DO DOLLARS MATTER BEYOND DEMOGRAPHICS?

DISTRICT CONTRIBUTIONS TO READING AND MATHEMATICS GROWTH FOR STUDENTS WITH DISABILITIES

by

JESSICA LYNN SAVEN

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

June 2015

DISSERTATION APPROVAL PAGE

Student: Jessica Lynn Saven

Title: Do Dollars Matter Beyond Demographics? District Contributions to Reading and Mathematics Growth for Students with Disabilities

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

Gerald Tindal	Chairperson
Joseph Stevens	Core Member
Charles Martinez	Core Member
Robert Horner	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 June 2015

DISSERTATION ABSTRACT

Jessica Lynn Saven

Doctor of Philosophy

Department of Educational Methodology, Policy, and Leadership

June 2015

Title: Do Dollars Matter Beyond Demographics? District Contributions to Reading and Mathematics Growth for Students with Disabilities

Growth modeling in education has focused on student characteristics in multilevel growth accountability models and has rarely included financial variables. In this dissertation, relations of several demographic and financial characteristics of Oregon school districts to the reading and mathematics growth of students receiving special education services in Grades 3-8 were explored after accounting for student level demographic characteristics. Previous research indicated that three variables were potentially related to student growth: district level aggregated student demographics, district geography (e.g., location in a remote area), and district funding. Three sources of data were used to investigate these relationships: institutional data reported by the Oregon Department of Education, the Common Core of Data gathered by the National Center for Education Statistics, and Oregon Assessment of Knowledge and Skills test data collected as part of the National Center on Assessment and Accountability in Special Education.

Multi-level models of student growth across Grades 3-8 were constructed for reading and mathematics, with time (level-1) nested within students (level-2) and districts (level-3). Results demonstrated that although student-level demographic factors account for the majority of meaningful differences in student growth, both district demographic characteristics and financial investment in students were related to growth for students who received special education services.

CURRICULUM VITAE

NAME OF AUTHOR: Jessica Lynn Saven

GRADUATE AND UNDERGRADUATE SCHOOLS ATTENDED:

University of Oregon, Eugene
Johns Hopkins University, Baltimore, MD

DEGREES AWARDED:

Doctor of Philosophy, Educational Methodology, Policy, and Leadership, 2015, University of Oregon
Master of Arts in Teaching, Secondary Chemistry, 2003 Johns Hopkins University
Bachelor of Science, Environmental Studies, 2001, University of Oregon

AREAS OF SPECIAL INTEREST:

Quantitative Research Methodology
Educational Accountability Systems

PROFESSIONAL EXPERIENCE:

Graduate Research Assistant, Behavioral Research and Teaching, 2011-2015
Secondary Science Teacher, Baltimore City Public School System, 2001-2011

PUBLICATIONS:

Tindal, G., Nese, J. F. T., Farley, D., & Saven, J. L. (2015). Documenting reading achievement growth for students with significant cognitive disabilities. *Exceptional Children.* Advance online publication. doi: 10.1177/0014402915585492

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

ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Tindal, for providing me with opportunities to grow professionally and expand my academic skillset in ways I could not have dreamed of before working with him at BRT. His combination of mentoring and challenging me to push myself at every turn made my pursuit of this degree a truly rewarding experience. I would also like to acknowledge the support and expertise of others at BRT and in the EMPL department as a whole throughout my experience in this program. This research was funded in part by a Cooperative Service Agreement from the Institute of Education Sciences (IES) establishing the National Center on Assessment and Accountability for Special Education – NCAASE (PR/Award Number R324C110004), and I thank the principal investigators for allowing me to use these data. On a more personal note, I'd like to thank my family. My parents have inspired me throughout my life to keep learning and do my best to effect positive outcomes for students – both when I was a teacher and now through my research. My sisters cheered me on and reminded me to take time to breathe along the way, and my life is better because of that. I'd also like to thank my partner Bill, who walked alongside me on this great adventure, for reminding me in my most chaotic moments that there was a light at the end of the tunnel, and for filling my world with so much joy.

For all of my former students, especially those who inspired me to ask more questions.

TABLE OF CONTENTS

Chapter Page

Chapter	Page

LIST OF FIGURES

Figure Page

LIST OF TABLES

CHAPTER I

INTRODUCTION

In December of 2001, the U.S. Congress passed the No Child Left Behind Act (NCLB), adding accountability in the form of Adequate Yearly Progress (AYP) to the agenda of each school, district, and state education agency. This NCLB legislation provided the basis for better documenting achievement for subgroups of students (disability status, English proficiency, socioeconomic status and race/ethnicity), thus ensuring the overall averages for schools and districts no longer masked effects for these subgroups. However, the use of a status model, rather than individual student growth to assess academic progress presented a number of problems in reflecting accountability, as each successive year was based on a different population of students. This status model basically confounded achievement effects with population differences. In the case of the diverse subgroup of students with disabilities (SWDs) the limitations of this status model were exacerbated due to the additional volatility from small sample sizes year-to-year.

Over the course of the next decade, meeting AYP based on reading and mathematics scores for students in Grades 3-8 and at least once in high school became a high priority for school districts across the country. Under NCLB (2002), 100% of students in states were required to achieve proficiency on state assessments by school year 2013-2014. The U. S. Department of Education (USED; 2014) interpretation of this proficiency mandate stated these standards were necessary in order to prevent some of the nation’s children from being denied a quality education. That interpretation, however, created difficulty for schools and districts to respond to “the massive challenge and statistical improbability of having 100% of any student population being rated proficient on the basis of a one-shot standardized examination" (Rentschler, 2005, p. 637). The specifics of NCLB legislation meant that educators and policymakers were documenting school increases in proficiency as the change in proficiency from previous years (with different students), rather than looking at gains of individual students.

In this section, I discuss post-NCLB accountability concerns faced by school districts, and concerns regarding district-level AYP for student subgroups within the state of Oregon. Pressures from two sources are explored: (a) district context and (b) financial investments in educating students. Within this AYP subgroup framework, longitudinal student growth is examined as a more suitable way to *adequately* measure student progress than the status model of NCLB. I conclude with specific research questions to better understand the relationships of district demographics and financial investments to growth for SWDs.

Adequate Yearly Progress

NCLB legislation required schools, districts, and states to meet increasing AYP targets yearly, so that 100% of the students were proficient in reading and mathematics by school year 2013-2014. However, the quantification of *proficient* was determined state by state based on cut-scores from unique state assessment systems. Universal proficiency did not necessarily mean a "common high level of student achievement" (Linn, 2008, p. 29).

Under NCLB (2002), AYP for both schools and districts is calculated using a status model, with the end result being the proportion of students attaining proficient status. To determine district AYP performance rates, proficient test scores from students in each district are divided by the expected number of proficient test scores from students enrolled on the first school day in May, not including students ineligible for testing (ODE, 2012a). Students, schools, and districts are evaluated by comparing student proficiency levels in a given grade to the number proficient in the previous year.

This status model is appealing to policy makers due to its transparency (e.g., it is easy to understand) and it attends to low-performing students in need of remediation (Figlio & Ladd, 2008; Rouse, Hannaway, Goldhaber, & Figlio, 2013). For example, using a regression-discontinuity design and data from principal surveys, Rouse, Hannaway, Goldhaber, and Figlio (2013) found that accountability pressures led to an increased emphasis on low-performing students in Florida.

However, a number of issues arise when using a status model as the only metric of progress. Schools designated ineffective based on aggregate performance levels may actually improve the achievement of individual students, yet be penalized for not meeting AYP (e.g., Clotfelter & Ladd, 1996; Kane & Staiger, 2002; Ladd & Walsh, 2002). In addition, pressure for 100% proficiency using a status model is likely to result in re-prioritization of resources at the school and district levels in order to improve student achievement. Perhaps the greatest consequence of using a status model is that all groups of students are compared to an "arbitrary score on a one-shot examination" rather than evaluated on their individual progress (Rentschler, 2005, p. 657). Finally, rather than ensuring access to a high-quality education for all students, the focus on AYP under NCLB often limits attention to (and focuses resources on) students at-risk, particularly those who *almost* meet proficiency standards (Goldhaber, 2002).

Student Subgroups for AYP

The overall effectiveness of NCLB depends both on measurement of school performance, and the treatment of student subgroups (Figlio & Ladd, 2008). Under NCLB, these four non-exclusive subgroups are defined as: students with disabilities (SWDs), race/ethnicity, limited English proficient (LEP), and economically disadvantaged. A 2004 letter to Chief State School Officers from the Assistant Secretary for the Office of Elementary and Secondary Education clarified that state performance rates for AYP needed to include reports on these four subgroups as determined by each state's accountability plan (Simon, 2004). Due to the non-exclusive nature of these subgroups, an individual student could potentially be a member of all subgroups reported on in a district (e.g., a Latino student with LEP, who is visually impaired and economically disadvantaged), and influence AYP not only as a member of the total population, but in four subgroup scores. This student's achievement would be weighted five times that of a typical student for accountability purposes.

A number of problems exist in addressing achievement of student subgroups. First, great variation exists in subgroup minimums nationwide, ranging from 5 to 100 students (Fulton, 2006; Porter, Linn, & Trimble, 2005). Second, although large subgroup sizes chosen by some states may increase the accuracy of measurements of performance, they may exclude those whom NCLB subgroup reporting was designed to help monitor (Figlio & Ladd, 2008). Finally, states defend these decisions based on concerns that smaller subgroup sizes risk over-identifying schools as missing AYP targets (Fulton, 2006). In what follows, description of and concerns regarding accountability for the

SWD subgroup are discussed, followed by discussion of key issues associated with the other three demographic subgroups that SWDs may also be members of.

Students with disabilities. SWDs make up just above 10% of the student population nationwide (Gordon, 2008). Of the 13 disability categories under IDEA (1997), 12 are used in Oregon: autism spectrum disorder, communication disorder, deaf-blind, developmental delay, emotional disturbance, hearing impairment, intellectual disability, orthopedic impairment, other health impairment, specific learning disability, traumatic brain injury, and visual impairment. A multiple disabilities category is not used: instead, students with more than one disability have each listed on their individualized education program (IEP). In order to qualify a child for services, a student's IEP team must determine that a student not only has one of these 12 disabilities, but also that the disability adversely affects their educational performance (IDEA, 1997).

Prior to amendments to the Individuals with Disabilities Education Act (IDEA) in 1997, the focus of accountability in special education (SpEd) was more on fiscal and procedural matters than student performance, except, in some cases, during the IEP review process (Harr, Parrish, & Chambers, 2008; McLaughlin & Thurlow, 2003). Administrators commonly placed SWDs and students with limited English proficiency (LEP) into non-tested groups at the state level in order to not lower the scores of schools and districts (Goldhaber, 2002). However, changes to IDEA (1997) required districts to include SWDs in local and statewide assessments with appropriate accommodations, set goals for these students, and report on their performance to the U.S. Secretary of Education. Partly because of this relatively recent focus on achievement for SWDs, Schulte and Stevens (2015) noted that "the paucity of studies and lack of consistency in methodology limit the confidence with which generalizations about achievement growth for SWDs can be made" (p. 3). The requirement under NCLB (2002) that at least 95% of students in each subgroup participate in testing each year provides a source of annual achievement data for SWDs, and can be used for further investigation of their academic gains in reading and mathematics.

One concern regarding including SWDs in NCLB accountability is that scores from this subgroup need not be disaggregated if a school or district has too few qualifying students (McLaughlin & Thurlow, 2003), which in turn limits what can be learned about SWDs in schools and districts where they are few in number. All students in such a district may take an assessment, but the SWD group results may not be reported (McLaughlin & Thurlow, 2003). For example, for school year 2000-2001, only 28 states reported both participation and performance rates for SWDs (Thurlow, Wiley, & Bielinski, 2003). Within the state of Oregon, minimum subgroup size at the school and district level is 42 students (ODE, 2012a). In cases where the SWD subgroup is not disaggregated from other students due to small subgroup size, a district may focus less attention on outcomes for SWDs, as their scores are more easily masked by students not receiving services (SWoD).

Membership in the SWD subgroup for NCLB accountability changes depending, in part, on perceived needs of students during annual IEP meetings (Ysseldyke & Bielinski, 2002). In addition to students changing disability classification, the type of test they are administered can also change (Saven, Anderson, Nese, Farley, & Tindal, in press) . Students with the most significant cognitive disabilities (SWSCDs) can take Oregon's Extended Assessment (ORExt), designed with reduced depth and breadth in order to be accessible to this population (ODE, 2012b). ORExt is scored on a different scale than the general assessment, which makes it difficult to include SWDs taking this measure in growth models (Ahearn, 2009). Despite this avenue for diverse learners, only 1% of the tested population "count" for AYP using scores from this assessment (USED, 2005). Many educators are concerned about the performance of SWDs that function at a level too high to qualify for ORExt, but not high enough for the general assessment to be an adequate measure of their achievement (Nagle, Yunker, & Malmgren, 2006).

SWDs present challenges for modeling student growth, as membership in this *official* subgroup changes when students enter and exit services (e.g., Ysseldyke & Bielinski, 2002). This has consequences with respect to the spirit of NCLB, as it is more difficult with changing membership to accurately model the achievement of SWDs as compared to their SWoD peers. Schulte and Stevens (2015), analyzed three cohorts of students based on being identified as SWDs in the first wave of measurement, at any point during Grades 3-7, and those continuously identified over five years. Their research demonstrated that while the achievement gap between SWDs and SWoD students was lower in their second, most inclusive model, SWDs showed lower average mathematics achievement and a slower rate of mathematics growth than SWoD students overall. Including students receiving SWDs services at any point during their education in models allows a better look at the full range of growth for SWDs. This practice also includes those students with disabilities frequently identified later in elementary school, as curriculum becomes more challenging. Without this consideration, the SWD subgroup will demonstrate limited gains in performance, as the top end of students who exit services are omitted.

Race/ethnicity. For purposes of NCLB (2002) accountability, analysis of the race/ethnicity subgroup centers on five specific categories: Black, Hispanic, White, American Indian, and Asian/Pacific Islander. In areas with low minority populations, students may not be disaggregated into the four non-White groups, which limits the information available about non-White students in majority White areas.

In Oregon, the race/ethnicity of students in Grades prekindergarten through 12 for school year 2006-2007 consisted of approximately 70% White, 3% Black, 16% Hispanic, 5% Asian/Pacific Islander, 2% American Indian/Alaska Native, 2% multiracial, and 2% unknown (http://www.ode.state.or.us/sfda/reports/r0067Select2.asp). This amounted to approximately half the national average for non-White students by 2008 (Aud et al., 2011). The number of non-White students in many small districts in Oregon was small enough not to necessitate NCLB reporting – particularly when considering specific racial/ethnic groups rather than non-White as a whole. For example, in 2006-2007, 79 of the 197 districts in Oregon had less than 42 non-White students total: far less than the required 42 students of a specific minority for subgroup reporting under NCLB. Of these 79 districts, two were majority non-White, yet were exempt from race/ethnicity subgroup reporting because the size of the districts was so small. An additional eight districts also were majority non-White, indicating substantial variability across districts.

A great deal of research has focused on the effect of race/ethnicity on student academic outcomes. For example, comparing National Assessment of Educational Progress reading and mathematics data for school year 2008-2009 across races, the scale scores for students in every race/ethnicity category except Asian/Pacific Islander were below the average scale scores of White students in Grades 4, 8, and 12 (Aud et al., 2011). Hanushek and Raymond (2004) found that while state-level accountability systems in the 1990s increased state achievement gain in reading and mathematics, and tended to narrow the Hispanic-White gap, they also widened the Black-White gap. In an investigation of the Black-White and Hispanic-White achievement gaps in reading, Lee and Reeves (2012) found no post-NCLB changes evident in either the status or growth rate for Grades 4 and 8.

Students with limited English proficiency. One purpose of NCLB was to reduce gaps in achievement between non-English proficient (LEP) students and their peers proficient in English. As of school year 2008-2009, 20% of students ages 5-17 in Oregon spoke a language other than English at home and spoke English with difficulty themselves (Aud et al., 2011). Of these students, the home languages spoken included approximately 74% Spanish, 8% Indo-European, and 12% Asian/Pacific Islander (Aud et al., 2011). Transitioning students who demonstrated fluency in English were placed on monitoring status until they no longer needed instructional services and methods provided by the district, or up to two years. Monitoring ended after a maximum of 5 years, at which time students were no longer reported as part of the LEP subgroup (ODE, 2012b). Teachers and instructional teams (including, in some cases, parents) decided which accommodations might have been necessary for students with LEP during testing.

Students with LEP are not required to take state reading and writing assessments during their first year of enrollment in U.S. schools, although they must take an English language proficiency measure instead (ODE, 2012). First-year students in Oregon take the English Language Proficiency Assessment, and are counted as participants in AYP reports in Reading or Writing. Students with LEP take state mathematics assessments even in their first year (with a Spanish version and a Russian version available as possible accommodations), which count for participation in AYP.

Overall, students with LEP lag behind children from English-only backgrounds in academic achievement. For example, students with LEP ranged from 0.86 to 1.14 standard deviations below English-only students on the 2005 NAEP reading and math tests for Grades 4 and 8 (Rumberger & Gándara, 2008). Nakamoto, Lindsey, and Manis (2007) modeled reading growth for Spanish-speaking students with LEP, and found that students with LEP who initially showed reading comprehension abilities in the normative range over Grades 1-2 did not necessarily remain in the normative range throughout elementary school.

Despite these lags, as soon as students are classified as proficient in English, many education supports are withdrawn (Rumberger & Gándara, 2008). Students classified as LEP at any point in their education have a reduced probability of accessing the language necessary to perform on measures of achievement. Unlike SWDs, students with LEP only have 5 years to gain the skills necessary to access the curriculum effectively before monitoring is removed. Similar to the SWD subgroup, the population of students that makes up the LEP subgroup is constantly changing (e.g., Abedi, 2004; Hopkins, Thompson, Linquanti, Hakuta, & August, 2013), which presents problems for assessing student growth. As students gain English proficiency they are exited from the program, which makes it difficult for the gaps between LEP and non-LEP students to ever be closed under the NCLB accountability framework.

This problem is magnified when a student with LEP is also a member of other subgroups. For example, Zehler et al. (2003) found that three quarters of district coordinators in districts that had SWDs receiving LEP services lacked sufficient numbers of teachers qualified to serve this group of students, which represented 9% of the LEP population in school year 2000-2001. A total of 24% of teachers in Grades K-12 worked with at least one student who had membership in both subgroups (Zehler et al., 2003). The effects of being a student with LEP may also interact with socioeconomic status. In a descriptive study of students with LEP in California, Gándara and Rumberger (2006) found that even if a student initially seemed to meet high standards once reclassified as fluent English proficient, they might fall behind native English speaking peers in secondary grades. Family income of non-native English speakers in California was 0.6 standard deviations below that of native English speakers, and Gándara and Rumberger (2006) did not investigate whether students with LEP underperformed their peers in secondary grades due to language proficiency or socioeconomic status.

