The higher-performing class likely exhibited more consistent academic behaviors due to generally higher levels of functioning. The lower performing class, alternatively, exhibited less predictable behaviors.

The performance of students with SLD, CD, and ED, and results from similar student populations in other states (Kearns et al., 2011), suggest that this state should likely reconsider its AA-AAS eligibility criteria to be more stringent, as the population of

SWSCD who participated in the OR AA-AAS did not match those from other states based on primary eligibility classifications (Quenemoen, 2008). Though there is no AA-AAS eligibility restriction based on IDEA exceptionality category, students who participate in AA-AAS based on national data typically are identified in the MD, ID, and ASD categories. The frequent participation of OR students who are SLD, CD, and ED is not common and contrary to some states' participation rules. Some of this discrepancy can be explained by the fact that only primary disability information was available in OR and the students served are likely more complex than suggested by a primary eligibility classification. In addition, OR does not have a MD eligibility category, which impacts IDEA eligibility profiles. Finally, it is possible that students who do not have "truly" significant cognitive disabilities participated in the OR AA-AAS during this time period.

These results also suggest that state education agencies (SEAs) may need to develop AA-AAS test designs that accommodate the varied student population of SWSCD who participate in these assessment systems. For example, SEAs might consider developing two assessments, one that targets students performing at the lower end of the academic continuum and another targeting those performing at higher levels. Given the wide differences in achievement between the two classes identified in the latent class models, it is unlikely that a singular approach to assessment could meet the academic assessment needs of both groups without unnecessarily expending valuable instructional time on items that are either too easy or too difficult.

This study provided information for policymakers to consider within accountability contexts for understanding growth for SWSCD. The results of both the conditional LGCM and the GMM demonstrated that the severity of the student's exceptionality was related to initial Grade 6 reading achievement, but not reading growth. These findings may be influenced by multiple factors, including measure and scale sensitivity, opportunity to learn, and day-to-day fluctuations in behavior in this populations (Ysseldyke & Olsen, 1999). The measures and scales employed may not have a sufficient number of items with adequate sensitivity to reliably capture reading achievement growth for this population of students (Osterlind, 2006). SWSCD are also known to exhibit limited attentional resources, which can give rise to fluctuating behaviors and learning responses (Kearns et al., 2011). These factors are worthy of further study and should be taken into consideration as states consider the inclusion of growth models within the context of the new ESSA (2015) for SWSCD.

Beyond evaluating the academic growth of SWSCD for substantive reasons, such as exploring typical rates of improvement for select student subgroups and the potential impact on measure design, the educational policy environment has placed considerable emphasis on evaluating students' growth as one component of accountability systems. As mentioned earlier, the *growth* included in accountability frameworks is tied to year-to-year gains, not true longitudinal growth, due at least in part to the need to generate annual reports for all students.

Given this context, modeling reading growth for SWSCD within statewide accountability systems is an ambitious undertaking. To support reliable estimates of growth with a large-scale, statewide AA-AAS that can lead to valid test use and interpretation, state departments of education will likely need to redesign status-based assessment systems to focus upon growth. This would require the development of numerous additional AA-AAS items across the grades tested. These items would need to reliably measure gains in academic achievement reflective of the developmental trajectories of academic growth for this population. New item delivery modes might also be needed to allow for more efficient use of time in testing to fulfill new ESSA (2015) proscriptions to protect instructional time. It is possible that computer adaptive assessment designs may meet this end, but adaptive instruments also present additional measurement decisions and challenges (van der Linden & Glas, 2008).

Results of the current study provided evidence that there was reading achievement growth over Grades 6 to 8 for SWSCD, and there were significant differences in initial achievement as a function of student exceptionality subgroup. This may be the most important finding from this study, as there is such a paucity of research on SWSCD reading growth (Farley et al., 2016; Tindal et al., 2015). Older discussions on the appropriate focus of academic instruction for SWSCD emphasized the development of functional skills. In the current educational assessment environment for SWSCD, we are no longer asking whether SWSCD can or should be learning academic content, but how much they can or should learn each year. This is a positive shift that denotes higher and more expansive expectations for these students. As mentioned by Kleinert et al. (2015), this population of students continues to meet or exceed educator expectations in the area of academic achievement, despite the fact that expectations have continually grown. In order to understand and evaluate academic achievement for SWSCD, we will need new assessments and models that can capture incremental student growth.