Although the causal factors behind former LEP students with initial success later underperforming native English speakers are unclear, Gándara and Rumberger's research demonstrated the importance of including students formerly classified as LEP as when investigating the achievement of students. In addition, including student socioeconomic status when investigating student achievement can help to clarify whether or not the needs of former LEP students can be attributed to lingering issues of language proficiency, or other factors.

Economically disadvantaged students. For AYP accountability purposes, economically disadvantaged students are identified through eligibility for free or reduced price lunch (FRL), which serves as a proxy for low-income status (Aud et al., 2011). A school is determined high-poverty if more than 75% of students are eligible for FRL. In

Oregon, the number of students eligible is increasing. Approximately 42% of students in Oregon were eligible to receive FRL for school year 2006-2007, as compared to approximately 31% in 1995-1996 (ODE, 2007a). Nationwide, the number of students identified as economically disadvantaged is also increasing: during school year 2008-2009, 19% of 5- to 17-year-olds lived in poverty, as compared to 15% in 2000 (Aud et al., 2011).

Although a great deal of research has been conducted on students with low socioeconomic status (SES), the construct itself is multi-dimensional, and membership varies much like disability status and LEP status (Reardon & Robinson, 2008). More recent research on SES focuses on a combination of family income, parental education, and parental occupation rather than father's education alone – the key operationalization of SES in research conducted in the 1960s and 1970s (Sirin, 2005). Parental income, specifically, reflects the potential for social and economic resources available to the student, albeit flexible over time (Sirin, 2005). FRL eligibility, the identifier for economically disadvantaged students under NCLB, is a proxy for parental income.

In a 2005 review of the literature on SES and academic achievement from journal articles published between 1990 and 2000 including 101,157 students from 128 school districts and 74 independent samples, Sirin identified a medium to strong relationship between SES and academic achievement, contingent on school level, minority status, and school location. Correlations were strongest with mathematics achievement, as opposed to verbal and science achievement (Sirin, 2005). However, in a multi-level investigation of academic growth on NAEP at the state level, Lee & Reeves (2012) determined that percent of economically disadvantaged students was not a significant predictor of academic growth for Grades 4 and 8, respectively.

Economic disadvantage impacts students at the school and district level, and can disproportionately influence non-White students. For example, Aud et al. (2011) found that across the country, greater percentages of Black, Hispanic, and American Indian/Alaska Native students attended high-poverty elementary and secondary public schools than did White or Asian/Pacific Islander students in 2008-2009. Aggregating this subgroup to the school and district level, as well as examining the effect of SES on individuals, makes a great deal of sense: Sirin (2005), found that SES was a stronger predictor of academic achievement for White students than non-White students, neighborhood and school. However, the weaker SES-achievement correlation was partly because non-White students had a greater chance of living in neighborhoods and attending schools with additional educational risk factors (Sirin, 2005). NCLB's subgroup reporting for economically disadvantaged students would benefit from factoring in the composition of the schools and districts assessed, rather than focusing solely on student-level information.

The group of students eligible to receive FRL changes from year to year, and concerns exist regarding potential long-term effects of economic disadvantage on student achievement, even for students who are not members of this subgroup every year (McLoyd, 1998). As such, consideration must be given to including students in this subgroup if they have *ever* been identified as eligible for FRL rather than just students who consistently receive these services.

Summary of student subgroups. Clearly, the requirements for reporting performance for subgroups of students under are an advancement in the NCLB accountability system. However, as noted in the sections above, a number of issues need to be addressed when analyzing outcomes from subgroups. Some of these issues are definitional, some are transitory in prevalence, and some are about access and participation. At the same time, research has proceeded, both before and after passage of NCLB, that indicates subgroup membership influences achievement. Most of this research has been done in the context of large-scale tests, using a status model and focusing on achievement gaps. The student-level factors that attribute them to these four subgroups are critical components for further investigation using growth models to ensure that students within these subgroups are not, in fact, being left behind.

District-Level Factors

Although the language of the NCLB legislation focuses on schools as the primary unit of analysis and accountability, districts set agendas for and disseminate funding to schools. In addition, accountability aggregates from the schools to districts, and eventually to states (NCLB, 2002). As a result, the effects of districts are relevant in investigations of student outcomes (Dee, Jacob, & Schwartz, 2013).

Districts are at greater risk than most schools for failing AYP in the area of student performance because larger numbers of students tested means a greater likelihood of having to report subgroup scores. Increased diversity may also result in less likelihood of meeting AYP at the district level (Rose, 2004). Aggregation of results to the district level may mask some of the differences in performance across schools in a district (e.g., Figlio & Ladd, 2008). However, by focusing solely on schools, the ability to properly assess the performance of student subgroups can be reduced in instances of subgroup membership below AYP reporting requirements (Fulton, 2006). In summary, by investigating district-level factors it is possible to properly document how student outcomes may be related not only to student-level factors but also to the presence of these aggregated subgroups.

When making comparisons between districts based on student outcomes, a variety of factors must be considered to help explain variance at both the student level and between districts. The location of a district (e.g., urban, suburban, rural) and number of students served, as well as the percentage of non-White, LEP, SWD and FRL eligible students are all variables of interest with potential to explain variance between districts. In addition, the financial investment of school districts in terms of total funding and funding for SWDs may influence district outcomes. A discussion of each follows.

District physical demographics: location and size. Twelve district location types are delineated in the School District Universe Survey for the Common Core of Data, distinguished by both location and population. These locations fall into four broad categorizations: cities, suburbs, towns that are not directly connected to urbanized areas, and rural areas. A district's location can influence both the opportunities available for student services, as well as the funding of a district.

For example, urban districts are often underfunded by state governments, and lack a sufficient local tax base to make up the difference (Rentschler, 2005). NCLB (2002) legislation was designed such that states and districts were not required to fund additional costs incurred by the requirements, resulting in insufficient funding (Gordon, 2008; Mathis, 2003). Rural districts may not be as underfunded by the state; yet, they often lack the resources present as a result of increased population and corresponding higher numbers of potential experts in urban districts. In spite of this potential lack of breadth of expertise, overall, teachers in rural schools average more years of experience than those in urban schools (Provasnik et al., 2007). One consequence of urban diversity not faced by rural schools is that urban school districts are more often required to report scores for NCLB subgroups, subjecting them to greater federal accountability pressures (Rentschler, 2005).

The effects of district location may not be the same for all groups of students. For example, Ladd and Ferguson (1996) used hierarchical linear modeling to determine that differences in test scores between African American and White students were larger in more urban districts, even after controlling for school and community inputs. Some of the relationship between SES and achievement varies by district location as well. For example, in a meta-analysis of ten years of studies examining SES, Sirin (2005) found the average effect size of SES on academic achievement was stronger for students in suburban schools (0.28) as compared to rural schools (0.17, $p < .001$).

Similar to physical location, the number of students a district serves can influence the availability of supports and services. The larger a district's student population, the more per-pupil revenue it receives from the state. As the number of students increases in a school district, the quantity of services available does as well as a direct result of increased funding. Provasnik et al. (2007) found that as a result, adjusted public school expenditures per student were higher in rural areas than in cities, suburbs, and towns, and that rural areas had a smaller smaller-teacher ratio than other locales. Students who received SpEd support services in rural areas were served by staff who had lower student caseloads than service providers in urban districts (Provasnik et al., 2007).

Evaluating the spending of districts nationwide, Chambers, Parrish, Esra, and Shkolnik (2002), found that districts with fewer than 2,500 students reported a 14% higher level of actual expenditure per SWD than that of districts with enrollment of 25,000 or more students. Although larger districts may not need to spend as much per pupil to fund necessary student services, increased school size also comes with a greater likelihood of needing to report disaggregated NCLB subgroups, putting districts at risk for greater accountability pressures. In comparison, small districts can face problems with increased year-to-year variability of scores (e.g., Zvoch & Stevens, 2006b), as assessment results are compared to students who took the exam in the previous school year (NCLB, 2002).

Although small districts may need to spend more to provide equivalent services, attending school in a smaller district was related to student test scores. For example, Driscoll, Halcoussis, and Svorny (2003) examined the effect of district size on student performance as measured by standardized test scores in California, and found that after controlling for student characteristics, class size, school size, and population density, students attending elementary and middle school in larger districts did not perform as well as those from smaller districts. Investigation of both district size and location using growth models is necessary to determine their potential influence on growth for SWDs.

District student demographics. In addition to district size and location, the demographic makeup of the students that districts serve can influence student growth. Given the potential NCLB sanctions associated with failing to meet AYP, attention to the four NCLB subgroups provides a logical starting point to use to examine student growth at the district level. The proportions of SWDs, students with LEP, students eligible for FRL, and non-White students in a district may influence the performance of SWDs above and beyond individual student characteristics. Along with total proportion of SWDs in a district, the proportion of SWDs who spend less than 40% of their time in the general education setting may be a factor, as these students may require more resources to educate than SWDs in more inclusive settings. If a district has a large percentage of SWDs requiring additional services to access education, this may impact the performance of SWDs in the district as a whole.

As an example of how aggregated demographics may influence student growth, when assessing the mathematics achievement of middle school students longitudinally across schools, Zvoch and Stevens (2006a) found the between-school effects for the proportion of non-White, FRL recipient, LEP, and SpEd students each had a significant negative influence on school mean achievement (b = -8.98, -9.37, -18.48, and -36.11 respectively, $p < .001$). The proportion of non-White and SpEd students also had a significant negative effect on school mean growth (b = -1.59, $p < .01$ and -4.15, $p < .001$, respectively). While assessed at the school rather than district level, each of these covariates had a negative influence on student growth (Zvoch & Stevens, 2006a), that will likely be present at the district level as well.

District funding. Neither IDEA nor NCLB have ever been funded adequately by the federal government (Mathis, 2004; Moores, 2005). Congress was authorized to fund 40% of the excess cost of educating SWDs in each state, but as of 2005 had never funded even 20% of those costs (Moores, 2005). As a result, states and districts have had to pick up the slack and fund both NCLB and IDEA requirements.

Questions about whether or not the quantity of money input into students' educations can impact educational outcomes date back to the Coleman report (1966). Between 1967 and 1987, a total of 38 studies examined the relationship between economic resources and student achievement using the production function approach. Meta-analyses conducted by Hanushek (1989) and by Hedges, Laine, and Greenwald (1994) using the same dataset came to opposite conclusions: Hanushek found that variations in expenditures were not systematically related to student performance, whereas Hedges, Laine, and Greenwald found per pupil expenditures and student outcomes were related. Specifically, Hedges, Laine, and Greenwald found that a $500 increase in per-pupil funding would be associated with a 0.7 standard deviation increase in student achievement.

The absence of a consensus about what appropriate educational outcomes are and challenges with identifying and measuring inputs complicates measuring the economic productivity of education (Rice & Schwartz, 2008). District resources include a great number of different inputs, including teacher salaries, school building maintenance, transportation, computers, and school supplies, among others. Because the inputs are so diverse, and quality of data vary greatly in all the categories, spending data is often used as "a summary measure of the amount of inputs used… which easily yields estimates of the improvements associated with each dollar spent" (Rice & Schwartz, 2008, p. 136). How funds are categorized may vary across districts, making total district funding a more appropriate measure of comparison.

Prior research has demonstrated that increases in education funding do not necessarily influence educational outputs (e.g., Hanushek, 1989; Hanushek, 2003; Wenglinsky, 1997), and that finance reform does not guarantee improved student achievement (Corcoran & Evans, 2008). However, investigating the potential effect of district funding on student achievement is critical in the context of student growth. Such insights may help districts set priorities with respect to raising educational outcomes for students.

Total district funding. Within states, there is a great deal of inequality in per-pupil expenditures across districts. For example, Corcoran and Evans (2008) used two panel datasets of school districts from 1972 to 2004, to compute measures of inequality in per-pupil expenditures. Overall, 60-70% of variance in spending was between states, with approximately 27% of the variance within-state in 2004. Substantial variation in per pupil expenditures across states was also observed by Zhou (2010), ranging from $3,886 in Utah to $11, 572 in New York for fiscal year 2008. This follows the findings of Ladd (1996), who demonstrated with a cross-sectional analysis using the 1990-1992 Common Core of Data that some districts needed to spend more money than others in order to provide an equivalent level of educational services. Although the greatest portion of variance was between-states, there is still substantial variance in spending across districts within states.

In a review of 10 states that had calculated the costs of bringing all children to NCLB standards, Mathis (2003) found seven of ten states projected cost increases of more than 24%. Furthermore, the federal government did not adequately compensate these costs. Dee, Jacob, and Schwartz (2013) used longitudinal data from roughly 10,000 unified school districts observed over 14 school years in an interrupted time-series design to determine that district spending increased by roughly $570 per student (2009 dollars) by 2008 (post-NCLB). The majority of these increases ($430 per student) funded instruction, with the rest going to per-pupil support services, as well as to increase salaries for teachers in high-poverty school districts (Dee et al., 2013). All of these increases were funded at the state and local level. Overall, district spending increased dramatically post-NCLB, even without factoring in the true cost of bringing students up to standards.

Additional funding pressures beyond NCLB and IDEA requirements exist in Oregon as a consequence of Ballot Measure 5, passed November 6, 1990, which reduced local district funding from property taxes. As a result, state funding necessarily increased for districts (Oregon Legislative Revenue Office, 2004). For school year 2007-2008, national average public school system revenue was divided as follows: the federal government funded 8.2%, states 48.3%, and local governments 43.5% (Zhou, 2010). In Oregon, the federal government funded 9.1%, the state 52.3%, and local governments 38.6% (Zhou, 2010).

Within states, considerable variation in district spending exists, part of which is related to median family income. Chambers, Parrish, Ezra, and Shkolnik (2002), found that districts whose families' incomes were lowest spent less on education per student. Districts with middle-income families spent $2,314 per student more in 1999-2000 than districts with the lowest-income families. The Federal Education Budget Project (2015) compared the per-pupil expenditures for 2010 to 2012 of all Oregon school districts to the state average, and weighted these according to size and poverty level. The resulting measure of school finance inequity among districts increased from 9.8% in 2010 to 12% in 2012. Districts can either try to do the best they can with available resources, or they can legally challenge the state to sufficiently fund education (Mathis, 2004).

In addition to direct district-level funding, Oregon districts are a part of Education Service Districts (ESDs) that provide services to districts that may be too costly for a single district to fund. Districts can coordinate programs and services through the ESDs and moderate their individual costs. ODE publishes this dollar amount annually, which must be taken into consideration when calculating total funding for a school district.

Special education funding. In school year 1999-2000, total spending for an average student receiving SpEd services was 91% higher nationally than for the average SWoD receiving no SpEd services (Chambers, Shkolnik, & Pérez, 2003; Harr et al., 2008). As a whole, SpEd funding makes up over 20% of all school funding nationwide (Hanushek, Kain, & Rivkin, 2002). In Oregon, SWDs made up approximately 13% of total public school enrollment as of school year 2008-2009 (Aud et al., 2011). Three major components make up total Oregon Special Education funding: district funding, federal IDEA Part B funding, and ESD funding.

Oregon district funding consisted of both state and local sources, and has been audited annually for Maintenance of Effort [MOE] in order to determine that the school district supports SWDs at the same level as previous years (ODE, 2013). Select services for SWDs in some districts have been provided by ESDs, although the specifics of these services vary across the state (ODE, 1999-2013). The district contribution combined with the ESD contribution together have typically comprised MOE (ODE, 2013). In an analysis of SpEd funding across the country, Chambers, Parrish, Esra, and Shkolnik found that the district spending ratio for SWDs relative to SWoDs in smaller districts was estimated to be greater than the national average, at 2.19 (2002).

Federal IDEA funding comprises a relatively small portion of total special education funding, as Congress has not adequately funded the legislated mandates (Harr et al., 2008; Mathis, 2004; Moores, 2005). While general funds have made a contribution to the education of SWDs, it has often been quite minor. For example, across disability categories, Special Education expenditures for 1999-2000 nationally ranged from 63% to 91% of the total average education expenditures for SWDs, depending on disability type (Chambers et al., 2003). Regardless of the state, Chambers, Shkolnik, and Pérez (2003) found considerable variation in average spending per SWD across districts, which was not always related to student poverty or percentage of high-cost disabilities in the district. For example, a previous study by Chambers, Parrish, Esra, and Shkolnik (2002) demonstrated spending on SWDs in smaller districts nationwide was 119% higher than the national average. The majority of funding for SpEd funds in Oregon fell into two categories across districts that accounted for more than 60% of school spending on SWDs across the country (Chambers et al., 2003): instruction and support services.

Within the category of instruction, the majority of funding was intended to establish a reasonable pupil-teacher ratio and for teacher salaries (Dee et al., 2013; Lee & Reeves, 2012; Levenson, 2012). Levenson (2012) found twice the level of variation in SpEd spending as there was for SWoDs across school districts nationwide. Additional instructional costs were incurred to educate SWDs requiring more restrictive placements based on the services students needed to appropriately access education. In Oregon, a separate budget line item has typically existed within the broad category of instruction specific to students in restrictive placements (based on health, behavioral, or other perceived need). Students in restrictive placements comprised approximately 15% of Oregon SWDs in school year 2007-2008 (Aud et al., 2011), and have been distinguished from their peers in more inclusive settings in Oregon test data files.

Supplementary services required to assist SWDs have included, but are not limited to, resource specialists, related services, and community based services (Chambers et al., 2003). Resource specialists either pull SWDs out of the general education setting or go into them to provide students with supports. Related services also have included a variety of specialists in areas such as speech/language, physical/occupational therapy, vision, audiology, psychology, social workers, personal health aides, and other providers. Community-based services have included extended-time (e.g. before- or after-school care) and summer school. Transportation has been a specific support service as a key part in ensuring education for SWDs is accessible. This has been particularly challenging in rural areas, as more students travel greater distances to school and SWDs may need to travel to access appropriate educational supports. In an analysis of national transportation data, Killeen and Sipple (2000) found patterns of per-pupil transportation expenditures in rural districts were twice as high as urban districts, and 50% more than suburban districts.

One problem with using special education funding averaged across SWDs in the district is that it can mask some of the variability in instructional costs and related services for SWDs across a district (Harr et al., 2008). The average expenditures for SWDs across states based on 1999-2000 Special Education Expenditure Project data illustrated that expenditures for a SWD range from 1.6 (specific learning disability) to 3.6

(multiple disabilities) times the average expenditure for a SWoD (Chambers et al., 2003). This finding suggests that other metrics should be considered beyond average expenditures per SWD when assessing the influence of SpEd funding.

Situating analyses of student outcomes in the context of federal accountability requirements, along with the resources districts have to support required educational progress, can provide a more complete story than student data alone. District funding in the context of district demographic variables affords a better understanding of what is taking place in the district to educate *all* children, whereas SpEd funding approximates a district's commitment to educating SWDs. Including demographic *and* fiscal covariates can help to determine whether a district is being provided insufficient resources to make progress given their student population (Figlio & Ladd, 2008).