APPENDIX A

CORRELATIONS AMONG VARIABLES

Table 8

Correlations Among Outcome and Demographic Variables

Variable	RIT G6	RIT G7	RIT G8	Female	Non-White	Econ-Dis	Cohort 2
RIT G6	1.00	0.86***	0.83***	- 0.00	0.06*	- 0.17***	0.01
RIT G7	-	1.00	0.86***	- 0.03	0.01	- 0.15***	0.03
RIT G8	-	-	1.00	- 0.02	0.02	- 0.18***	- 0.01
Female	-	-	-	1.00	- 0.04	0.00	- 0.01
Non-White	-	-	-	-	1.00	- 0.24***	0.02
EconDis	-	-	-	-	-	1.00	- 0.01
Cohort 2	-	-	-	-	-	-	1.00
ID	-	-	-	-	-	-	-
CD	-	-	-	-	-	-	-
ED	-	-	-	-	-	-	-
OI	-	-	-	-	-	-	-
OHI	-	-	-	-	-	-	-
ASD	-	-	-	-	-	-	-
SLD	-	-	-	-	-	-	-

Note. RIT scores are Rasch Interval Unit scores for the ORExt Reading assessment. The exceptionality abbreviations are: ID = Intellectual Disability, CD = Communication Disorder, ED = Emotional Disturbance, OI = Orthopedic Impairment, OHI = Other Health Impairment, ASD = Autism Spectrum Disorder, SLD = Specific Learning Disability.

* $p < .05$

** $p < .001$

Table 9

Correlations Among Outcome and Demographic Variables, cont.

Variable	ID	CD	ED	OI	OHI	ASD	SLD
RIT G6	- 0.21***	0.13***	0.10***	- 0.18***	0.08**	- 0.18***	0.33***
RIT G7	- 0.17***	0.13***	0.07*	- 0.19***	0.08**	- 0.17***	0.33***
RIT G8	-0.18***	0.12***	0.08*	- 0.14***	0.08**	- 0.13***	0.30***
Female	0.16***	- 0.00	- 0.02	0.04	- 0.02	- 0.13***	- 0.04
Non-White	- 0.08***	0.09***	- 0.03	- 0.05	- 0.04	- 0.07**	0.16***
EconDis	- 0.00	- 0.06*	0.01	0.06**	0.05	0.15***	- 0.17***
Cohort 2	- 0.03	- 0.01	0.03	0.02	0.01	0.02	- 0.00
ID	1.00	- 0.19***	- 0.12***	- 0.14***	- 0.24***	- 0.34***	- 0.39***
CD	-	1.00	- 0.05	- 0.05*	- 0.09***	- 0.13***	- 0.15***
ED	-	-	1.00	- 0.03	- 0.06*	- 0.08***	- 0.10***
OI	-	-	-	1.00	- 0.07**	- 0.10***	- 0.11***
OHI	-	-	-	-	1.00	- 0.17***	- 0.19***
ASD	-	-	-	-	-	1.00	- 0.27***
SLD	-	-	-	-	-	-	1.00

Note. RIT scores are Rasch Interval Unit scores for the ORExt Reading assessment. The exceptionality abbreviations are: ID = Intellectual Disability, CD = Communication Disorder, ED = Emotional Disturbance, OI = Orthopedic Impairment, OHI = Other Health Impairment, ASD = Autism Spectrum Disorder, SLD = Specific Learning Disability.

* $p < .05$

** $p < .01$

** $p < .001$

APPENDIX B

TEST TAKING PATTERNS

Table 10

Longitudinal Test-taking Patterns, Count, and Percentages for each Pattern

Test taking pattern	*n*	%
AA-AA-GA	63	0.04
GA-AA-GA	48	0.03
GA-AA-AA	95	0.06
GA-AA-NA	9	0.01
GA-NA-AA	10	0.01
AA-GA-GA	182	0.11
AA-GA-AA	21	0.01
AA-GA-NA	25	0.02
AA-AA-GA	101	0.06
AA-AA-AA	874	0.54
AA-AA-NA	93	0.06
GA-NA-AA	13	0.01
AA-NA-AA	25	0.02
AA-NA-NA	53	0.03
Total	1,612	

Note. GA means that the student took the general state assessment that year, AA means that they took the alternate assessment, and NA means that data were missing for that year.