Individual Growth for Students with Disabilities

Three main sources of data are available for investigating student growth: large national databases (e.g., Wei, Blackorby, & Schiller, 2011; Wei, Lenz, & Blackorby, 2012), curriculum-based measures (CBMs; e.g., Deno, Fuchs, Marston, & Shin, 2001; J. Shin, Espin, Deno, & McConnell, 2004), and statewide accountability measures. National databases make it easy to generalize to the country as a whole, but these gains may come at the expense of specificity, and cost to access restricted-use data may in some cases be prohibitive (Strayhorn, 2009). CBMs were not designed to be part of accountability systems, and without accounting for differences in scale across years, limit investigations of growth to within-year studies (Deno et al., 2001). State accountability measures, in contrast, were designed with vertically linked scales that lend themselves to longitudinal investigation of student growth.

Unlike the status-based accountability system applied to evaluate AYP under NCLB, investigating individual student growth provides incentives for districts to focus attention on the entire student body, rather than just the disadvantaged students (Ladd & Walsh, 2002). Although applicable to the SWoD population, the range of academic performance within the SWDs subgroup makes examining student growth especially appropriate. The extent to which students who are academically successful exit SpEd services limits observed gains for the SWD subgroup (e.g., Schulte & Villwock, 2004; Ysseldyke & Bielinski, 2002). Use of student growth as a metric for AYP subgroup accountability can help to address the fact that, on average, SWDs start with scores lower than their SWoD peers (e.g., T. Shin, Davison, Long, Chan, & Heistad, 2013), but still are held to the same status-based accountability standards (Chudowsky & Chudowsky, 2009; Eckes & Swando, 2009). A focus on student growth can also address the intent of NCLB to monitor the progress of underperforming students without exacerbating the problem of measurement error associated with student subgroup size (Figlio & Ladd, 2008).

When longitudinally modeling student growth over time, students are compared to themselves rather than to others. This can be more informative than a status model, which is likely to misjudge the performance of schools and districts (Figlio & Ladd, 2008). To look at the growth of individual students over time gives insight into the effectiveness of the education they receive. Failing to consider growth in an accountability system means that students who are already above proficiency, as well as those so far below the benchmark that teachers deem them a lost cause, are excluded from efforts to improve scores (Goldhaber, 2002). Additionally, accountability systems based solely on year-to-year gains rather than long-term student growth "may interfere with the ability of schools to make the investments in capacity that could potentially led to higher student achievement in the future" (Figlio & Ladd, 2008, p. 175). Use of a longitudinal growth model to evaluate student progress can help to illuminate potential long-term educational outcomes for students that a status model can not.

Modeling student growth necessitates assessments with a scale vertically linked scale grades to allow for comparison of student scores across time. While it may be tempting to evaluate student growth solely from the perspective of state NCLB proficiency levels, as that is what concerns stakeholders the most, Dunn and Allen (2009) champion a combination of student growth via vertically scaled scores in addition to a model including status to accurately evaluate which schools truly have at-risk student populations. Using growth as a metric, students who are below proficiency cut-points can still demonstrate annual achievement gains (Ladd & Lauen, 2010). Within a status model, such gains are hidden.

Assessing growth for SWDs using a vertically scaled measure aligned to state content standards can give a more detailed picture of how students are performing, on average, within the context of their educational programs. In the post-NCLB era, student gains in both reading and mathematics are a constant topic of conversation for AYP. Additionally, the performance of SWDs is a topic of concern, with respect to subgroup accountability (Eckes & Swando, 2009). Few studies have been conducted on growth for SWDs in reading and mathematics using state accountability measures; yet, these investigations of student growth provide insight into what is taking place beyond NCLB status-based claims. This is particularly important for the population of SWDs, many of whom may demonstrate growth but still not be identified by a status model as making progress.

Reading growth. Few studies exist that investigate SWDs reading growth using accountability measures as outcome variables. Some of those do not generalize to the entire AYP subgroup, as they only include students with specific disability types. For example, in a study of 461 students with learning disabilities (LD) across 5 years in one district, Schulte, Villwock, Whichard, and Stallings (2001) found that for the two years of the study that North Carolina's state growth standards were available, average growth for students with LD in the Grade 3 cohort did not meet the growth standard, whereas students in Grades 4 and 5 met the standard, and in some cases exceeded it. A student "met" the state growth standard by having a end of year gain score (calculated by subtracting previous year performance from present year performance) to an expected progress standard. However, the relatively large variance in the observed growth rate (*SD* = 4.3 to 8.2 scale score units) indicated that growth for students with LD was heterogeneous (Schulte et al., 2001). Instead of relying on a proficiency rating to determine student progress, the authors utilized a measure of growth (gain score) to evaluate student progress.

Later investigations of student reading growth using end of year state assessments as outcome measures included SWDs as a whole rather than only subsets of this AYP subgroup, and explored longitudinal student growth as opposed to gain scores (e.g., Schulte, 2010; Schulte & Villwock, 2004). Schulte and Villwock (2004) cumulated results from three longitudinal cohorts of students identified as SWDs at any point during the study, and found steady positive reading growth across all grades in all schools, with substantial differences in Grade 5 outcomes across the six schools F (5, 220) = 4.41, $p <$ .001. Introducing students' Grade 3 achievement level (initial achievement) as a covariate accounted for approximately 45% of the variance in achievement at Grade 5 [F (5, 196) = 188.89, $p < .001$], whereas differences across schools were no longer significant F (5, 196) = 0.93, $p >.05$. Although the percentage proficient for AYP for Grade 5 at these schools ranged from 50 to 91%, SWDs in all six schools demonstrated growth. Investigations of reading growth for SWDs allow for a more nuanced view of student progress than relying on a status model alone.

To reinforce the importance of examining growth for SWDs, Schulte (2010) examined the growth two cohorts of students across Grades 3-5, from 1,347 schools and found that tracking SWDs using residual gain scores was a better reflection of their reading performance than the NCLB status model. In addition, Schulte found growth for SWDs was similar to that of students in the SWoD population (r = .24 and .28, respectively), even though initial reading achievement levels were quite different.

Extending studies of reading growth over Grades 3-8 (the span of consecutive NCLB accountability years) will inform the field about the actual performance of SWDs above and beyond the simple metric of the NCLB status model. Longitudinal investigation of student growth allows for opportunities to investigate nonlinear growth, and to learn more about the progression of students in reading over these elementary and middle school years.

Mathematics growth. Similar to reading, a limited number of studies investigate growth of SWDs in mathematics growth using accountability measures as outcome variables. Zvoch and Stevens (2005) investigated sample exclusion and student attrition effects on estimates of student, school, and district mathematics performance across grades 6-8 by comparing a complete cohort sample of all students who ever took a mathematics assessment ($N = 6{,}098$, SWDs = 1,092) to an accountability sample including only students who attended the same middle school for all three years and had complete test data from regular test administration ($N = 3{,}334$, SWDs = 101). In both samples, overall student initial achievement status was a positive predictor of mathematics growth rate ($\beta = .06, p < .001$ and $\beta = .02, p < .05,$ respectively). However, SWD status in the accountability sample was not a significant predictor of mathematics growth rate ($\beta = 1.10$) whereas in the complete cohort SWD status was a negative predictor of mathematics growth rate ($\beta = -2.81, p < .001$).

Schulte and Stevens (2015) demonstrated that SWDs who were identified as *always*, *ever*, or *Wave 1* recipients of SpEd services from Grades 3-7 differed in mathematics performance from SWoDs. Initial status for students receiving SpEd services was 6.21, 5.30, and 5.01 scale score points below SWoDs at Grade 3 ($p < .001$) depending on classification. The rate of initial change for SWoDs was approximately 7 points a year, regardless of how students receiving SpEd Services were identified, with initial change for SWDs *always* in SpEd at 0.92 points lower, *Wave 1* SpEd at 0.64 points lower and *ever* in SpEd at 0.39 points lower than SWoDs. Quadratic deceleration for SWoDs was approximately -0.05 points per year, and the only group of SWDs that was significantly different were the *ever* SpEd group, which had a small moderation in deceleration.

Stevens, Schulte, Elliott, Nese, and Tindal (2015) explored the mathematics growth of 92,045 students with and without disabilities over Grades 3-7, classifying

SWDS into seven separate exceptionality categories. Students from all exceptionality categories had initial mathematics performance that was lower than for SWoDs. After accounting for exceptionality type and student-level demographics, the average initial growth rate was 6.99 scale score points per year, with an average deceleration of 0.54 points per year. There were also significant differences in initial growth and deceleration rate for some exceptionality types. Student race/ethnicity, sex, parental education, free/reduced lunch status, and English language proficiency were found to be significant predictors of achievement.

Across both reading and mathematics, the limited literature on the growth of SWDs indicated that although their initial achievement fell below SWoDs, growth rates for both groups were similar (e.g., Schulte, 2010; Schulte & Stevens, 2015; Schulte & Villwock, 2004; Schulte et al., 2001; Zvoch & Stevens, 2005). Investigations of student achievement using a status model would characterize SWDs as a consistently low-performing group based on a proficiency metric. Inferences drawn from cross-sectional models "of the achievement gap frequently reported in the research and policy literature may underestimate academic growth for SWDs" (Schulte & Stevens, 2015, p. 19). A focus on longitudinal growth is necessary to characterize gains for SWDs across years, and more appropriately demonstrate the academic gains of this subgroup.

Summary and Study Context

It is critical for school districts to be able to determine where they can devote resources in order to improve student achievement, particularly within the confines of NCLB accountability. The analysis of subgroups of students, made specific (while flexible) under NCLB, as well as the additional resources (both financial and instructional time) necessary to ensure appropriate access to education makes the SWDs subgroup a key constituent in such an accountability system. Including student-level inputs when evaluating a student's growth accounts for the varied experiences students bring to their education process (Rice & Schwartz, 2008). Additional investigation of the district-level demographic and funding factors that may influence student growth is necessary in order for districts to determine what areas on which to focus immediate attention. Providing districts with insights to help make determinations of how to allocate limited resources to improve student growth is an important resource for improving educational outcomes for SWDs. Conducting an investigation into student growth using scores from a vertically scaled measure across Grades 3-8 allows for a more nuanced look at growth for SWDs, as well as the district-level factors that may influence that growth in the elementary/middle school years.

Research Questions and Contribution

The purpose of this dissertation is to assess the extent to which district financial investments are related to the growth of SWDs from Grades 3-8 in reading and mathematics above and beyond what is predicted by district demographic characteristics. To evaluate these relationships, the following research questions will be addressed:

1. Are individual student demographic variables (e.g. race/ethnicity, LEP status, FRL eligibility) related to reading and mathematics growth for SWDs for Grades 3-8 above and beyond the influence of time eligible to receive SpEd services?
2. Do district-level demographics (e.g., size, percent non-White, SWD, LEP, FRL-eligible students) moderate the effect of time eligible for SpEd services on student growth?

3. Does district-level funding (total and SpEd funding) moderate the effect of time eligible for SpEd services on student growth?

Discerning these relationships was be a three-step process: first, the impact of student demographics was examined, followed by district demographic information, and finally adding funding to the model. These research questions were not investigated to seek causal predictors, but to discern which student and district-level factors were related to growth for SWDs.

CHAPTER II

METHODS

This study used extant data from a total of 54 data sets from three different sources collected for the 2006-2007 through 2011-2012 academic years. The Oregon Assessment of Knowledge and Skills (OAKS) reading and mathematics data for all school districts were provided by the Oregon Department of Education through a federal grant funding the National Center on Assessment and Accountability for Special Education (NCAASE). District demographic information including the number of SWDs in each district and how remote the district was in location were obtained from the Common Core of Data's Local Education Agency (School Universe) Survey Data for the same academic years. Additional district-level demographic and financial information was drawn from a series of annual reports published online by ODE.

Sample

Students' data were assigned to the district in which they *received* their instruction in (their attending district), which was not necessarily the district for which their test scores "counted" for AYP (their residential district). Less than 1% of students had attending and residential districts that did not match. This sample included students who initially took OAKS in Grade 3 and also those who took Oregon's Extended Assessment (ORExt) in Grade 3 but took OAKS in later years (only OAKS scores for these students were used). This sampling strategy provided a more comprehensive look at SWDs in Oregon than if the cohort had been limited to only students with a valid OAKS score for Grade 3.

The preliminary sample for reading included 51,908 students with at least one valid score on the Oregon Assessment of Knowledge and Skills (OAKS) reading test from Grades 3-8 (see column 1 of Table 1). This group of students was approximately 51% male, 35% non-white, and included approximately 20% SWDs. Approximately 17% of students were limited in English proficiency (LEP) at any point during Grades 3-8, and 61% received free or reduced lunch (FRL) in at least one year. For mathematics, the initial sample included 51,030 students (see column 1 of Table 1). A slightly higher percentage of these students were eligible for FRL (62%) or received services for LEP (17%) at any point during Grades 3-8 than in the initial reading sample.

Analytic sample. Students from both initial samples were systematically excluded to construct samples appropriate for analysis at both the student and district levels (see column 2 of Table 1). Students with non-standard grade progression (e.g., those who were promoted or retained in any year; n = 570, <1% for reading, n = 606, <1% for mathematics) also were removed, as they went through their education in a manner systematically different than other students. Students missing any demographic predictor variables (sex, ethnicity/race, SWD, FRL, LEP; n = 332 for reading, n = 342 for mathematics) made up less than 1% of the initial sample and also were eliminated prior to analysis (Enders, 2010).

Students who switched district more than once *and* had less than three years of data in a district also were removed prior to analysis (n = 245,< 1% for reading, n = 256, <1% for mathematics) to most appropriately nest student characteristics within attended districts. Districts with fewer than 15 students (n = 314, <1% for reading, n = 287 <1% for mathematics) were eliminated from analyses to avoid imprecise parameter estimates at the district level (American Institutes for Research, 2012; McCaffrey, Sass, Lockwood, & Mihaly, 2009). Finally, students tested with Oregon Department of Education or one of the Extended Services Districts listed were eliminated from the sample (n = 1,104, 2% for reading, n = 1,026, 2% for mathematics) as these entities are not actual Oregon "districts" and did not have available district-level data for purposes of this study (funding variables, demographics). Overall, 4.49% of students were eliminated for reading, and 4.41% for mathematics.

The analytic sample detailed in Table 1 included 49,578 students from 148 districts for reading and 48,779 students from 149 districts for mathematics. For both subject areas, approximately 20% of students were eligible to receive SpEd services at some point during Grades 3-8, 51% were male, and 17% were eligible to receive LEP services at some point during Grades 3-8. Students eligible for FRL made up 61% of the analytic sample for reading and 62% for mathematics.

To explore the differences between the initial and analytic samples, Cohen's h, a measure of effect size for proportions, was computed to determine how different the initial and analytic samples were from each other. Using Cohen's rules of thumb for h to interpret effect sizes (Cohen, 1988, pp. 184-185), all values of h were well under the .20 designation of a *small* effect, ranging from a h of 0.00 to 0.01. This demonstrated that the compositions of the analytic samples were not substantially different than the initial samples.

Table 1

Student Demographic Characteristics by Sample and Subject

Characteristic	Total Sample		Analytic Sample		
	N	%	*N*	%	*h*
Reading					
n	51,908		49,578		
SWD	10,227	19.70	9,692	19.55	0.00
Male	26,535	51.12	25,300	51.03	0.00
FRL	31,889	61.43	30,477	61.47	0.00
LEP	8,396	16.17	8,223	16.59	0.01
Non-White	17,887	34.46	17,405	35.10	0.01
Mathematics					
n	51,030		48,779		
SWD	10,019	19.81	9,594	19.67	0.00
Male	26,101	51.15	24,925	51.10	0.00
FRL	31,543	61.81	30,160	61.83	0.00
LEP	8,429	16.52	8,248	16.91	0.01
Non-White	17,587	34.46	17,103	35.06	0.01

Cohen's *h* also was computed to determine if the group of SWDs within the analytic samples differed in demographic composition from SWoDs (see Table 2). There were very small differences in composition between groups in the categories of LEP and non-white. For reading, approximately 17% more SWDs were male ($h = 0.34$), and 13% more SWDs received FRL ($h = 0.29$) than SWoDs. For mathematics, SWDs included 17% more males ($h = 0.35$) and 12% more SWDs were eligible for FRL ($h = 0.25$). These differences are between *small* and *medium* using Cohen's rules of thumb, indicating disproportionate representation of students receiving FRL and males in the SWD group.

Table 2

Student Demographic Characteristics by Group and Subject

Characteristic	SWoD		SWD		
	N	%	*N*	%	*h*
Reading					
n	39,886		9,692		
Male	19,066	47.80	6,234	64.32	0.34
FRL	23,483	58.88	6,994	72.16	0.29
LEP	6,758	16.94	1,465	15.12	0.01
Non-White	14,100	35.35	3,305	34.10	0.03
Mathematics					
n	39,185		9,594		
Male	18,738	47.82	6,187	64.49	0.35
FRL	23,229	59.28	6,931	71.24	0.25
LEP	6,788	17.32	1,460	15.22	0.06
Non-White	13,844	35.33	3,259	33.97	0.00

Missing data. Students in the analytic samples had missing data for a variety of reasons. Some students switched test participation between OAKS and ORExt based on the recommendations of their IEP teams (ODE, 2013; Saven et al., 2015) and therefore test data were missing from the OAKS data in some years. Students also changed districts over the course of Grades 3-8. Where students attended more than one district, student data were investigated to determine the district in which the student attended the majority of the time and that was assigned as their "majority" district. In cases of equal membership between two or more districts, their first district attended was used. Test data were eliminated for specific years in which students were tested in a district outside of their "majority" district.

Approximately 53% of students had a valid test score for every grade in reading (n =26,007) and 55% of students in mathematics (n = 26,778). In both subjects, approximately 11% were missing one test score (n = 5,669 for reading, n = 5,181 for mathematics), approximately 10% were missing two years of data (n = 5,669 and 4,991) as well as three years of data (*n* = 4,941 and 4,856). For reading, approximately 8% of students were missing four years of data (*n* = 3,731) and also five years of data (*n* = 4,111). For mathematics, approximately 7% of students were missing four years of data (*n* = 3,584) and five years of data (*n* = 3,380).

Measures

The Oregon Assessment of Knowledge and Skills (OAKS) served as the outcome measures in reading and mathematics for Grades 3-8. At that time, NCLB (2002) accountability for Oregon hinged on these scores. The majority of SWDs took the vertically scaled OAKS measures with or without accommodations, although a small percentage of SWDs with significant cognitive impairments took ORExt instead (1% of the total student population). For this study, only the OAKS measures were used, because ORExt was neither vertically scaled nor equated with OAKS (Oregon Department of Education, 2012b).

Oregon Assessment of Knowledge and Skills. The state of Oregon began developing a comprehensive assessment system including all students in 1996 (ODE, 2012b). The test was individually administered using a computer-adaptive testing system that presented items contingent on students' previous responses (ODE, 2012b). Each assessment included 40-50 operational multiple-choice items as well as 5-6 field test items, which were not included in student scores. Approximately 40,000 to 45,000 students per grade were tested annually. Students had three opportunities to take the test online each academic year, from October through May. Of these three potential administrations, the highest score was reported for accountability purposes.

The OAKS reading/literature assessment consisted of multiple-choice items corresponding to passages designed for three purposes: reading for literary experience, reading to gain information, and reading to perform a task (Choi & Tinkler, 2002). These measures were only administered in English, and were designed to assess six literacy skills: vocabulary, reading to perform a task, demonstrate general understanding, develop an interpretation, and examine the content and structure of both informational and literary text.

OAKS mathematics tests included 40 multiple-choice items focused on five state reporting categories aligned with the National Council of Teachers of Mathematics' *Principles and Standards for School Mathematics* (National Council of Teachers of Mathematics, 2006): Calculation and estimation; measurement, statistics and probability; algebraic relationships; and geometry. Mathematics assessments were available in English, Spanish, and Russian, with English and Spanish provided online.

Accessibility. Teachers and instructional teams who knew individual student strengths and weaknesses decided, in consultation with parents or guardians, which accommodations and/or modifications if any, were necessary for students SWDs. Students' IEP teams determined not only whether OAKS was appropriate for students (rather than ORExt) but also which state-authorized accommodations and/or modifications were needed to access the content appropriately. In addition to the online OAKS option, paper and pencil, dual language (Russian/English and Spanish/English), Braille and large print versions were available for students with specific testing accommodations requiring them (ODE, 2012b). Staring school year 2008-2009, the OAKS Online system included the ability to change screen magnification and to print reading passages, which reduced the number of paper and pencil tests required for students to properly access content (ODE, 2012b). To ensure items on OAKS functioned without bias for both LEP students and SWDs, items were analyzed for differential item functioning (ODE, 2007) and found to be functioning without bias.

Updated performance standards for mathematics and reading/literature were applied to OAKS for tests administered during 2006-2007 and beyond. Additionally, the K-8 mathematics content standards were revised in 2007, and implemented during school year 2010-2011. As the OAKS was constructed using a vertically linked scale represented by a Rasch Unit, or RIT score, measuring growth of individual students across time was possible (ODE, 2012b). Item difficulties generated using Rasch modeling were used to generate a raw score to RIT score conversion for student OAKS scores (ODE, 2014).

Reliability evidence. Several studies have been conducted on the consistency, stability, and accuracy of OAKS test scores documenting the standard errors of measurement, and classification accuracy, both overall and for subgroups (ODE, 2012b). Based on the operational data from school year 2003-2004, OAKS online provided similarly reliable test scores across the range of ability, except for the extreme ends of the distribution, relative to the paper and pencil test (ODE, 2007b). The standard error of measurement was, on average, 3 RIT points, except for students in the 99^{th} percentile, where the standard error increased.

A second study by the Northwest Evaluation Association (NWEA; 2005) demonstrated reliability of achievement classification, and indicated high classification accuracy for all reading and mathematics forms for 2003-2004 across all performance levels. Overall, reliabilities ranged from .84 to .99, with the majority of coefficients falling above .90. In a study using 2005-2006 data, Doran and Cohen also indicated each test form was similar in classification accuracy across the range of ability levels (McCall, 2005).

Test validity inferences. Oregon's test scores were correlated with three external tests measuring the same constructs. High correlations were present between the OAKS and the nationally normed California Achievement Test in all grades for both reading and mathematics, with the average correlation being $r = .77$ (Doran & Cohen, 2006). Additionally, for a random sample of 200 students from each of Grades 3, 5, and 8 who took both OAKS reading and mathematics tests in 1998, as well as the Iowa Test of Basic Skills, the correlations were high, ranging from .76 to .85 (ODE, 2001). High correlations also were present between the OAKS reading and mathematics tests and the NWEA tests in the same subject with correlations ranging from .71 to .81 in reading, and .66 to .83 in mathematics (ODE, 2001).

Such high correlations with three external tests measuring the same construct provided strong evidence for the concurrent validity of OAKS. Additional studies demonstrated comparability of tests administered using OAKS Online to its pencil-and-paper predecessor prior to full implementation. Correlations of the two measures at Grade 3 were .95 for reading and .96 for mathematics (McCall, 2005). Computer adaptive tests also provided more information about students at every level of achievement (ODE, 2007).

Data Preparation

Data from multiple ODE sources and the district-level Common Core of Data were merged to create a district-level data file consisting of district demographic and financial information using SPSS (version 22) prior to analysis. Subsequently, student-level data from ODE obtained through NCAASE were cleaned and merged into a longitudinal data file. The district-level variables were merged with the student-level data file prior to analysis.

Student-level variables. To address the first research question of whether or not student-level demographic factors were related to the growth of SWDs, student demographic information was included in the model building process. Students' demographic variables (sex and race/ethnicity) were investigated longitudinally, and in cases of conflicting demographic information, the value used was that reflecting the majority of the six years. Students with information that could not be resolved were removed from the analytic sample.

If a student received FRL or services for LEP at any point from Grades 3-8 they were classified as a member of that subgroup. SWD status, however, was coded as the percentage of years of test data a student was eligible to receive SpEd services. Resulting values for this *Time in SpEd* variable ranged from 0 (never eligible to be served) to 100 (eligible to be served in all years tested), with a mean of 14.30% for reading and 14.35% for mathematics. To put this in context, a student with six years of data who was eligible for SpEd services one out of six years would be have been eligible 16.67% of the time. "Average" time eligible for services was less than one year.

Mean reading and mathematics scores from Grades 3-8 are displayed in Table 3 for the analytic samples as a whole, for SWoDs, and for students who were eligible to receive SpEd services broken up into five categories: eligible 1% to 20% of the time, 21% to 40% of the time, 41% to 60% of the time, 61% to 81% of the time, and 81% to 100% of the time. Of the 9,692 SWDs included in the analytic sample for reading, and 9,594 included for mathematics, approximately 8% of SWDs were eligible 1% to 20% of the time, 12% were eligible 21% to 40% of the time, 13% were eligible 41% to 60% of the time, 14% were eligible 61% to 90% of the time, and 53% were eligible 81% to 100% of the time (Table 3) across subjects. Overall, SWDs who were eligible to receive SpEd services for any amount of time in Grades 3-8 had lower mean scores than SWoDs. In addition, students who were eligible for services in more years had lower means in reading and mathematics than students who received fewer years of services, with the exception of the groups of SWDs eligible for services 1% to 20% of the time, and 20% to 40% of the time. These two groups had very similar trajectories for both subjects. These relationships are displayed in Figures 1 (reading) and 2 (mathematics).

District-level variables. To address the research question on the influence of district-level demographic factors on growth for SWDs, a series of variables was investigated during the model-building process. Total district size, proportion of non-white students, as well as the percentage of the district receiving FRL and services for SpEd or LEP were investigated as potential predictors of student growth. Values for districts were calculated as averages across the six-year span for each variable, with descriptive statistics shown in Table 4.

Table 3

Mean RIT Score by Grade and Student Group, for Reading and Mathematics

Group		Grade 3		Grade 4		Grade 5		Grade 6		Grade 7		Grade 8	
	N	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Reading total	49,578	214.98	12.49	220.44	11.17	223.50	9.64	227.97	9.52	234.10	9.83	234.61	8.62
SWoD	39,886	216.86	11.57	222.13	10.45	225.10	8.79	229.56	8.68	235.61	9.10	235.90	7.90
SpEd 1-20%	793	212.64	11.94	218.51	10.82	221.83	8.81	226.32	8.82	232.34	9.62	233.05	8.68
SpEd 21-40%	1,140	213.08	12.71	218.78	11.22	221.80	10.04	226.21	9.93	232.55	10.31	233.40	9.16
SpEd 41-60%	1,239	210.48	12.61	216.86	10.91	220.20	9.60	224.41	9.94	230.89	10.26	231.89	9.00
SpEd 61-80%	1,345	207.10	12.54	213.46	11.07	217.29	9.85	222.02	9.68	228.32	9.61	229.72	8.97
SpEd 81-100%	5,175	203.87	12.71	210.38	10.66	214.07	9.71	218.67	9.42	224.90	9.51	226.52	8.89
Mathematics total	48,779	209.55	11.45	217.75	10.10	224.03	10.04	226.77	10.09	234.33	9.13	237.07	10.69
SWoD	39,185	210.99	10.95	219.15	9.44	225.51	9.48	228.30	9.47	235.60	8.65	238.53	10.01
SpEd 1-20%	778	207.77	10.72	216.08	9.96	222.21	9.78	224.87	9.42	232.52	8.89	234.68	10.65
SpEd 21-40%	1,141	208.25	12.02	216.21	10.66	222.27	10.36	226.96	10.41	232.79	9.49	235.47	11.23
SpEd 41-60%	1,252	206.00	11.75	214.53	10.47	220.70	9.80	223.28	10.48	231.53	9.22	233.52	11.25
SpEd 61-80%	1,327	203.35	11.03	212.43	10.09	218.58	9.84	221.13	9.76	229.53	9.01	231.39	11.16
SpEd 81-100%	5,096	201.35	11.03	209.73	10.46	215.65	9.40	218.01	9.50	226.82	8.50	228.00	10.50

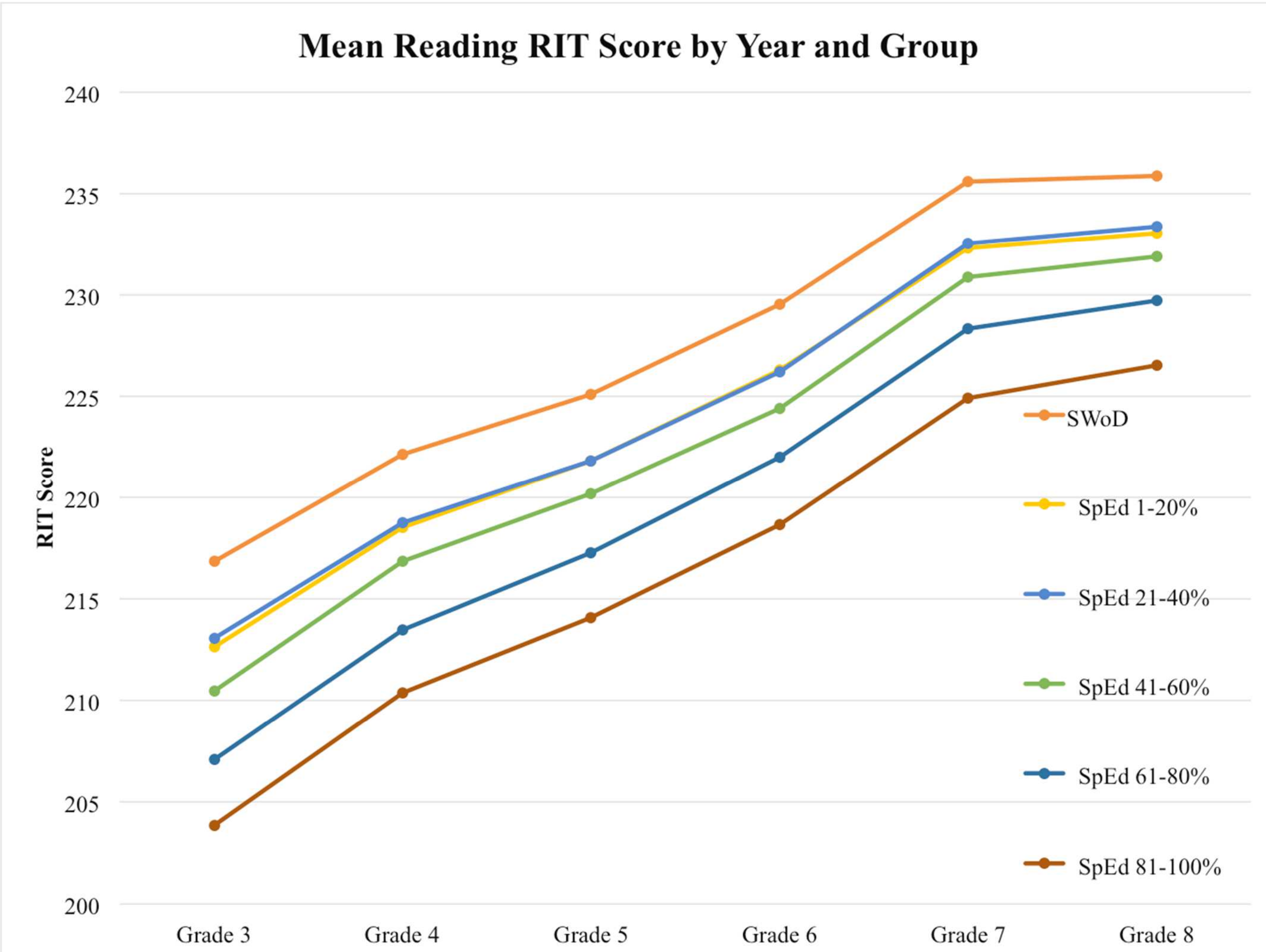

Figure 1. Mean Reading RIT Score by Grade and Group

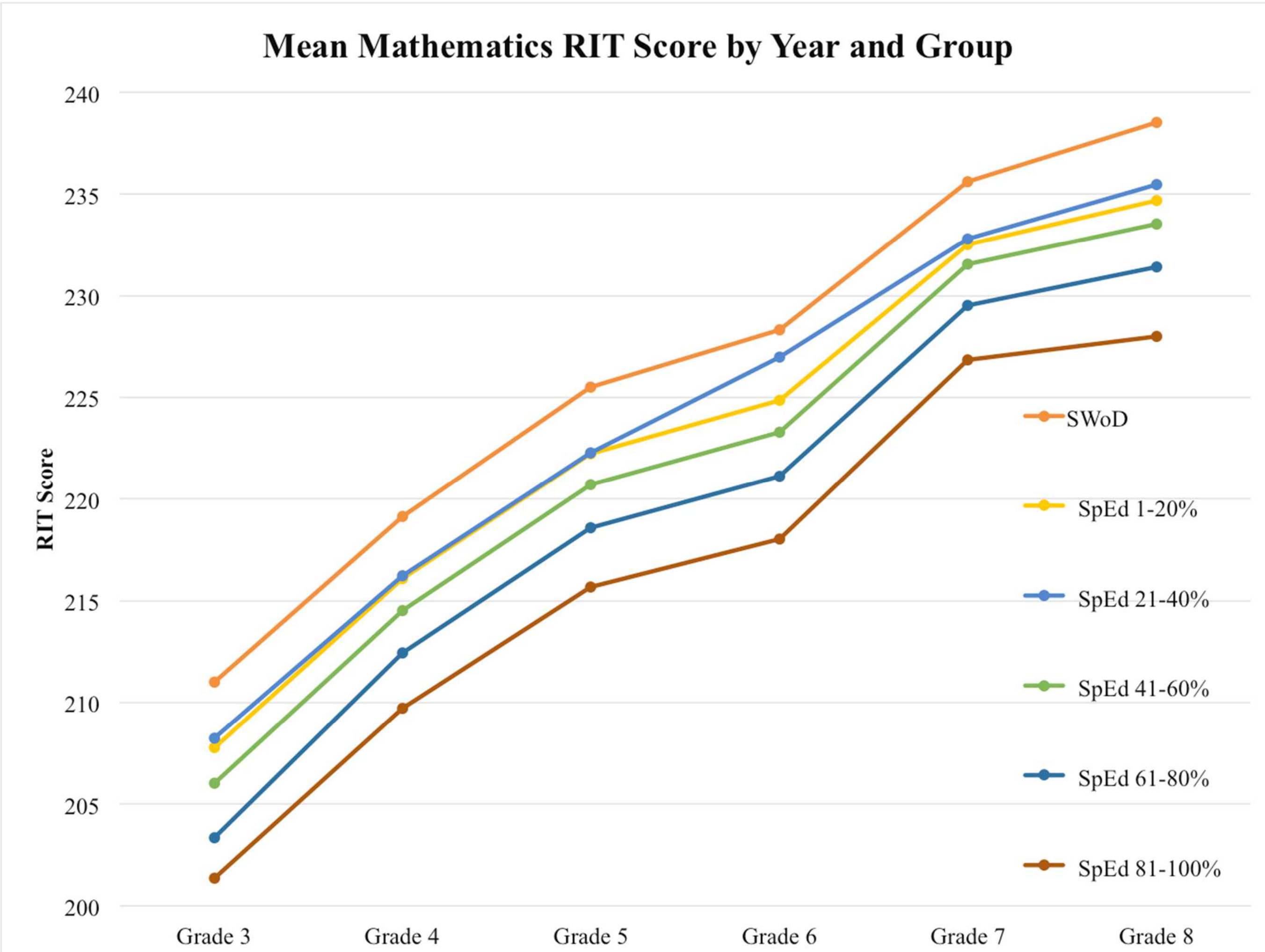

Figure 2. Mean Mathematics RIT Score by Grade and Group

To explore whether or not there was a relationship between district location and the performance of SWDs, the variable “remote” was computed using the Common Core of Data variable “ULOCAL.” This 12-level variable was recoded to have only two categories: remote, which included towns and rural areas labeled as “distant” or “remote” (values of 32, 33, 42, and 43 in the ULOCAL variable); and non-remote, which included all other region types. In total, approximately 81% of districts were categorized as remote (*n* = 120 for reading, *n* =121 for mathematics).

To investigate the third research question on the relation of the growth of SWDs to district financial variables, four variables were constructed using extant financial data. Total funding per student, total SpEd funding per SWD, percent increase in total funding per student, and the percent increase in SpEd funding per student were calculated for each district, as described below.

One consideration when looking at total funding over this time period is that the recession of 2008 played a part in total district funding. Figure 3 tracks mean district funding state-wide over this 6-year period, which illustrates a funding increase after school year 2006-2007, and an obvious drop following school year 2008-2009 coinciding with the 2008 recession.

Given this demonstrated variability over time in both per pupil and total district funding, two variables were created: the first represented average district expenditures per student, and the second represented average district SpEd expenditures per SWD. For average district expenditures per student, first, the average total funding obtained from ODE reports was computed per student for each district across the six-years of data, and then averaged across the six years of data.