APPENDIX C

LGCM MODEL FIT STATISTICS AND PARAMETER ESTIMATES

Table 11

Model 1 Fit Statistics for all Latent Growth Curve Models

Model	Label	*AIC*	*BIC*	*AIC Weight*	*BIC Weight*	χ^2*p-value*	*RMSEA*	*CFI*	*TLI*	*SRMR*
Linear	1A	30,773.94	30,817.02	0.000	0.000	0.13	0.029	1.000	0.999	0.002
Nonlinear	1B	30,773.63	30,822.09	0.000	0.000	0.00	0.000	1.000	1.000	0.000
Quadratic	1C	30,773.63	30,822.09	0.000	0.000	0.00	0.000	1.000	1.000	0.000
Linear Student Demos	1D	30,722.47	30,807.86	0.000	0.000	0.16	0.020	0.999	0.997	0.005
Linear EconDis Exceptionality	1E	30,382.15	30,500.62	1.000	1.000	0.19	0.016	0.999	0.997	0.004

Table 12

Model 1 Intercepts, Slopes, Time, and Intercept and Slope Variances

Model	Model Label	*Intercept*	*Slope*	*Time*	*Intercept* σ^2	*Slope* σ^2
Linear	1A	102.99*	2.79*	0-1-2	438.16	14.18
Nonlinear	1B	102.86*	3.26*	0-1-1.70	445.93	21.56
Quadratic	1C	102.86*	3.75*	0-1-2/0-1-4	441.25	16.03
Linear Student Demos	1D	105.37*	3.32*	0-1-2	425.75	14.45
Sex		105.12	3.04			
Race		105.60	2.60			
EconDis		97.51*	2.77			
Linear Econdis Exceptionality	1E	97.40*	2.72*	0-1-2	334.35	13.76
Econdis		93.12*	2.39			
CD		114.42*	2.85			
ED		115.95*	1.87			
OI		83.86*	4.43			
OHI		109.96*	2.60			
ASD		96.41	2.81			
SLD		116.49*	3.90			

Note. ID = Intellectual Disability (reference group), CD = Communication Disorder, ED = Emotional Disturbance OI = Orthopedic Impairment, OHI = Other Health Impairment, ASD = Autism Spectrum Disorder, SLD = Specific Learning Disability, and EconDis = Economic Disadvantage.