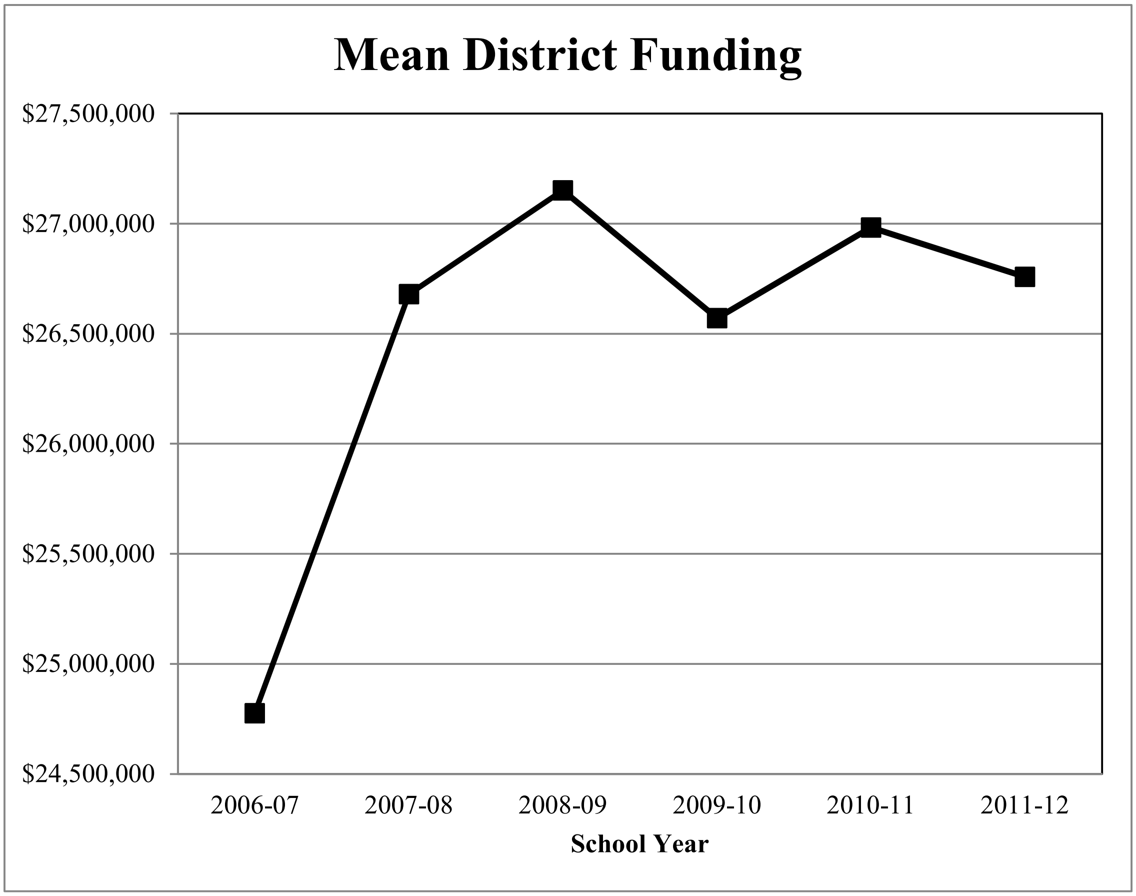

Figure 3. Oregon mean district funding, school years 2006-2007 through 2011-2012

The final variable, total funding, was scaled with $500 increments so that a one-unit increase corresponded with a $500 increase in total funding per student. For average district SpEd spending per SWD, three sources of data were obtained for each district: federal IDEA funding, district-level SpEd funding, and the ESD contribution to SpEd funding. These three funding sources were combined for each year. Average SpEd funding per SWD was computed for each district annually, and then averaged across the six-years of data. The final variable, average SpEd funding, was scaled in $500 increments per SWD.

Table 4

Continuous District Variables: Means and Standard Deviations

Variable	*M*	*SD*	Minimum	Maximum
Reading (*n* = 140 districts)				
Size	3,745.41	6,628.62	105.67	46,021.67
% SWD	14.59	2.69	3.85	22.32
% Non-White	22.95	15.08	5.22	77.23
% LEP	7.66	9.92	0	54.69
% FRL	50.59	14.09	0.75	82.33
Total funding per student	$10,165	$2,055	$7,625	$16,455
Total SpEd funding per SWD	$9,240	$1,682	$6,140	$18,785
% change total funding per student	10.82	10.65	-26.82	50.71
% change SpEd funding per SWD	18.39	20.96	-23.54	111.67
Mathematics (*n* = 141 districts)				
Size	3,721.29	6,612.75	105.67	46,021.67
% SWD	14.60	2.69	3.85	22.32
% Non-White	22.87	15.06	5.22	77.23
% LEP	7.61	9.91	0	54.69
% FRL	50.67	14.07	0.75	82.33
Total funding per student	$10,184	$2,064	$7,625	$16,455
Total SpEd funding per SWD	$9,236	$1,678	$6,140	$18,785
% change total funding per student	10.89	10.65	-26.82	50.71
% change SpEd funding per SWD	19.02	22.29	-23.45	113.28

Note. All demographic and financial values above are expressed in actual units

Both variables that represented percent increase in funding were calculated at the per-student level. The first represented percent change in total funding per student (year 1 funding per student subtracted from year 6 funding per student) at the district level. The second represented percent change in SpEd funding per SWD, and was calculated using the same procedure. These four financial variables accounted for both the average funding available per student as well as increase/decrease in funding over the 6-year period, and are displayed in Table 4.

Analyses

Two, three-level hierarchical linear models were constructed to address all three research questions at different points during the model building process, one for reading, and one for mathematics. The model building process was identical for both content areas, with slight variations in final models, discussed below. These models included a baseline model of time (level-1) nested within students (level-2) and districts (level-3). All models were estimated using HLM 7 for Windows (Bryk, Raudenbush, & Congdon, 2004), using full maximum likelihood estimation, and all model parameters specified as random effects.

HLM unconditional models. Both the reading and the mathematics model-building processes began by fitting an unconditional model, followed by an unconditional growth model. Previous research has suggested that growth across years may be nonlinear (e.g., Bloom, Hill, Black, & Lipsey, 2008). The functional form of the mean growth trajectory was explored by including a quadratic term in the models. Although the quadratic terms were significant for both models ($p < .001$) and accounted for a significant reduction in model deviance (Reading: x^2 (7) = 10,638.12, $p < .001$; Mathematics: x^2 (7) = 12654.11, $p < .001$), addition of a quadratic term accounted for less than a 1% reduction in deviance for both subjects. Visual inspection of plots of individual student scores for a random sample of 100 students in each subject did not support the inclusion of a quadratic term. Given the large sample sizes used in this study, which increased the likelihood that even very small differences were deemed significant, the quadratic term was eliminated from future modeling for purposes of parsimony. Full results for the reading and mathematics quadratic models are shown in Appendix A.

Following the decision to model linear student growth, the percentage of time a SWD was eligible for SpEd services was added to the model. This variable ranged from 0-100%, with a mean of 14.30% (*SD* =31.70) in reading, and 14.35% (*SD* =31.67) in mathematics. This SpEd model was defined at level 1 following the notation of Raudenbush and Bryk (2002) as

$$Y_{tij} = \pi_{0ij} + \pi_{1ij}(Year)_{tij} + e_{tij} \quad (1a)$$

where Y_{tij} represents the outcome measure of OAKS reading or mathematics at time *t* for student *i* nested in district *j*, π_{0ij} represents the initial status of student *i* nested in district *j* at Grade 3, and π_{1ij} is the linear growth rate for student *ij*. The e_{tij} term represents unmodeled residual variance. It is assumed that this term is normally distributed, with variance σ^2.

At level 2, the between-student variability was modeled for the intercept and the linear growth parameter as follows:

$$\begin{aligned} \pi_{0ij} &= \beta_{00j} + \beta_{01j} * (\boldsymbol{time\ in\ SpEd}) + r_{0ij} \\ \pi_{1ij} &= \beta_{10j} + \beta_{11j} * (\boldsymbol{time\ in\ SpEd}) + r_{1ij} \end{aligned} \quad (1b)$$

Time in SpEd was entered grand-mean centered, such that the β_{00j} term represents the average initial student achievement for students who were eligible to receive SpEd services for an average percentage of time at Grade 3 (14.30% for reading and 14.35% for mathematics), and β_{10j} represents the average student linear growth rate for students who were eligible to receive SpEd services for an average percentage of time. The terms β_{01j} and β_{11j} represents the difference between the average intercept and the average linear growth rate for a student given a 10% change in time eligible to receive SpEd services. The terms r_{0ij} and r_{1ij} represent individual student variation around the average intercept and linear growth. By including these residuals, each student was allowed to have his/her own growth trajectory and a random effects model was specified. These random effects were assumed to be distributed according to the multivariate normal distribution, with an unstructured variance-covariance matrix.

At level 3, the between-district variability was modeled, as follows

$$\begin{aligned} \beta_{00j} &= \gamma_{000} + u_{00j} \\ \beta_{01j} &= \gamma_{010} + u_{01j} \\ \beta_{10j} &= \gamma_{100} + u_{10j} \\ \beta_{11j} &= \gamma_{110} + u_{11j} \end{aligned} \tag{1c}$$

where γ_{000}, and γ_{100} represent average district achievement and average district linear growth for students who were eligible to receive SpEd services for an average amount of time at Grade 3, and γ_{010}, and γ_{110} represent average district achievement and average district linear growth for students who were eligible to receive SpEd services 10% more than average during Grades 3-8. The u_{00j}, u_{01j} u_{10j}, and u_{11j} terms represent variations in parameters between districts, and were assumed to be distributed normally.

Model-building process. The model-building process consisted of adding variables to the SpEd model in three phases: student-level characteristics, district demographics, and district financial variables. To address the first research question of whether or not student-level demographic factors were related to the growth of SWDs, all additional student demographics (sex, FRL, LEP, non-white) were added simultaneously to the intercept and slope of the SpEd model. These demographics were entered as un-centered dummy-coded variables, with "1" representing the corresponding characteristic. Chi-square deviance tests were used to determine whether the model including student demographics fit the data better than the SpEd model. Pseudo-R^2s also were computed to assess the percent of variance explained by the student demographic model as compared to the SpEd model.

To address the second research question, regarding the extent to which district demographic variables were related to reading and mathematics growth for SWDs from Grades 3-8, district-level demographic variables (remote, district size, and demographic percentages) were added to the model at level-3 as predictors of both the overall slope and intercept as well as on the effect of time eligible for SpEd services on students' intercept and slope. In this district demographic model, the dichotomous "remote" variable was uncentered, and all other predictors were entered grand-mean centered, so that the coefficients for these predictors represented the effect of a 10% change relative to a non-remote district of average size and demographic characteristics. As with the student demographic characteristics, all predictors were added simultaneously, and model fit evaluated. Variables that were not significant were removed using backwards elimination. The total variance explained at level 3 (pseudo-R^2) was compared before and after elimination of predictors to ensure that the percentage of variance explained did not change substantially as a result of using backwards elimination. Chi-square deviance tests were used to determine whether the district demographic model fit the data better than the student demographic model. Pseudo-R^2s were computed to assess the percentages of variance explained by each component of the district demographic model as compared to the student demographic model. In this manner, the extent to which district demographic variables account for variance in student growth above and beyond student-level demographics was evaluated.

To address the third research question of whether financial investment in students and SWDs was related to growth for SWDs above and beyond district demographics, district funding variables (average total funding per student, average SpEd funding per SWD, percent change in total funding per student, and percent change in SpEd funding per SWD) were added as predictors of the intercept and linear slope, and the SpEd intercept and linear slope. Other scales of the funding variables were investigated (increases of $100 and $1,000 in total and SpEd funding) to explore how to appropriately scale those variables, with the end decision being to retain $500 increments of funding for modeling. All financial factors were added simultaneously to this district funding model and non-significant predictors were removed using backwards elimination to improve model fit. Variance explained in the full and final model was compared to ensure similarity, and final model fit was evaluated using deviance testing compared to the district demographic model.

Effect sizes. Achievement gap effect sizes were calculated (Cohen's *d*) between SWD and SWoD were calculated (Bloom et al., 2008) between each SWD group and SWoD students. The mean difference between the five SWD groups described above (e.g., 1-20% time eligible for SpEd services, 21-40% time eligible for services) and SWoDs was calculated for each year, and divided by the standard deviation of all students from that year. The resulting effect sizes were interpreted using the same rules of thumb for Cohen's *h*, discussed above.

CHAPTER III

RESULTS

Full model results are displayed for reading, with the fixed effects in Table 5, and the variance components and model fit statistics in Table 6. For mathematics, results are displayed in Tables 7 and 8. The following sections describe the model building process and the interpretation of model results for the variables of interest with respect to research questions.

Reading

The first model for reading was a baseline model, estimating only the grand mean and variance components. Subsequently, an unconditional linear growth model was estimated by including "Year" as a Level 1 predictor, as shown in the first set of columns in Table 5 (fixed effects). This model resulted in a statistically significant improvement in model fit ($p < .001$) as shown in Table 6 (random effects). In this model, both intercept and linear slope were allowed to vary randomly across students and districts. On average, students had a score of 214.44 RIT points in Grade 3 and gained 4.15 points per year, which were both significantly different from zero ($t = 862.74$, $SE = 0.25$, $p < .001$; $t = 99.80$, $SE = 0.04$, $p < .001$, respectively). A deviance test showed the linear growth model fit the data better than the baseline model, $\chi^2\ (5) = 212862.97$, $p < .001$.

SpEd model. To establish the SpEd model, the *Time in SpEd* variable was added as a predictor of both students' intercepts and slopes. The effect of *Time in SpEd* was specified as varying randomly across districts. Fixed effects from this model are shown in the second set of columns of Table 5. The reference group for this model consisted of students who had an average percentage of time eligible to receive SpEd services

(14.30%). The results of this model indicated that students who had an average percentage of time eligible for SpEd had an average score of 214.48 points in Grade 3 and progressed at 4.15 points per year across Grades 3-8. A 10% increase in the percentage of time eligible for SpEd services corresponded to a predicted decrease in initial achievement of 1.37 points, with an average rate of growth 0.08 RIT points higher than the reference group, per year.

Variance components for the SpEd model are presented in Table 6. The addition of the SpEd term accounted for approximately 17% of the unexplained variance in student intercepts, and 5% of the variance in students' slopes, as compared to the unconditional growth model. A deviance test indicated that the SpEd model better fit the data than the unconditional growth model, χ^2 (9) = 7,171.54, $p < .001$.

Student demographics. Student-level demographic predictors (sex, FRL, LEP, non-white) were added to the model as predictors of both students' intercepts and slopes. All demographic predictors were allowed to vary between districts, with fixed effects shown in the third column of Table 5. The reference group for this model consisted of students who were eligible to receive SpEd services 14.30% of the time who were female, White, not LEP, and not FRL-eligible. The results of this model indicated that students in the reference group had a score of 218.23 points in Grade 3 and progressed at 4.04 points per year across Grades 3-8.

Holding all other predictors constant, a 10% increase in the time students were eligible to receive SpEd services above corresponded with a 1.33 point decrease in students' initial reading performance in Grade 3, and slightly higher linear growth (0.08) than the reference group. In addition, male (-0.29), non-White (-1.69), LEP (-7.71), and

FRL students (-3.30) all had significantly lower initial reading performance than the reference group. Linear growth for male students (-0.13) was significantly lower than the reference group. Non-White students (0.06) and students eligible to receive LEP services (0.86) had linear growth rates that were significantly higher than the reference group.

Variance components for the student demographic model (Table 6) indicated that the student demographic predictors accounted for an additional 21% of the variance in students' intercepts, and 10% of the variance in students' slopes, as compared to the model that only included SpEd. A deviance test indicated that the student demographic model better fit the data than the SpEd model, χ^2 (76) = 10,781.06, $p < .001$.

District demographics. Following estimation of the student demographic model, district-level demographic characteristics were added as predictors of the overall intercept and slope. District demographics were also entered as predictors of the effect of *Time in SpEd* on students' intercept and slope (i.e., cross-level interactions). The model first was run including all district demographics (remote, district size, and the percent of SWD, LEP, FRL, and non-White students in the district). For the sake of parsimony, demographic predictors that were not statistically significant were removed individually using backwards elimination, and the model was re-run. Eliminating these predictors increased the percentage of variance explained at level-3 by only 1% of its original value as compared to the pseudo-R^2 for the full model with all district-level demographic variables that is presented in Appendix A. District demographic variables included in the final model were remote, the percentage of LEP, and the percentage of FRL students in the district. Fixed effects from the final district demographic model are shown in the fourth column of Table 5. The reference group for this model consisted of students eligible to receive SpEd services 14.30% of the time who were female, not LEP, and not FRL-eligible, from non-remote districts with average percentages of LEP and FRL students. The results of this model indicated that students in the reference group had an average Grade 3 score of 218.95 points and progressed at 3.93 points per year across Grades 3 to 8. A 10% increase in the percentage of time eligible for SpEd corresponded to a predicted decrease in initial achievement of 1.32 points, with an average rate of growth 0.07 RIT points higher than the reference group, per year.

At the district level, holding all other predictors constant, students in remote districts had lower Grade 3 reading scores (-1.09) and higher linear growth (0.18), on average, than students in non-remote districts, both of which were statistically significant ($p < .01$). For every 10% increase in students eligible for FRL in a district above the overall district average, students had significantly lower initial reading scores (-0.83) than students in districts with an average percentage of FRL students. In contrast, for every 10% increase in students with LEP in a district above the overall district average, students had significantly higher initial reading scores (0.74) than students in districts with an average percentage of LEP students. In addition, the effect of *Time in SpEd* on student growth was moderated by remote district location (<0.01) although this value was so small that it is likely not practically significant.

Variance components for the district demographic model (Table 6) indicated that the district demographic terms accounted for approximately 24% of the variance in intercepts between districts, and 4% of the variance in slopes between districts. A deviance test indicated that the district demographic model fit the data better than the student demographic model, χ^2 (6) = 57.56, $p < .001$.

District funding. In the last model, district funding predictors were added as predictors of intercept and slope. District funding variables were also entered as predictors of the effect of *Time in SpEd* on students' intercept and slope. The model first was run including all district funding variables (average funding per student, average SpEd funding per SWD, percent increase in total funding per student and percent increase in SpEd funding per SWD); these results are presented in Appendix A. To obtain a more parsimonious model, funding predictors that were not statistically significant were removed using backwards elimination. The variance explained by the full and final models was compared, and the differences in variance explained at level three changed only slightly with the removal of non-significant predictors. Fixed effects from the final district funding model including only significant predictors are shown in the final column of Table 5. The reference group for this model consisted of students who were eligible to receive SpEd services 14.30% of the time who were female, White, not LEP, and not FRL-eligible, from a district with the average percent of LEP and FRL-eligible students, average total funding per student, average SpEd funding per SWD, average increases in total funding per student and average increases in SpEd funding per SWD. The results of this model indicated that students in the reference group had an average score of 219.05 points in Grade 3 and progressed at 3.93 points per year across Grades 3-8. For every 10% increase in the time students were eligible to receive SpEd services, their initial achievement decreased by 1.32 points and they progressed 0.07 points more per year than the reference group.

Table 5

Fixed Effects, Longitudinal HLM Models, Reading Grades 3 to 8.

Fixed effect	Unconditional growth		SpEd model		Student demographics		District demographics		District funding	
	Intercept	Growth	Intercept	Growth	Intercept	Growth	Intercept	Growth	Intercept	Growth
Mean	214.44**	4.15**	214.48**	4.15**	218.23**	4.04**	218.95**	3.93**	219.05**	3.93**
	(0.25)	(0.04)	(0.24)	(0.04)	(0.24)	(0.04)	(0.27)	(0.06)	(0.27)	(0.06)
Student Level										
Time in SpEd			-1.37**	0.08**	-1.33**	0.08**	-1.32**	0.07**	-1.32**	0.07**
			(0.02)	(0.01)	(0.02)	(<0.01)	(0.02)	(0.01)	(0.02)	(0.01)
Male					-0.29**	-0.13**	-0.29*	-0.13**	-0.28*	-0.13**
					(0.10)	(0.02)	(0.10)	(0.02)	(0.10)	(0.02)
Non-White					-1.69**	0.06*	-1.74**	0.07*	-1.73**	0.06*
					(0.19)	(0.03)	(0.19)	(0.03)	(0.19)	(0.03)
LEP					-7.72**	0.86**	-7.94**	0.86**	-7.92**	0.86**
					(0.22)	(0.04)	(0.24)	(0.04)	(0.24)	(0.04)
FRL					-3.30**	0.03	-3.19**	0.03	-3.19**	0.03
					(0.17)	(0.02)	(0.17)	(0.02)	(0.18)	(0.02)
District level										
Remote							-1.09*	0.18*	-1.16*	0.17*
							(0.33)	(0.07)	(0.32)	(0.07)
% LEP							0.74**	--	0.78**	--
							(0.16)		(0.16)	
% FRL							-0.84**	0.06*	-0.90**	0.06*
							(0.13)	(0.03)	(0.13)	(0.03)
Remote on SWDs							--	<0.01*	--	<0.01*
								(0.01)		(0.01)
Average funding per *n*									0.07*	--
									(0.04)	
Average SpEd funding per SWD on SWDs									--	<-0.01*
										(<0.01)
% Δ SpEd funding per SWD on SWDs									--	<0.01*
										(<0.01)

Note. Standard errors shown in parentheses. Coefficients for SpEd, %LEP and %FRL are reported in increments of 10%. * $p < .05$, ** $p < .001$

Table 6

Random Effects and Model Fit, Longitudinal HLM Models, Reading Grades 3 to 8.

Random effect	Unconditional growth		SpEd model		Student demographics		District demographics		District funding	
	Intercept	Growth	Intercept	Growth	Intercept	Growth	Intercept	Growth	Intercept	Growth
Level-2 Students	112.23**	1.27**	93.57**	1.21**	73.90**	1.08**	73.92**	1.08**	73.91**	1.08**
Level-1 Residual	23.54		23.55		23.55		23.55		23.55	
Level-3 Districts	7.62**	0.21**	7.33**	0.21**	6.16**	0.16**	3.15**	0.15**	2.98**	0.15**
in time in SpEd			<0.01**	<0.01**	<0.01**	< 0.01**	<0.01**	<0.01**	<0.01**	<0.01**
in Male					0.17	0.01	0.19	0.01	0.19	0.01
in Non-White					1.77**	0.04**	1.85**	0.04**	1.85**	0.04**
in LEP					1.86**	0.04*	1.92**	0.04*	2.02**	0.04**
in FRL					1.69**	0.01	1.72**	0.01	1.73**	0.01
Student-level pseudo-R^2			16.63	4.72	21.02	10.74				
District-level pseudo-R^2							24.21	4.00	0.71	-0.84
Δ Deviance, χ^2 (*df*)	212862.97 (5)**		8,514.75 (9)**		10,799.31 (76)**		57.56 (6)**		11.90 (3)*	

Note. Pseudo-R^2 values were compared to previous model, and expressed as percentages. * $p < .05$, ** $p < .001$

Holding all other predictors constant, some district funding variables were related to students' initial scores and rates of growth. For every $500 increase in average total funding per student above the overall district mean, students' Grade 3 reading RIT scores were 0.07 units higher than students in districts with average funding per student. The effect of time in SpEd on linear growth was moderated by districts having higher SpEd funding per SWD (<-0.01 per $500 higher), and by increases in SpEd funding over time (<0.01 per 10% increase). These relations to district funding, although significant, are likely not practically significant due to their magnitude.

Variance components for the district funding model (last column of Table 6) demonstrated that the additional district demographic terms accounted for approximately 1% of the variance between district intercepts, yet corresponded with an increase of 1% in the variance between average slopes between districts. A deviance test indicated that the district funding model fit the data better than the district demographic model, χ^2 (3) = 11.90, $p < .05$.

Mathematics

Adding a linear growth trajectory to the baseline model to estimate an unconditional growth model resulted in a statistically significant improvement in model fit ($p < .001$) as shown in the first set of columns in Tables 7 (fixed effects) and 8 (random effects). In this model, both intercept and linear slope were allowed to vary randomly across students. On average, students had a score of 210.23 RIT points in Grade 3, and gained 5.27 points per year, which were both significantly different from zero ($t = 876.37$, $SE = 0.24$, $p < .001$; $t = 108.43$, $SE = 0.05$, $p < .001$, respectively). A deviance test indicated that the linear growth model fit the data better than the baseline model, χ^2 (5) = 276,309.93, $p < .001$.

SpEd model. The *Time in SpEd* variable was added as a predictor of both students' intercepts and slopes, with its' effect specified as varying randomly across districts. Fixed effects from this model are shown in the second set of columns of Table 7. The reference group for this model represented students who had an average percentage of time eligible to receive SpEd services (14.35%). The results of this model indicated that students who had an average percentage of time eligible for SpEd had an average score of 210.27 points in Grade 3 and progressed at 5.26 points per year across Grades 3-8. A 10% increase in the percentage of time eligible for SpEd corresponded to a predicted decrease in initial achievement of 1.01 points, with an average rate of growth that was not significantly different than the reference group.

Variance components for the SpEd model are presented in Table 8. The addition of the SpEd term accounted for approximately 12% of the variance in students' intercepts, and 1% of the variance in students' slopes, as compared to the unconditional growth model. A deviance test indicated that the SpEd model fit the data better than the unconditional growth model, χ^2 (9) = 6,633.45, $p < .001$.

Student demographics. Student-level demographic predictors (sex, FRL, LEP, non-white) were added to the model as predictors of both students' intercepts and slopes. All demographic predictors were allowed to vary between districts, with fixed effects shown in the third column of Table 7. The reference group for this model consisted of students who were eligible to receive SpEd services 14.35% of the time who were female, White, not LEP, and not FRL-eligible. These students had a score of 12.01 RIT points in Grade 3, and progressed at a rate of 5.44 RIT points per year across Grades 3-8.

Holding all other predictors constant, a 10% increase in the time students were eligible to receive SpEd services corresponded to a 1.01 point decrease in students' initial mathematics performance in Grade 3, and slightly higher linear growth (0.01) than the reference group. Students who were male had significantly higher initial mathematics performance (1.89) than the reference group. Students who were non-White (-1.44), LEP (-5.06), and FRL-eligible students (-2.65) had significantly lower initial mathematics performance than the reference group. Linear growth for male students (-0.23) and FRL students (-0.22) was significantly lower than the reference group. Non-White students (0.05) and students eligible to receive LEP services (0.61) had linear growth rates that were significantly higher than the reference group.

Variance components for the student demographic model (Table 8) indicated that the student demographic predictors accounted for an additional 16% of the unexplained variance in student intercepts, and 6% of the unexplained variance in student slopes, as compared to the model that only included SpEd. A deviance test indicated that the student demographic model was a better fit to the data than the SpEd model, χ^2 (76) = 8,168.89, $p < .001$.

District demographics. District-level demographic characteristics were added as predictors of the overall intercept and slope. District demographics were also entered as predictors of the effect of *Time in SpEd* on students' intercepts and slopes, as well as to the difference for students who were eligible for above average percentage of SpEd services from the intercept and slope. The model first was run including all district demographics (remote, district size, and the percent of SWD, LEP, FRL, and non-White students in the district). For the sake of parsimony, demographic predictors that were not statistically significant were removed individually using backwards elimination, and the model was re-run. Removing these predictors did not change the percentage of variance explained at any level of the model by more than 5% of the original value. Results of the full model with all district-level demographic variables are presented in Appendix A. District demographic variables that remained in the final model were remote, the percentage of LEP, and the percentage of FRL-eligible students in the district. Fixed effects from the final district demographic model are shown in the fourth column of Table 7. The reference group for this model consisted students eligible to receive SpEd services 14.35% of the time who were female, not LEP, and not FRL-eligible, from a non-remote district with average percentages of LEP and FRL-eligible students. These students had an average Grade 3 score of 212.61 points and progressed at a rate of 5.45 RIT points per year across Grades 3 to 8.

At the district level, controlling for all other predictors, students in remote districts had significantly lower Grade 3 mathematics scores (-0.84), on average, than students non-remote districts. For every 10% increase in students eligible to receive FRL above the overall district average, students had on average significantly lower (-0.73) than students in districts with an average percentage of FRL students. In contrast, for every 10% increase in students eligible for LEP services above the overall district average, students had on average significantly higher initial mathematics scores (0.62) and significantly higher linear growth (0.14) than students in districts with average percentages of LEP students.

Variance components for the district demographic model (Table 8) indicated that the district demographic terms accounted for approximately 16% of the variance in intercepts between districts, and 13% of the variance in slopes between districts. A deviance test indicated the district demographic model fit the data better than the student demographic model, χ^2 (5) = 87.83, p < .001.

District funding. In the last model, district funding predictors were added as predictors of intercept and slope, and as predictors of the effect of *Time in SpEd* on students' intercept and slope. The model first was run including all district funding variables (average funding per student, average SpEd funding per SWD, percent increase in total funding per student and percent increase in SpEd funding per SWD); these results are presented in Appendix A. To obtain a more parsimonious model, funding predictors that were not statistically significant were removed using backwards elimination. Fixed effects from the final district funding model including only significant predictors are shown in the final column of Table 7. The reference group for this model consisted of students who were eligible to receive SpEd services 14.35% of the time who were female, White, not LEP, and not FRL-eligible, from a district with the average percent of LEP and FRL students, average total funding per student, average SpEd funding per SWD and average increases in total funding per student. These students had an average score of 212.62 points in Grade 3, and progressed at a rate of 5.45 RIT points per year across Grades 3-8.

Table 7

Fixed Effects, Longitudinal HLM Models, Mathematics Grades 3 to 8.

Fixed effect	Unconditional growth		SpEd model		Student demographics		District demographics		District funding	
	Intercept	Linear	Intercept	Linear	Intercept	Linear	Intercept	Linear	Intercept	Linear
Mean	210.23** (0.24)	5.27** (0.05)	210.27** (0.24)	5.26** (0.05)	212.01** (0.25)	5.44** (0.05)	212.61** (0.27)	5.45** (0.05)	212.62** (0.28)	5.45** (0.05)
Student Level										
Time in SpEd			-1.01** (0.02)	<0.01** (<0.01)	-1.01** (0.02)	0.01* (<0.01)	-1.08** (0.02)	0.01* (<0.01)	-1.07** (0.02)	0.01* (<0.01)
Male					1.89** (0.08)	-0.23** (0.02)	1.96** (0.09)	-0.24** (0.02)	1.96** (0.09)	-0.24** (0.02)
Non-White					-1.44** (0.19)	0.05 (0.03)	-1.50** (0.19)	0.05* (0.03)	-1.49** (0.19)	0.05* (0.03)
LEP					-5.06** (0.21)	0.61** (0.04)	-5.00** (0.21)	0.56** (0.04)	-5.01** (0.21)	0.56** (0.04)
FRL					-2.65** (0.17)	-0.22** (0.03)	-2.67** (0.17)	-0.22** (0.03)	-2.66** (0.17)	-0.23 (0.03)
District level										
Remote							-0.84* (0.27)	--	-0.85* (0.27)	--
% LEP							0.62* (0.21)	0.14** (0.04)	0.62* (0.21)	0.14** (0.04)
% FRL							-0.73** (0.12)	--	-0.74** (0.12)	--
Remote on SWDs							0.01** (<0.01)	--	0.01* (<0.01)	--
Average SpEd funding per SWD on SWDs									<0.01** (<0.01)	< -0.01* (<0.01)
% Δ funding per *n* on SWDs									<0.01** (<0.01)	--

Note. Standard errors shown in parentheses. Coefficients for SpEd, %LEP and %FRL are reported in increments of 10%. * $p < .05$, ** $p < .001$

Table 8

Random Effects and Model Fit, Longitudinal HLM Models, Mathematics Grades 3 to 8.

Random effect	Unconditional growth		SpEd model		Student demographics		District demographics		District funding	
	Intercept	Linear	Intercept	Linear	Intercept	Linear	Intercept	Linear	Intercept	Linear
Level-2 Students	84.63**	1.15**	74.09**	1.14**	62.43**	1.07**	62.41**	1.07**	62.41 **	1.07**
Level-1 Residual	25.52		25.52		25.53		25.53		25.53	
Level-3 Districts	7.29**	0.31**	7.17**	0.31**	6.87**	0.34**	4.67**	0.29**	4.65**	0.29**
in time in SpEd			<0.01**	<0.01**	<0.01**	< 0.01**	<0.01**	<0.01**	<0.01**	<0.01**
in Male					0.13	0.01	0.15	0.01	0.15	0.01
in Non-White					1.73**	0.02	1.80**	0.01	1.78**	0.01
in LEP					1.74*	0.24	1.86*	0.05	1.87*	0.05
in FRL					1.36**	0.04**	1.81**	0.04**	1.83**	0.04**
Student-level pseudo-R^2			12.44	0.87	15.75	6.14	0.03	0	0	0.38
District-level pseudo-R^2			1.65	0.00	-71.69	-49.39	16.41	13.04	0.13	0.74
Δ Deviance, χ^2 (*df*)	276,309.93 (5)**		6,633.45 (9)**		8,168.89 (76)**		87.83 (5)**		9.37 (3)*	

Note. Pseudo-R^2 values were compared to previous model, and expressed as percentages. * $p < .05$, ** $p < .001$

Controlling for all other predictors, some district funding variables were related to students' initial achievement and rate of growth. District SpEd funding per SWD moderated the effect of *Time in SpEd* on students' initial mathematics scores (0.01 per $500 higher) and on linear growth (<-0.01 per $500 higher). The effect of *Time in SpEd* on students' initial mathematics scores was also moderated by increases in total funding per student (<0.01 per 10% increase).

Variance components for the district funding model (last column of Table 8) demonstrated that the additional district demographic terms accounted for less than 1% of the variance between district intercepts, and approximately of 1% of the variance in average slopes between districts. A deviance test indicated that the district funding model was a better fit to the data than the district demographic model, χ^2 (3) = 9.37, $p <$ .05.

Achievement Gap Effect Sizes

In comparison to SWoDs, all SWD groups were lower by 0.21 to 1.14 standard deviations in reading and by 0.24 to 1.02 standard deviations in mathematics at Grade 3. As shown in Table 9 for reading, these gaps were considered small according to Cohen's rules of thumb for students eligible to receive services up to 60% of the time, medium for students eligible to receive services 61-80% of the time, and large for students eligible to receive services 81-100% of the time. In Mathematics, these gaps were considered small for students eligible to receive services from 1-40% of the time, medium for students eligible to receive services 41-80% of the time, and large for students eligible to receive services 81-100% of the time. Overall, differences remained fairly stable within group over Grades 3-8, with Grade 8 scores 0.29 to 1.09 standard deviations lower in reading and 0.29 to 0.98 standard deviations lower in mathematics.

Table 9

Achievement Gap Effect Sizes Compared to SWoDs by Subject, Grades 3 to 8.

Group	Grade 3	Grade 4	Grade 5	Grade 6	Grade 7	Grade 8
Reading						
SpEd 1-20%	0.34	0.32	0.34	0.34	0.34	0.33
SpEd 21-40%	0.26	0.21	0.25	0.28	0.24	0.23
SpEd 41-60%	0.30	0.30	0.34	0.35	0.31	0.29
SpEd 61-80%	0.51	0.47	0.51	0.54	0.48	0.47
SpEd 81-100%	1.04	1.05	1.14	1.14	1.09	1.09
Mathematics						
SpEd 1-20%	0.28	0.30	0.33	0.34	0.34	0.36
SpEd 21-40%	0.24	0.29	0.32	0.33	0.31	0.29
SpEd 41-60%	0.44	0.46	0.48	0.50	0.45	0.47
SpEd 61-80%	0.63	0.67	0.69	0.71	0.67	0.67
SpEd 81-100%	0.84	0.93	0.98	1.02	0.96	0.98

Gaps for students eligible to receive services 1-40% of the time were similar in magnitude across subjects, whereas students eligible to receive services at least 41% of the time had gaps that differed across subjects. Overall, students eligible to receive services 41-80% of the time had larger gaps in mathematics than in reading, whereas students eligible to receive services 81-100% of the time had larger gaps in reading than mathematics. These trends are displayed in Figure 4.

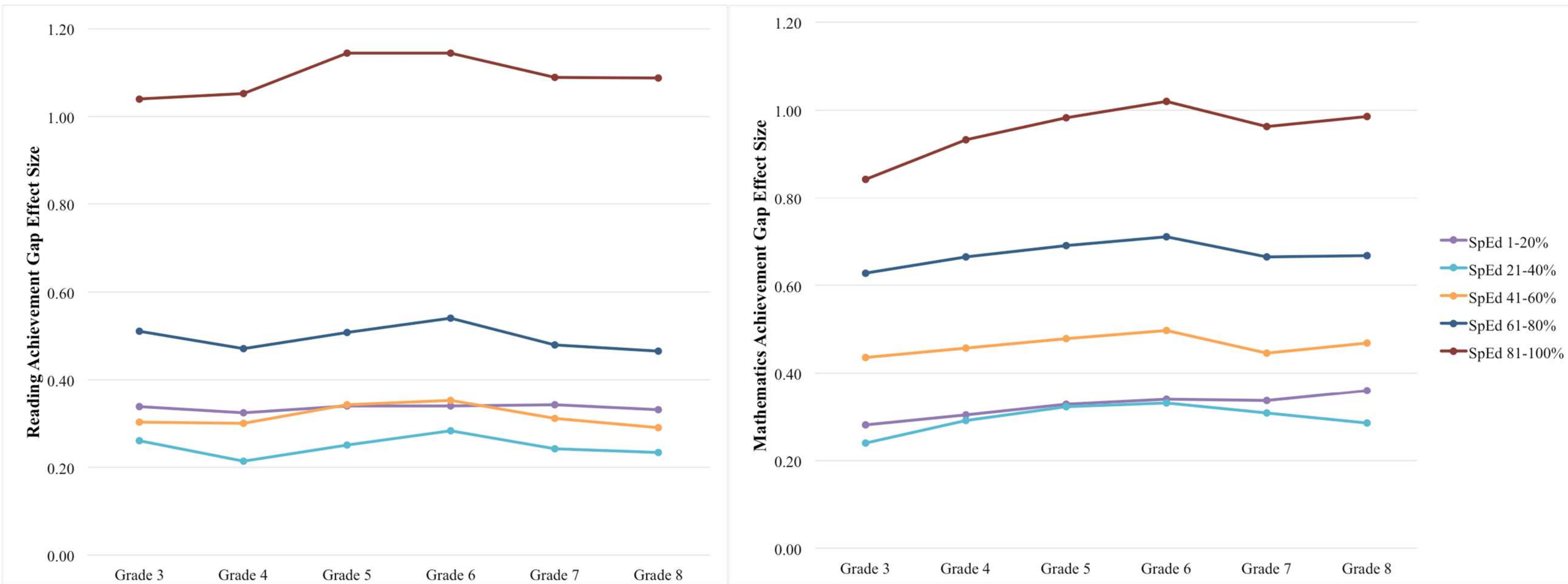

Figure 4. Comparison of reading and mathematics achievement gap effect sizes for SWD groups. In each case, the SWD groups were compared to SWoDs.