* $p < .001$

APPENDIX D

EFFECT SIZE GAPS GRADES 6 AND 7

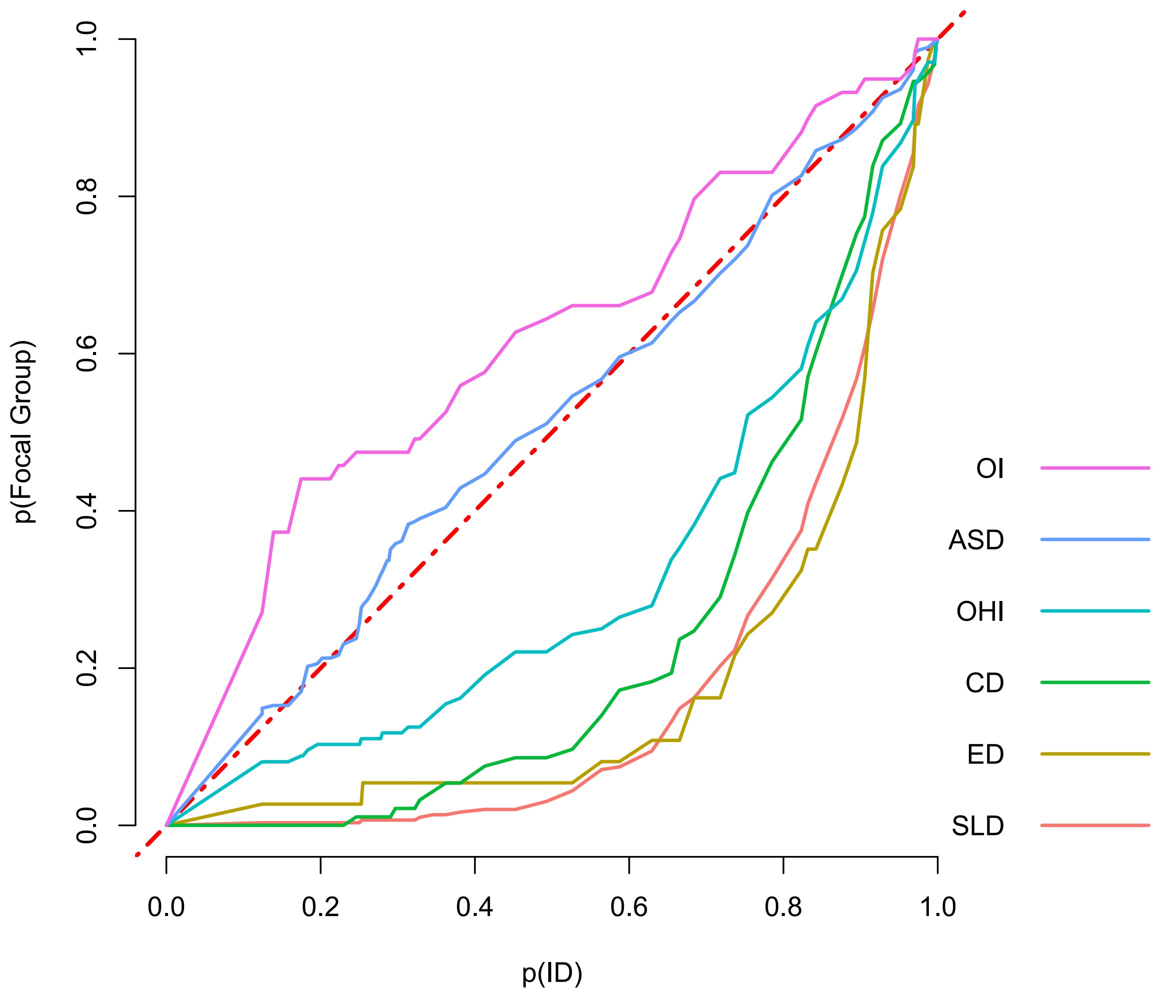

Figure 10. Effect sizes gaps in Grade 6. Gaps shown between student exceptionality groups with students identified as ID as the reference group. ID = Intellectual Disability, CD = Communication Disorder, ED = Emotional Disturbance OI = Orthopedic Impairment, OHI = Other Health Impairment, ASD = Autism Spectrum Disorder, SLD = Specific Learning Disability.