CHAPTER IV

DISCUSSION

The purpose of this dissertation study was to examine reading and mathematics growth for SWDs from Grades 3-8, accounting for student- and district-level characteristics. Previous research on growth for SWDs using accountability measures has been limited mainly to student characteristics (e.g., Schulte & Stevens, 2015; Wei et al., 2011), rather than examining the interplay between the two using a multi-level approach. Some studies have investigated school-level factors (e.g., Schulte & Villwock, 2004), yet, policies exist at the district level that influence these schools, in terms of curricula, professional development, and funding. Failing to explore the potential impact of this nested relationship on student growth leads to assumptions about the factors associated with the growth of SWDs. To better understand how student and district level factors relate to growth for SWDs, I discuss the substantive findings in relation to previous research, and then consider limitations. I finally present conclusions and recommendations for future research.

Substantive Findings

All else constant, across models and subjects, increased time eligible to receive SpEd services was related to higher linear growth compared to students with average time eligible to receive services (reference group for these models). For this study, "average" was approximately 14% of the time across subjects, and amounted to less than one year of eligibility for SpEd services. These differences were not large in magnitude, particularly for SWDs with 10% more *Time in SpEd* than the reference group (0.01). However, for students eligible to receive services 100% of the time the slope for reading was on average approximately 0.70 units per year higher in reading and 0.10 units per year higher in mathematics than the reference group. The practical significance of these differences is limited given their small magnitude, particularly for mathematics.

Achievement gap effect sizes (Bloom et al., 2008) computed to determine the effects of 20% increases in time a SWD was eligible to receive services on achievement as compared to SWoDs also showed gaps between SWD groups showed an effect of time eligible for SpEd services on student outcomes. This gap remained within two tenths of a standard deviation within group across time. This demonstrated that there was an effect of time eligible to receive SpEd services on reading and mathematics achievement for SWDs as compared to SWoDs. Overall, this effect was greater for SWDs who were eligible to receive services for longer periods of time.

Research question one: student demographics. Student demographic variables were related with overall student growth in both subjects. Increased time eligible for SpEd services was associated with slightly higher average initial reading achievement (-1.33 as opposed to -1.37), and slightly higher linear growth in mathematics (0.01 as opposed to <0.01) after the introduction of student demographics, as compared to the model that only included time eligible to receive SpEd services. This finding is consistent with previous research indicating that the introduction of student demographic variables can change estimates for SWDs on state accountability measures (e.g., Stevens et al., 2015).

The addition of student demographic characteristics did not change the magnitude of estimates for reading growth of SWDs based on the percentage of time eligible to receive SpEd services, and only had a slight influence on estimates of mathematics growth (0.01 instead of < 0.01). Estimates of growth for the reference group did change with the introduction of student demographics. For reading, the growth of the reference group was lower after the introduction of student demographics (difference of -0.11, *SE* = 0.04) and for mathematics the growth of the reference group was higher (difference of 0.18, *SE* = 0.05). In addition, more student-level variance was explained following the addition of these demographic models (approximately 11% in reading and 6% in mathematics).

In reading, each of the other student demographic characteristics had lower initial scores than the reference group ranging from -0.29 for male students to -7.72 for LEP students. In mathematics, all student demographic characteristics were associated with lower initial scores (ranging from -1.44 non-White to -5.06 for LEP) except male students, who scored higher initially (1.89). These student groups with lower initial scores were all identified as AYP subgroups based on an achievement gap between each respective group and students in the general population. These results support that these achievement gaps were still in existence almost a decade after NCLB was drafted. The differences in intercept for SWDs based on time eligible for SpEd services seems to fit in the middle of these student demographic groups: However, SWDs eligible for SpEd services 100% of the time had lower initial intercepts than any demographic group (-11.57 in reading, -8.66 in mathematics).

Although some demographic groups had lower growth than the reference group (male in reading, -0.13; FRL-eligible in mathematics, -0.22), non-White (0.06) and LEP students (0.86) had higher reading growth, and students with LEP had higher mathematics growth. Thinking in terms of closing achievement gaps, these gains compared to students in the general population were not of great enough magnitude to ever "close" the achievement gap for students in these demographic groups. SWDs eligible for SpEd services 100% of the time grew at a rate 0.70 points higher than the reference group in reading and 0.09 points higher in mathematics, which was higher than the growth of all groups except for students with LEP. Although this rate of growth was higher, it was not nearly enough to close achievement gaps between SWDs and SWoDs.

Growth for students overall changed with the addition of student demographics, such that the effect of eligibility to receive SpEd services 10% of the time added on to the overall estimate for student growth decreased in total by 0.11 points per year in reading and increased by 0.19 points per year in mathematics. Standard errors for the reference group and SpEd students in the SpEd Growth model were 0.04 and 0.01 in reading and 0.05 and less than 0.01 in mathematics, respectively. In the student demographics model these remained the same. The changes in student growth were larger than the standard errors in both subjects. Failure to consider student demographic characteristics when modeling the growth of SWDs may lead to biased estimates of annual growth, and should not be ignored when considering alternative methods for AYP reporting.

Research question two: district demographics. Three district-level demographic factors were related to reading and mathematics achievement: the remote nature of a district, percentage LEP, and percentage FRL, with remote location moderating the effect of time eligible for SpEd services on student growth. Overall, in remote districts, students performed lower initially in both subjects than students in non-remote districts. In contrast, the growth of students in remote districts was higher in reading, on average, than students in non-remote districts. This difference may be attributable to the fact that teachers in rural areas average more years experience and smaller class sizes than those in urban schools (Provasnik et al., 2007) allowing for students to grow at a higher rate in rural districts. Systematic mechanisms beyond location were present in these remote districts that influenced outcomes for students.

The effect of *Time in SpEd* on linear growth in reading (<0.01) and the effect of *Time in SpEd* on initial mathematics achievement were moderated slightly (<0.01) by a district being remote. These relationships may be due to the fact that SWDs in rural areas likely received related services from individuals with smaller caseloads than those in urban areas (e.g., Provasnik et al., 2007), a metric that was not available to include in analyses given the extant nature of the data. Additionally, SWDs eligible for increased services in these districts on average started lower than students who spent less time eligible for SpEd services, and may have had more room to grow to meet academic expectations. SWDs in remote districts had experiences that influenced their education, rather than the actual remote location having an influence on outcomes for students.

There were relations between the percentage of LEP students in districts and higher initial achievement for the reference group in both subjects as well as higher mathematics growth. This may be attributed to increased awareness of evidence-based practices in districts providing services to a greater number of students with LEP. For example, districts with greater number of students with LEP, may be implementing more strategies to assist struggling readers (e.g., Rosenshine, 2012) which may influence the performance of all students in these districts. Alternately, the student demographic model demonstrated that the growth of students with LEP was approximately a point higher than the reference group in both subjects. Increasing the concentration of students in a district that are growing at a higher rate per year than their peers may influence the performance of students district-wide.

The percentage of students that were FRL-eligible in districts was related to lower average initial reading and mathematics achievement, and slightly higher reading growth. This aligns with Sirin's (2005) findings that SES and academic achievement were related, yet may stand in contrast to part of the work of Lee & Reeves (2012), who found that the percent of economically disadvantaged students was not a significant predictor of academic growth in reading **or** mathematics at the state level. As the magnitude of the increase in reading growth was small (0.06 points per year), while statistically significant it is likely not a practically significant relationship.

Introducing district-level demographics to the models resulted in a slight change in the magnitude of the effect of *Time in SpEd* on mathematics (difference of -0.07, *SE* = 0.02) achievement, as well as lowering the slope for reading (difference of -0.01, *SE* = 0.01) as compared to the reference group. These relations between district demographic variables and growth for SWDs based on eligibility to receive services were no larger in magnitude than the standard errors. Introduction of district demographic variables explained approximately 4% variance in reading growth and 13% of the variance in math growth between districts. Combined with estimates for the reference group from the district demographic model, the total estimate for student growth factoring in time eligible for services (reference group + 10% *Time in SpEd*) was 0.12 points per year lower in reading (*SE* = 0.06, 0.01 for reference group and for *Time in SpEd,* respectively) and was not significantly higher than the reference group than in the student demographic model.

Modeling the growth of SWDs including district demographics indicated that there are relationships between district-level factors and overall student outcomes, as well as moderation of the effect of time eligible for SpEd services on reading growth and mathematics initial achievement. Holding all other factors constant, district context can influence outcomes for students. These relationships are very small in magnitude, which is not surprising, as district factors are more distal to the actual day-to-day learning of students than their own demographic characteristics.

Research question three: district funding. The only funding variable that moderated the effect of time eligible for SpEd on both reading and mathematics outcomes was district average SpEd funding per SWD. This was related to slightly higher initial mathematics achievement, and slightly lower reading and mathematics growth than students with similar needs in districts with average SpEd funding per SWD. In terms of this lower rate of growth, higher spending per SWD in districts might have corresponded with a percentage of SWDs that required greater monetary supports to access education appropriately (e.g., Chambers et al., 2003; Harr et al., 2008), rather than an intent by the district to allocate additional SpEd funds to each SWD to improve student performance. If a district had some SWDs with more costly required service, higher average SpEd funding might not have corresponded with higher funding *per* SWD, and thus did not relate to a corresponding increase in performance.

Some increases in funding over the six-year period were related to outcomes for SWDs. District increases in SpEd funding per SWD moderated the effect of *Time in SpEd* on reading growth (<0.01 per 10% increase) and increases in total funding per student moderated the effect of *Time in SpEd* on initial mathematics achievement (<0.01

per 10% increase). These finding provide some support for those of Hedges et al. (1994) who found a systematic relationship between inputs into education and outcomes, although those from the current study are very small in magnitude.

To add some nuance to the conflicting positions of Hanushek (1989) and Hedges, Laine, and Greenwald (1994), students in districts with higher total funding per student, on average, had higher initial reading scores than students in districts with average total funding per student. A relationship existed between this source of funding and student initial achievement, which may be attributable to years of time spent in a higher funded district (e.g., Grades K-3) before the first year of data were collected at the end of Grade 3. However, this relationship was so small in magnitude (0.07 points) that it is likely not practically significant, which aligns with the position of Hanushek (1989) that variations in expenditures were not systematically related to student performance – particularly not at the level of Hedges, Laine, and Greenwald (1994) who found that a $500 increase in per pupil funding was related to a 0.7 standard deviation increase in student achievement. Although relationships existed between district funding variables and outcomes for students, the extent to which these relationships can inform district fiscal policy is limited. There was no change in the magnitude of growth for the reference group or based on time eligible to receive SpEd services following the addition of district funding variables.

Limitations

Many factors were considered when designing this study, with the intent of appropriately documenting district-level factors and student-level characteristics. Despite these precautions, there are limitations to the method used to categorize SWDs, and associated with the use of proxies to account for student- and district-level demographics in this study. These factors are discussed in detail below.

Ideally, when examining growth for SWDs it would not be under that broad umbrella, but instead within specific exceptionalities, as growth for students with different exceptionality types can be different not only within-group but between group (e.g., Stevens et al., 2015, Wei et al., 2011). However, exceptionality classifications can shift as students progress through school (e.g., Ysseldyke & Bielinski, 2002). Accounting for changes in exceptionality type would require either use of a cross-classified design, or "fixing" a student's classification at a specific point in time, which limits how nuanced an investigation using exceptionality type actually is. This becomes more complicated when introducing a third level (district, in the case of this study) into multilevel modeling, as some exceptionality types have low membership, such as students with deaf-blindness, traumatic brain injury, or visual impairment. Classifying students by exceptionality type and also nesting students within district would further reduce numbers of students being compared across districts within student groups, and district-level effects would be estimated based on small numbers of students. The purpose of this study, in part, was to explore district-level sources of variance. Although small in magnitude, this study indicated that there are factors that contribute to variation between districts in terms of both demographics and funding.

The choice made to group students by the percentage of time SWDs were eligible to receive SpEd services rather than by exceptionality type was an effort to provide a level of differentiation to SWDs while avoiding some of the constraints of low membership associated with some exceptionality types. Results from this study suggest that looking at SpEd status in this manner was valid, both in terms of results from hierarchical linear models, as well as achievement gap effect sizes. In this study I provided no information about the effects of specific student exceptionality types on student achievement. However, using time eligible to receive SpEd services is more fine-grained than classifying SWDs at a specific point in time and not adjusting for changes in eligibility.

Although I examined growth for SWDs, accounting for student-level demographics and district-level characteristics, many additional sources of variance existed that were not modeled in this study. Fine-grained information about students' lives outside of school, and a spectrum of student-level demographics beyond proxies such as FRL and LEP status may provide more information about growth for SWDs. In particular, FRL is a rather poor proxy for the actual socioeconomic status of students (e.g., Reardon & Robinson, 2008), and membership in this subgroup accounted for approximately 70% of SWDs. Although FRL is commonly used as a proxy for socioeconomic status at the district and state level, previous research has determined that parental education, income, and parental occupation provide better understanding of socioeconomic status than the FRL label (e.g., Reardon & Robinson, 2008; Sirin, 2005). This choice to use such proxies as FRL instead of seeking out more fine-grained information was to mirror AYP subgroups – the actual student groups districts were concerned about with respect to NCLB reporting.

In addition, while funding variables were explored in this study, limited variability in district funding per student (M = 20.37, SD = 4.13) and district SpEd funding per SWD (M = 18.47, SD = 3.36) limited the extent to which the influence of these variables on student growth could be explored. There was more variability in the percent change in these two variables (*M* = 10.89, *SD* = 10.65; *M* = 19.02, *SD* = 22.29, respectively) between districts. Nationwide, Corcoran & Evans (2008) found more variability in funding between-state than within-state. The extent to which the identified relationships between funding and growth for SWDs in this study generalize beyond Oregon are likely limited.

Beyond student-level factors, a host of other structural and situational variables may influence growth for SWDs at the district level, such as curriculum choices, professional development, and whether or not a district is using multi-tiered systems of support (MTSS) in order to identify and support students academically and behaviorally. Although beyond the scope of this study, these factors exist, and may influence growth for SWDs. In addition, current transitions from state accountability measures to consortia-based assessments of proficiency in the Common Core State Standards may limit the generalizability of these results to future investigations of growth for SWDs.

Conclusions and Recommendations for Future Research

Do dollars matter beyond demographics? In a nutshell, it depends. The magnitude of relationships between district-level funding and reading and mathematics outcomes for students was very small, and the amount of variance between districts associated with the amount of time SWDs were eligible for SpEd services was also very small (<0.01). This indicates that a focus on relationships between-districts for SWDs may not really be warranted.

However, aspects of this study suggest that the relationship of district funding to student outcomes may be more complex than these results indicate. For example, there were relationships for SWDs (reading growth and mathematics initial achievement) and for the reference group (reading and math initial achievement, reading growth) for students situated in a remote district. This is not to say that resituating all students in remote districts would lead to positive relationships with reading growth. Instead, there are systematic influences in remote districts that influence reading growth of students, which may include class size, number of SWDs on service providers' caseloads, relationships built living in smaller communities, as well as other potential school and district-level characteristics. These factors were not available in the extant datasets used for this research, but should be considered when planning future investigations into the relations between funding and growth for SWDs.

Although there would have been more variability in funding at the school level, the choice to use district demographic and financial data rather than school data for this study was due, in part, to the fact that SpEd funding, one of the primary variables of interest, was not reported at the school level in Oregon, but at the district level. The choice to nest students within districts was finalized after considering other district-driven sources of variation that were not explored in this study, but provided some consistencies in educational experiences for students within districts. Local-control of education in Oregon means that, even with state-led initiatives, local school districts and their school boards control the execution of these initiatives. Choice of curriculum, types of professional development available to educators (some of which would be determined by whether or not districts led schools to use MTSS to support students), or the decision to use curriculum-based measures (e.g., easyCBM or Dynamic Indicators of Basic Early Learning Skills [DIEBELS]) to monitor within-year progress of students and use data to make decision about instruction were all influenced by districts. Nesting SWDs in districts had not been explored previously, and in doing so, this study filled a necessary gap in the literature.

As the district-level factors related to growth for SWDs based on time eligible to receive SpEd services have been demonstrated to be small in magnitude, future studies should examine the relationships between these same demographic and funding factors at the school level rather than the district level. This would help to clarify whether the variables themselves have relations to growth for SWDs that are worth exploring in detail. It may be that nesting SWDs within schools would perhaps better speak to the relations between demographics, funding and growth for SWDs than district-level factors. The greater variability in funding at the school level would perhaps relate to student outcomes to a different degree than district-level factors. Although overall policy is generally set at the district-level, and trickles down to schools (e.g., Levenson, 2012), school-level variables are more proximal to students, and may have stronger relations to growth for SWDs.

Part of the reason that the included funding variables had a limited influence on outcomes in this study may be attributed to the lack of a clear picture of where the funding actually goes. For example, average SpEd funding per SWD was examined, yet some students require extensive services that are costly. Information about the actual cost of services per SWD could help to provide a better understanding of the relation of district funding to growth for SWDs. These data were not available for purposes of this study, but could be gathered during a school year at either the district or school level. The relationship between actual SpEd funding per SWD and within-year growth of

SWDs could be investigated using scores on curriculum based measures such as easyCBM as outcome variables. Additional variables to consider when investigating whether or not money matters at the school level could include class size, quality of personnel, evidence-based practices implemented at the school level, and school-wide utilization of student data to drive instruction.

Modeling growth for SWDs requires a host of decisions up front about how to most appropriately group students within this AYP subgroup. Researchers have begun to examine SWDs in the context of state accountability measures in both reading and mathematics (e.g., Schulte, 2010; Schulte & Stevens, 2015; Schulte & Villwock, 2004; Stevens et al., 2015; Zvoch & Stevens, 2005) and to explore different methods for grouping SWDs within these analyses (e.g., Schulte & Stevens, 2015). However, even this recent focus on growth for SWDs using state accountability measures provides a relatively narrow view of these students.

Future studies are needed to assess growth for SWDs who are also members of one or more of the other AYP subgroups examining the interactions between these student demographics and SWD status rather than using student demographics as control variables. In this manner, greater understanding can be gained as to what typical growth for SWDs may be, taking into account student demographic information. In addition to multilevel approaches, achievement gap effect sizes (Bloom et al., 2008) could be constructed for groups of SWDs who are also members of other subgroups – an "interaction" achievement gap effect size so to speak. Examination of the performance of students who fall into multiple AYP subgroups could provide a useful resource for educators when assessing how SWDs perform relative to what is considered average growth for students with similar characteristics, and inform decisions around the impact of intervention.