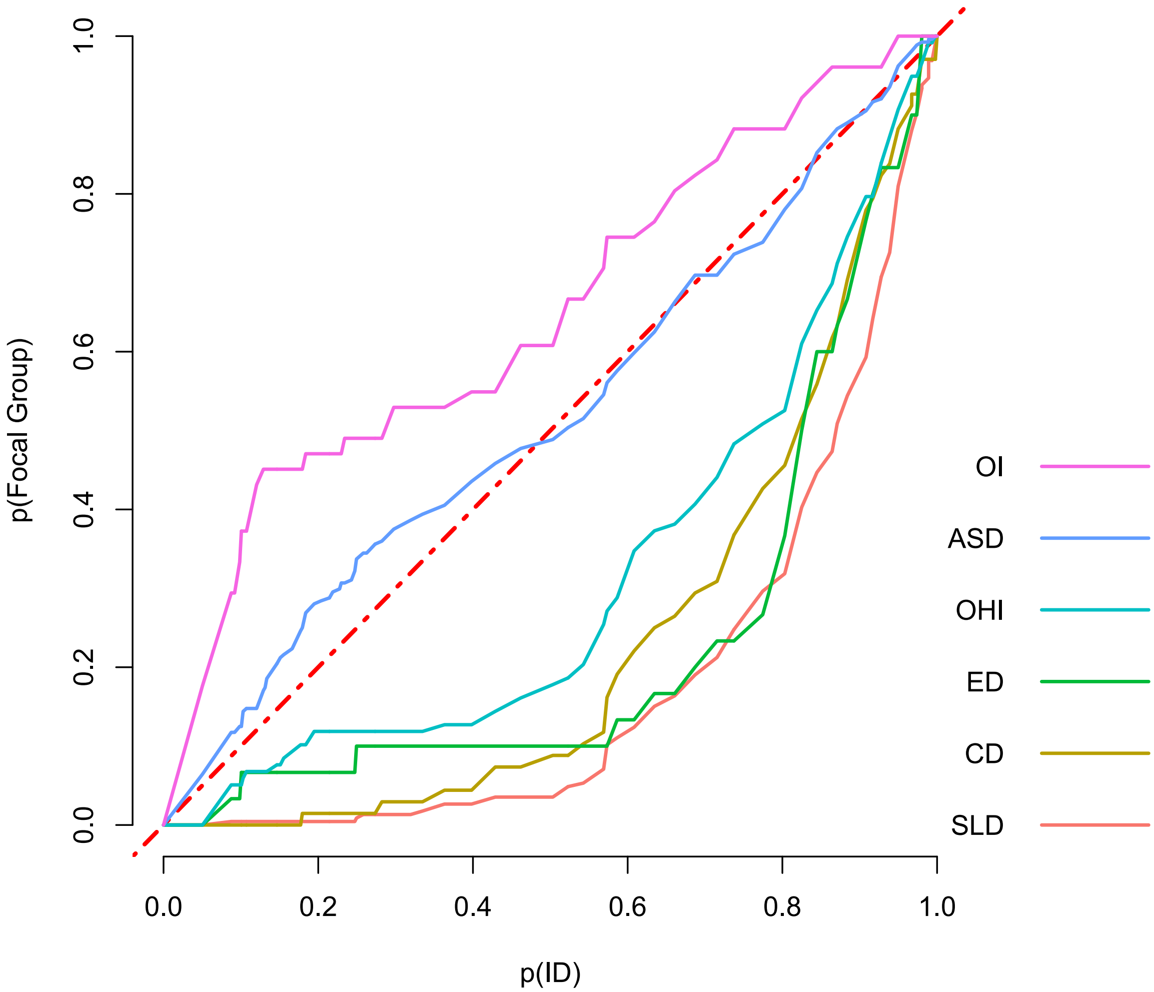

Figure 11. Effect sizes gaps in Grade 7. Gaps shown between student exceptionality groups with students identified as ID as the reference group. ID = Intellectual Disability, CD = Communication Disorder, ED = Emotional Disturbance OI = Orthopedic Impairment, OHI = Other Health Impairment, ASD = Autism Spectrum Disorder, SLD = Specific Learning Disability.

APPENDIX E

LCGA FIT STATISTICS AND PARAMETER ESTIMATES

Table 13

Model 2 Latent Class Growth Analysis Results

Model	*-LL*	*AIC*	*BIC*	*R1*	*R2*	*R3*	*LMR-LRT p-value*	*BLRT*	*Entropy*
C(1)	- 16,851.89	33,713.79	33,740.72	536.18	539.77	530.57	N/A	N/A	N/A
Δ	1,460.24	- 2,914.49	- 2,898.33	- 384.17	- 358.97	- 352.67			
C(2)	- 15,391.65	30,799.30	30,842.38	152.01	180.80	177.90	<.001	<.001	0.979
Δ	269.16	- 532.32	- 516.17	- 33.62	- 56.89	- 68.35			
C(3)	-15,122.49	30,266.98	30,326.21	118.39	123.91	109.55	0.002	<.001	0.885
Δ	158.15	- 310.29	- 294.14	- 48.39	- 20.46	3.85			
C(4)	- 14,964.34	29,956.69	30,032.08	70.00	103.46	113.40	0.052	<.001	0.821
Δ	144.10	- 282.20	- 266.05	- 1.75	2.59	- 42.71			
C(5)	- 14,820.24	29,674.49	29,766.03	68.25	106.05	70.69	0.006	<.001	0.861
Δ	89.68	- 173.36	- 157.20	- 17.25	1.18	-2.68			
C(6)	- 14,730.56	29,501.13	29,608.83	51.00	107.23	68.02	0.007	<.001	0.868
Δ	77.32	- 148.64	- 132.48	- 8.01	- 10.04	- 12.32			
C(7)	- 14,653.24	29,352.49	29,476.35	42.99	97.19	55.70	0.167	<.001	0.830

Table 14

Model 2 Latent Class Growth Analysis n-sizes, Percentages, Intercept, and Slope Estimates

LCGA Model	Class	*Ave. Latent Class Probability*	*n*	*%*	*Intercept*	*Slope*	*SE Intercept*	*SE Slope*
One-Class Solution	Class 1	1.00	1,612	1.00	101.41	1.80	0.62	0.27
Two-Class Solution	Class 1	0.982	252	0.16	59.88	3.62	1.15	0.84
	Class 2	0.997	1,360	0.84	110.59	1.82	0.36	0.19
Three-Class Solution	Class 1	0.881	253	0.16	90.11	3.48	2.77	0.90
	Class 2	0.982	187	0.12	55.35	2.16	1.38	1.24
	Class 3	0.959	1,172	0.73	113.83	2.09	0.61	0.19
Four-Class Solution	Class 1	0.894	957	0.59	116.59	2.19	0.56	0.21
	Class 2	0.868	378	0.23	101.20	2.19	0.93	0.31
	Class 3	0.913	126	0.08	76.07	3.11	1.92	2.31
	Class 4	0.979	151	0.09	50.07	4.47	1.27	1.51

REFERENCES CITED

Altman, J. R., Lazarus, S. S., Quenemoen, R. F., Kearns, J., Quenemoen, M., & Thurlow, M. L. (2010). *2009 survey of states: Accomplishments and new issues at the end of a decade of change.* Retrieved from: http://www.cehd.umn.edu/NCEO/onlinepubs/2009StateSurvey.pdf

Anderson, D. (2017). *esvis: Calculation and visualization of effect sizes*. R package version 0.0.0.9000. Installed from https://github.com/DJAnderson07/esvis.

Anderson, D., Farley, D., & Tindal, G. (2015). Test design considerations for students with significant cognitive disabilities. *The Journal of Special Education, 49* (1), 3-15. doi:10.1177/0022466913491834

Bilir, M. K., Binici, S., & Kamata, A. (2008). Growth mixture modeling: Application to reading achievement data from a large-scale assessment. *Measurement and Evaluation in Counseling and Development, 41*.

Blackorby, J., Wagner, M., Cameto, R., Davies, E., Levine, P., Newman, L., …Sumi, C. (2005). Engagement, academics, social adjustment, and independence: The achievement of elementary and middle school students with disabilities. Retrieved from http://www.seels.net/designdocs/engagement/All_SEELS_outcomes_10-04-05.pdf

Bloom, H. S., Hill, C. J., Black, A. R., & Lipsey, M. W. (2008). *Performance trajectories and performance gaps as achievement effect-size benchmarks for educational interventions*. New York: MDRC.

Bollen, K. A., & Curran, P. J. (2006). *Latent curve models: A structural equation perspective*. Hoboken, NJ: John Wiley & Sons, Inc.

Browder, D. M., Ahlgrim-Delzell, L., Flowers, C., Karvonen, M., Spooner, F., & Algozzine, R. (2002, April). *How states define alternate assessments for students with disabilities*. Paper presented at the Annual Meeting of the American Educational Research Association, Charlotte, NC.

Burnham, K. P., & Anderson, D. R. (2004). Understanding AIC and BIC in model selection. *Sociological Methods & Research, 33*(2), 261-304. doi:10.1177/0049124104268644

Buzick, H. M., & Laitusis, C. C. (2010). Using growth for accountability: Measurement challenges for students with disabilities and recommendation for research. *Educational Researcher, 39*, 537-544.

Cohen, J. (1992). A power primer. *Psychological Bulletin, 112*, 155-159.

Cohen, J. (1988), *Statistical Power Analysis for the Behavioral Sciences, 2nd Edition*. Hillsdale: Lawrence Erlbaum.

Dunn, J., Roussos, L., Lonczak, T., & Sukin, T. (2012). *An exploration of growth models for an alternate assessment.* Paper presented at the American Educational Research Association Annual Conference, Vancouver, Canada.

Elliott, S. N. (2015). Measuring opportunity to learn and achievement growth: Key research issues with implications for the effective education of all students. *Remedial and Special Education, 36*(1), 58-64. doi:10.1177/0741932514551282

The Every Student Succeeds Act of 2015, Pub. L. No. 114-95 § 1111 (ESSA; 2015).

Farley, D., Anderson, D., Irvin, P. S., & Tindal, G. (2016). Modeling Reading Growth in Grades 3-5 with an Alternate Assessment. *Remedial and Special Education.* doi: 10.1177/0741932516678661

Faul, F., Erdfelder, E., Lang, A-G., & Buchner, A. (2009). G*Power, version 3.1.9.2.

George, D., & Mallery, M. (2010). *SPSS for Windows step by step: A simple guide and reference*, 17.0 update (10a ed.) Boston: Pearson.

Hu, L., & Bentler, P. M. (1999). Cutoff criteria for fit indices in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling, 6*(1), 1-55.

IBM Corp. (2016). IBM SPSS Statistics for Windows, Version 24.0. Armonk, NY: IBM Corp.

Individuals with Disabilities Education Act amendments, Pub. L. No. 105-17 (IDEA; 1997).

Individuals with Disabilities Education Act, 20 U.S.C., U. S. Department of Education, Pub. L. No. 108-446 § 1400 *et seq.* (IDEA; 2004).

Jamshidian, M., Jalal, S., & Jansen, C. (2014). MissMech: An R package for testing homoscedasticity, multivariate normality, and missing completely at random (MCAR). *Journal of Statistical Software, 56*(6), 1-31.

Judge, S. & Bell, S. M. (2010). Reading achievement for students with learning disabilities during the elementary school years. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 27, 153-178. doi 10.1080/10573569.2011.532722

Judge, S., & Watson, S. M. R. (2011). Longitudinal outcomes for mathematics achievement for students with learning disabilities. *The Journal of Educational Research, 104*(3), 147-157. doi:10.1080/00220671003636729

Jung, T., & Wickrama, K. A. S. (2008). An introduction to latent class growth analysis and growth mixture modeling. *Social and Personality Psychology Compass, 2*(1), 302-317.

Kamata, A., Nese, J. F. T., Patarapichayatham, C., & Lai, C.-F. (2013). Modeling nonlinear growth with three data points: Illustration with benchmarking data. *Assessment for Effective Intervention, 38*(2), 105-116.

Karvonen, M., Flowers, C., & Wakeman, S. Y. (2013). *An exploration of methods for measuring academic growth for students with significant cognitive disabilities*. Paper presented at the Annual Meeting of the American Educational Research Association, San Francisco, CA.

Kearns, J. F., Towles-Reeves, E., Kleinert, H. L., Kleinert, J. O., & Thomas, M. K.-K. (2011). Characteristics of and implications for students participating in alternate assessments based on alternate academic achievement standards. *Journal of Special Education, 45*(1), 3-14.

Kettler, R. J., Elliott, S. N., Beddow, P. A., Compton, E., McGrath, D., Kaase, K. J., . . . Hinton, K. (2010). What do alternate assessments of alternate academic achievement standards measure? A multitrait-multimethod analysis. *Exceptional Children, 76*, 457-474.

Kleinert, H. L., Browder, D. M., & Towles-Reeves, E. A. (2009). Models of cognition for students with significant cognitive disabilities: Implications for assessment. *Review of educational research, 79*(1), 301-326.

Kleinert, H. L., Kennedy, S., & Kearns, J. F. (1999). The impact of alternate assessments: A statewide teacher survey. *The Journal of Special Education, 33*, 93-102. doi:10.1177/002246699903300203

Kleinert, H. L., Towles-Reeves, E., Quenemoen, R., Thurlow, M. Fluegge, L. Weseman, L. & Kerbel, A. (2015). Where students with significant cognitive disabilities are taught: Implications for general curriculum access. *Exceptional Children, 81*(3), 312-328. doi:10.1177/0014402914563697

Kline, R. B. (2016). *Principles and practices of structural equation modeling* (4th ed.). New York: The Guilford Press.

Koretz, D. M., & Hamilton, L. S. (2006). Testing for accountability in K-12. In R. L. Brennan (Ed.), *Educational Measurement* (4th ed., pp. 531-578). Westport, CT: Praeger Publishers.

Little, R. J. A., & Rubin, D. B. (2002). *Statistical analyses with missing data*. Hoboken, NJ: John Wiley & Sons.

Masters, G. N. (1982). A Rasch model for partial credit scoring. *Psychometrika, 47*(2), 149-174. doi:10.1007/BF02296272

Morgan, P. L., Farkas, G., & Wu, Q. (2011). Kindergarten children's growth trajectories in reading and mathematics: Who falls increasingly behind? *Journal of Learning Disabilities, 44*(5), 472-488. doi:10.1177/0022219411414010

Muthén, L. K., & Muthén, B. O. (1998-2015). *Mplus user's guide*. Los Angeles, CA: Muthén & Muthén.

Oregon Department of Education (ODE; 2009). 2008-09 technical report: Oregon's alternate assessment system.

Oregon Department of Education (ODE; 2010). 2009-10 technical report: Oregon's alternate assessment system.

Oregon Department of Education (ODE; 2011A). State Consolidated State Application Accountability Workbook.

Oregon Department of Education (ODE; 2011B). 2010-11 technical report: Oregon's alternate assessment system.

Oregon Department of Education (ODE; 2012). 2011-12 technical report: Oregon's alternate assessment system.

Osterlind, S. J. (2006). Modern measurement: Theory, principles, and applications of mental appraisal. Upper Saddle River, New Jersey: Pearson.

The No Child Left Behind Act of 2001, Pub. L. No. 107-110 § 115, Stat. 1425 (NCLB; 2002).

Pennington, R. C., & Courtade, G. R. (2015). An examination of teacher and student behaviors in classrooms for students with moderate and severe intellectual disability. *Preventing School Failure: Alternative Education for Children and Youth*, *59*(1), 40-47. doi:10.1080/1045988X.2014.919141

Perie, M. (2007). *Setting alternate achievement standards*. National Center for the Improvement of Educational Assessment, Inc.

Quenemoen, R. (2008). *A brief history of alternate assessments based on alternate achievement standards (Synthesis Report 68)*. National Center on Educational Outcomes, Minneapolis, MN.

R Core Team (2016). R: A language and environment for statistical computing. R Foundation and Statistical Computing, Vienna, Austria. URL: https://www.R-project.org.

Raudenbush, S. W. (2001). Comparing personal trajectories and drawing causal inferences from longitudinal data. *Annual Review of Psychology, 52*, 501-525. doi:10.1146/annurev.psych.52.1.501

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

Rieke, R., Lazarus, S. S., Thurlow, M., & Dominguez, L. M. (2013). *2012 Survey of states: Successes and challenges during a time of change*. Minneapolis, MN: University of Minnesota, National Center on Educational Outcomes.

Rogosa, D. R., Brandt, D., & Zimowski, M. (1982). A growth curve approach to the measurement of change. *Psychological Bulletin, 92*, 726-748. doi: 10.1037/0033-2909.92.3.726

Rogosa, D. R., & Willett, J. B. (1985). Understanding correlates of change by modeling individual differences in growth. *Psychometrika, 50*, 203-228. doi:10.1007/BF02294247

Saven, J. L., Anderson, D., Nese, J. F. T., Farley, D., & Tindal, G. (2015). Patterns of state assessment participation for students with significant cognitive disabilities. *Journal of Special Education*, Advanced online publication. doi:10.1177/0022466915582213

Saven, J. L., Farley, D., & Tindal, G. (2013). Constructing alternate assessment cohorts: An Oregon perspective. Retrieved from the National Center on Assessment and Accountability for Special Education (NCAASE) website: http://www.ncaase.com/publications/in-briefs

Schulte, A., & Stevens, J. (2015). Once, sometimes, or always in special education: Mathematics growth and achievement gaps. *Exceptional Children*, 1-18. doi:10.1177/0014402914563695

Schulte, A., Stevens, J., Elliott, S., Tindal, G., & Nese, J. F. T. (2016). Achievement gaps for students with disabilities: Stable, widening, or narrowing on state-wide reading comprehension test? *Journal of Educational Psychology, Advance online publication*. doi:http://dx.doi.org/10.1037/edu0000107

Schulte, A., & Stevens, J. J. (2013). *Special education growth: Contrasting stable and variable identification of special education student status across grades*. Paper presented at the annual meeting of the National Council for Measurement in Education, San Francisco, CA.

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

Stevens, J., Schulte, A., Elliott, S. N., Nese, J. F. T., & Tindal, G. (2015). Growth and gaps in mathematics achievement of students with and without disabilities on a statewide achievement test. *Journal of School Psychology, 53*, 45-62.

Stevens, J. P. (2002). *Applied multivariate statistics for the social sciences* (Fourth ed.). Mahwah, NJ: Lawrence Ehrlbaum Associates.

Thompson, S., & Thurlow, M. (2001). *2001 State special education outcomes: A report on state activities at the beginning of a new decade*. Retrieved from: http://www.cehd.umn.edu/NCEO/OnlinePubs/StateSpecEducOutcomes2001.pdf

Tindal, G., Nese, J. F. T., Farley, D., Saven, J. L., & Elliott, S. N. (2015). Documenting reading achievement and growth for students taking alternate assessments. *Exceptional Children*, 1-16. doi:10.1177/0014402915585492

Tindal, G., Schulte, A. C., Elliott, S. N., & Stevens, J. J. (2011). *Research and Development Grant Application*. Retrieved from: http://www.ncaase.com/publications/tech-reports

United States Department of Education. (USDE; 2003). Title 1 - Improving the academic achievement of the disadvantaged; final rule, 34 CFR § 200 (2003).

United States Department of Education. (USDE; 2005). *Alternate achievement standards for students with the most significant cognitive disabilities: Non-regulatory guidance*. Washington, D. C. Retrieved from http://www2.ed.gov/policy/elsec/guid/altguidance.pdf

United States Department of Education. (USDE; 2015). ESEA Flexibility. Retrieved from http://www2.ed.gov/policy/elsec/guid/esea-flexibility/index.html

van der Linden, W. J., & Glas, C. A. W. (2010). Elements of Adaptive Testing. New York: Springer.

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.

Wei, X., Lenz, K. B., & Blackorby, J. (2012). Math growth trajectories of students with disabilities: disability category, gender, racial, and socioeconomic status differences from ages 7 to 17. *Remedial and Special Education, 34*(3), 154-165. doi:10.1177/0741932512448253

Ysseldyke, J., & Olsen, K. (1999). Putting alternate assessments into practice: What to measure and possible sources of data. *Exceptional Children, 65*, 175-185.