These examinations of the interactions between AYP subgroups should be replicated across states in multiple regions of the country in order to determine whether patterns of growth for SWDs who are members of additional subgroups are consistent across regions. Results from the current study indicate that district location, as well as the proportion of LEP and FRL-eligible students in a district may be related to overall student growth. These relations may vary depending on the region of the country in which SWDs receive services. Investigating and comparing the growth of SWDs in multiple regions of the country would provide insight into whether or not the average performance of SWDs varies by region, which could help teachers make more informed decisions about how to meet the needs of the SWDs they serve.

Within the context of AYP status models, a focus on students in isolated subgroups was inevitable. As states design and adapt accountability systems to meet local needs and best serve students, it is critical to consider factors that influence growth for SWDs, beyond eligibility to receive services. Further investigations into the ways in which funding influences outcomes for SWDs are necessary in order to more appropriately determine the relations between investments and student outcomes. In addition, moving towards describing all students in terms of comprehensive demographics rather than splitting them up by individual traits is necessary in order to more accurately investigate student growth.

APPENDIX

FULL HLM MODELS

Table A.1

Fixed and Random Effects, Quadratic HLM Model, Reading Grades 3 to 8.

	Intercept	Linear	Quadratic
Fixed effects	211.35 (0.25)**	5.37 (0.04)**	-0.24 (0.02)**
Random effects	Intercept	Linear	Quadratic
Level-2 Students	182.23**	11.14**	0.19**
Level-1 Residual	20.89		
Level-3 Districts	15.63**	1.40**	0.04**
Δ Deviance, x^2 (*df*)	10638.12(7)**		

Note. Standard errors shown in parentheses. Pseudo-R^2 value was compared to baseline model, and is expressed a percent. * $p < .05$, ** $p < .001$

Table A.2

Fixed and Random Effects, Full District Demographic Model, Reading Grades 3 to 8.

Fixed Effects	Intercept	Linear
Mean	218.96 (0.43)**	3.98 (0.08)**
Student Level		
SpEd	-1.35 (0.05)**	0.07 (0.01)**
Male	-0.28 (0.11)*	-0.13 (0.02)**
Non-White	-1.73 (0.20)**	0.07 (0.03)*
LEP	-7.95 (0.24)**	0.86 (0.04)**
FRL	-3.17 (0.17)**	0.02 (0.02)
District level		
Size	< 0.01 (< 0.01)	< 0.01 (< 0.01)*
Remote	-1.11 (0.44)*	0.12 (0.10)
% SWD	-0.13 (0.69)	-0.19 (0.16)
% NW	-0.30 (0.32)	-0.02 (0.06)
% LEP	0.94 (0.54)	0.05 (0.09)
% FRL	-0.74 (0.17)**	0.07 (0.04)
Size on SWDs	< 0.01 (< 0.01)	< -0.01 (< 0.01)
Remote on SWDs	< 0.01 (< 0.01)	< 0.01 (< 0.01)
% SWD on SWDs	0.01 (0.01)	< 0.01 (< 0.01)
% NW on SWDs	< -0.01 (< 0.01)	< -0.01 (< 0.01)
% LEP on SWDs	0.01 (0.01)	< 0.01 (< 0.01)
% FRL on SWDs	< 0.01 (< 0.01)	< -0.01 (< 0.01)
Random Effects	**Intercept**	**Linear**
Level-2 Students	73.91**	1.08**
Level-1 Residual	23.55	
Level-3 Districts	2.97**	0.14**
in SpEd	<0.01**	< 0.01**
in male	0.20	0.01
in non-White	1.87**	0.04**
in LEP	1.93**	0.04*
in FRL	1.72**	0.01
Student-level pseudo-R^2	-0.02	0.01
District-level pseudo-R^2	25.40	5.08
Δ Deviance, x^2 (*df*)	78.65 (24)**	

Note. Standard errors shown in parentheses. Coefficients for SpEd, %NW, %LEP and %FRL are reported in increments of 10%. Pseudo-R^2 value was compared to student demographic model, and is expressed as a percent. * $p < .05$, ** $p < .001$

Table A.3

Fixed Effects, Full District Funding Model, Reading Grades 3 to 8.

Fixed Effects	Intercept	Linear
Mean	219.15 (0.27)**	3.91 (0.06)**
Student Level		
SpEd	-1.32 (0.03)**	0.08 (0.01)**
Male	-0.28 (0.10)*	-0.13 (0.02)**
Non-White	-1.73 (0.19)**	0.06 (0.03)*
LEP	-7.93 (0.24)**	0.86 (0.04)**
FRL	-3.18 (0.18)**	0.02 (0.02)
District level		
Remote	-1.25 (0.31)**	0.20 (0.07)*
% LEP	0.78 (0.16)**	--
% FRL	0.93 (0.13)**	0.07 (0.03)*
Remote on SWDs	--	< 0.01 (< 0.01)*
Total funding per student	0.11 (0.04)*	-0.02 (0.01)
Sped funding per SWD	0.01 (0.05)	< -0.01 (0.01)
10 % Δ funding per student	0.28 (0.18)	-0.03 (0.05)
10 % Δ Sped funding per SWD	-0.11 (0.08)	0.04 (0.02)
Total funding per student on SWDs	< -0.01 (< 0.01)	< 0.01 (< 0.01)
SpEd funding per SWD on SWDs	< 0.01 (< 0.01)*	< -0.01 (0.01)
10 % Δ funding per student on SWDs	0.01 (< 0.01)*	< -0.01 (< 0.01)
10 % Δ Sped funding per SWD on SWDs	< -0.01 (<0.01)	< 0.01 (< 0.01)*

Note. Standard errors shown in parentheses. Total funding per student and SpEd funding per SWD were estimated in increments of $500. Pseudo-$R^2$ value was compared to district demographic model, and is expressed as a percent. * $p < .05$, ** $p < .001$

Table A.4

Random Effects and Model Fit, Full Funding HLM Model, Reading Grades 3 to 8.

Random Effects	Intercept	Linear
Level-2 Students	73.90**	1.08**
Level-1 Residual	23.55	
Level-3 Districts	2.72**	0.14**
in SpEd	< 0.01**	< 0.01**
in male	0.22	0.01
in non-White	1.87**	0.04**
in LEP	2.02**	0.05*
in FRL	1.32**	0.01
Student-level pseudo-R^2	0.02	0.05
District-level pseudo-R^2	3.09	-1.68
Δ Deviance, x^2 (*df*)	30.20 (16)**	

Note. Pseudo-R^2 value was compared to the final district demographics model, and is expressed as a percent. * $p < .05$, ** $p < .001$

Table A.5

Fixed and Random Effects, Longitudinal Quadratic HLM Model, Mathematics Grades 3 to 8.

	Intercept	Linear	Quadratic
Fixed effects	208.79 (0.26)**	7.39 (0.13)**	-0.42 (0.02)**
Random effects	Intercept	Linear	Quadratic
Level-2 Students	97.78**	9.40**	0.20**
Level-1 Residual	21.95		
Level-3 Districts	8.37**	2.12**	0.07**
Δ Deviance, x^2 (*df*)	14,045.28 (7)**		

Note. Standard errors shown in parentheses. Pseudo-R^2 value was compared to baseline model, and is expressed a percent. * $p < .05$, ** $p < .001$

Table A.6

Fixed and Random Effects, Full District Demographics, Mathematics Grades 3 to 8.

Fixed Effects	Intercept	Linear
Mean	212.67 (0.43)**	5.38 (0.11)**
Student Level		
SpEd	-1.11 (0.04)**	0.02 (0.01)*
Male	1.96 (0.09)**	-0.24 (0.02)**
Non-White	-1.49 (0.18)**	0.05 (0.03)
LEP	-5.00 (0.21)**	0.56 (0.04)**
FRL	-2.65 (0.17)**	-0.23 (0.03)**
District level		
Size	< 0.01 (< 0.01)	< 0.01 (< 0.01)*
Remote	-0.93 (0.48)	0.08 (0.13)
% SWD	1.13 (0.85)	-0.55 (0.23)*
% NW	0.12 (0.30)	-0.10 (0.07)
% LEP	0.64 (0.46)	0.19 (0.10)
% FRL	-0.91 (0.18)**	0.08 (0.05)
Size on SWDs	< 0.01 (< 0.01)**	< -0.01 (< 0.01)
Remote on SWDs	0.02 (< 0.01)**	< -0.01 (< 0.01)
% SWD on SWDs	< 0.01 (0.01)	< 0.01 (< 0.01)
% NW on SWDs	-0.01 (< 0.01)	< 0.01 (< 0.01)
% LEP on SWDs	0.01 (< 0.01)	< -0.01 (< 0.01)*
% FRL on SWDs	< -0.01 (< 0.01)	< -0.01 (< 0.01)
Random Effects	**Intercept**	**Linear**
Level-2 Students	62.41**	1.07**
Level-1 Residual	25.52	
Level-3 Districts	4.75**	0.27**
in SpEd	<0.01**	< 0.01**
in male	0.14	< 0.01
in non-White	1.85**	0.01
in LEP	1.91**	0.05
in FRL	1.83**	0.04**
Student-level pseudo-R^2	-0.01	0
District-level pseudo-R^2	14.87	17.39
Δ Deviance, x^2 (*df*)	109.22 (24)**	

Note. Standard errors shown in parentheses. Coefficients for SpEd, %NW, %LEP and %FRL are reported in increments of 10%. Pseudo-R^2 value was compared to student demographic model, and is expressed as a percent. * $p < .05$, ** $p < .001$

Table A.7

Fixed Effects, Full District Funding Model, Mathematics Grades 3 to 8.

Fixed Effects	Intercept	Linear
Mean	212.72 (0.28)**	3.91 (0.06)**
Student Level		
SpEd	-1.07 (0.03)**	0.08 (0.01)**
Male	1.96 (0.09)*	-0.13 (0.02)**
Non-White	-1.50 (0.19)**	0.06 (0.03)*
LEP	-5.01 (0.21)**	0.86 (0.04)**
FRL	-2.65 (0.17)**	0.02 (0.02)
District level		
Remote	-0.94 (0.26)**	--
% LEP	0.61 (0.02)*	0.13 (0.04)*
% FRL	-0.73 (0.11)**	--
Remote on SWDs	0.01 (<0.01)**	--
Total funding per student	0.04 (0.07)	< 0.01 (< 0.01)
Sped funding per SWD	0.02 (0.07)	< 0.01 (0.02)
10 % Δ funding per student	0.27 (0.02)	0.01 (0.05)
10 % Δ Sped funding per SWD	-0.04 (0.01)	-0.01 (0.03)
Total funding per student on SWDs	< 0.01 (< 0.01)	< 0.01 (< 0.01)
SpEd funding per SWD on SWDs	< 0.01 (< 0.01)*	< -0.01 (0.01)*
10 % Δ funding per student on SWDs	0.01 (< 0.01)*	< -0.01 (< 0.01)*
10 % Δ Sped funding per SWD on SWDs	< -0.01 (<0.01)	< 0.01 (< 0.01)

Note. Standard errors shown in parentheses. Total funding per student and SpEd funding per SWD were estimated in increments of $500. Pseudo-$R^2$ value was compared to district demographic model, and is expressed as a percent. * $p < .05$, ** $p < .001$

Table A.8

Random Effects and Model Fit, Full Funding HLM Model, Mathematics Grades 3 to 8.

Random Effects	Intercept	Linear
Level-2 Students	62.41**	1.07**
Level-1 Residual	25.53	
Level-3 Districts	4.42**	0.29**
in SpEd	< 0.01**	< 0.01**
in male	0.17	0.01
in non-White	1.79**	0.01
in LEP	1.88*	0.05
in FRL	1.83**	0.04**
Student-level pseudo-R^2	0	0
District-level pseudo-R^2	1.85	0
Δ Deviance, x^2 (*df*)	21.09 (16)**	

Note. Pseudo-R^2 value was compared to the final district demographics model, and is expressed as a percent. * $p < .05$, ** $p < .001$

REFERENCES CITED

Abedi, J. (2004). The No Child Left Behind Act and English language learners: Assessment and accountability issues. *Educational Researcher, 33*, 4-14. doi: 10.3102/0013189X033001004

Ahearn, E. (2009). Growth models and students with disabilities: Report of state interviews (Vol. 15, pp. 2012). Alexandria, VA: National Association of State Directors of Special Education.

American Institutes for Research. (2012). 2011-12 growth model for educator evaluation technical report: Final: New York State Education Department.

Aud, S., Hussar, W., Kena, G., Bianco, K., Frohlich, L., Kemp, J., & Tahan, K. (2011). The condition of education 2011. NCES 2011-033: National Center for Education Statistics.

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. *Journal of Research on Educational Effectiveness, 1*(4), 289-328. doi: 10.1080/19345740802400072

Bryk, A.S., Raudenbush, S.W., & Congdon, R. (2004). HLM for Windows. Chicago, IL: Scientific Software International, Inc.

Chambers, J.G., Parrish, T.B., Esra, P.E., & Shkolnik, J.L. (2002). How Does Spending on Special Education Students Vary across Districts? An Analysis of Spending by Urbanicity, District Size, Median Family Income, and Student Poverty Levels in 1999-2000. *Special Education Expenditure Project (SEEP)* (pp. 1-25). Palo Alto, CA: Center for Special Education Finance, American Institutes for Research in the Behavioral Sciences.

Chambers, J.G., Shkolnik, J., & Pérez, M. (2003). Total expenditures for students with disabilities, 1999-2000: Spending variation by disability. In J. J. Harr (Ed.), *Special Education Expenditure Project (SEEP)* (pp. 1-131). Washington, DC: Center for Special Education Finance, American Institutes for Research.

Choi, S.W., & Tinkler, T. (2002). *Evaluating comparability of paper and pencil and computer based assessment in a K-12 setting.* Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, L.A.

Chudowsky, N., & Chudowsky, V. (2009). State test score trends through 2007-08, part 4: Has progress been made in raising achievement for students with disabilities? Washington, DC: Center on Education Policy.

Clotfelter, C.T., & Ladd, H.F. (1996). Recognizing and rewarding success in public schools. In H. F. Ladd (Ed.), *Holding Schools Accountable: Performance-Based Reform in Education* (pp. 23-64). Washington DC: Brookings Institution Press.

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences (2nd ed.)*. Mahwah, NJ: Erlbaum.

Coleman, J.S., Campbell, E.Q., Hobson, C.J., McPartland, J., Mood, A.M., Weinfeld, F.D., & York, R.L. (1966). Equality of educational opportunity. Washington, D.C.: U.S. Government Printing Office.

Corcoran, S.P., & Evans, W.N. (2008). Equity, adequacy and the evolving state role in education finance. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of Research in Education Finance and Policy* (pp. 332-356). New York: Routledge.

Dee, T.S., Jacob, B., & Schwartz, N.L. (2013). The effects of NCLB on school resources and practices. *Educational Evaluation and Policy Analysis, 35*(2), 252-279. doi: 10.3102/0162373712467080

Deno, S.L., Fuchs, L.S., Marston, D., & Shin, J. (2001). Using curriculum-based measurement to establish growth standards for students with learning disabilities. *School Psychology Review, 30*(4), 507-524.

Doran, H., & Cohen, J. (2006). Oregon technical report 1. Technical report for the Oregon Department of Education.

Driscoll, D., Halcoussis, D., & Svorny, S. (2003). School district size and student performance. *Economics of Education Review, 22*(2), 193-201. doi: 10.1016/S0272-7757(02)00002-X

Dunn, J.L., & Allen, J. (2009). Holding schools accountable for the growth of nonproficient students: Coordinating measurement and accountability. *Educational Measurement: Issues and Practice, 28*(4), 27-41.

Eckes, S., & Swando, J. (2009). Special education subgroups under NCLB: Issues to consider. *The Teachers College Record, 111*(11), 2479-2504.

Enders, C.K. (2010). *Applied missing data analysis*. New York, NY: Guilford Press.

Federal Education Budget Project. (2015, April 5). Oregon.   Retrieved April 5, 2015, from http://febp.newamerica.net/k12/or - Funding

Figlio, D.N., & Ladd, H.F. (2008). School accountability and student achievement. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of Research in Education Finance and Policy* (pp. 166-182). New York: Routledge.

Fulton, M. (2006). Minimum subgroup size for adequate yearly progress (AYP): State trends and highlights *State Notes: No Child Left Behind*. Denver, CO: Education Commission of the States.

Gándara, P.C., & Rumberger, R.W. (2006). Resource needs for California's English learners. Stanford, CA: Institute for Research on Education Policy & Practice.

Goldhaber, D. (2002). What might go wrong with the accountability measures of the "No Child Left Behind Act?" *No Child left Behind: What Will It Take* (pp. 89-102): Thomas B. Fordham Foundation.

Gordon, N.E. (2008). The changing federal role in education finance and governance. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of Research in Education Finance and Policy* (pp. 295-313). New York: Routledge.

Hanushek, E.A. (1989). The impact of differential expenditures on school performance. *Educational Researcher, 18*, 45-51+62.

Hanushek, E.A. (2003). The failure of input‐based schooling policies. *The economic journal, 113*(485), F64-F98. doi: 10.1111/1468-0297.00099

Hanushek, E.A., Kain, J.F., & Rivkin, S.G. (2002). Inferring program effects for special populations: Does special education raise achievement for students with disabilities? *Review of Economics and Statistics, 84*(4), 584-599. doi: 10.1162/003465302760556431

Hanushek, E.A., & Raymond, M.E. (2004). The effect of school accountability systems on the level and distribution of student achievement. *Journal of the European Economic Association, 2*(2-3), 406-415.

Harr, J.J., Parrish, T.B., & Chambers, J.G. (2008). Special Education. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of Research in Education Finance and Policy* (pp. 573-590). New York: Routledge.

Hedges, L.V., Laine, R.D., & Greenwald, R. (1994). An exchange: Part I: Does money matter? A meta-analysis of studies of the effects of differential school inputs on student outcomes. *Educational researcher, 23*, 5-14.

Hopkins, M., Thompson, K.D., Linquanti, R., Hakuta, K., & August, D. (2013). Fully accounting for English learner performance a key issue in ESEA reauthorization. *Educational Researcher*, 0013189X12471426.

Individuals with Disabilities Education Act 20 U.S.C., Pub. L. No. 105-117 § 1400 et seq. (1997).

Kane, T.J., & Staiger, D.O. (2002). The promise and pitfalls of using imprecise school accountability measures. *The Journal of Economic Perspectives, 16*(4), 91-114. doi: 10.1257/089533002320950993

Killeen, K., & Sipple, J. (2000). School consolidation and transportation policy: An empirical and institutional analysis (pp. 1-43). Washington, DC: Rural School and Community Trust.

Ladd, H.F. (1996). How school districts respond to fiscal constraint. In W. J. Fowler (Ed.), *Selected papers in school finance* (pp. 39-59). Washington, DC: National Center of Education Statistics.

Ladd, H.F., & Ferguson, R. (1996). How and Why Money Matters: An Analysis of Alabama Schools' *Holding Schools Accountable: Performance-Based Reform in Education*. Washington, DC: Brookings Institution.

Ladd, H.F., & Lauen, D.L. (2010). Status versus growth: The distributional effects of school accountability policies. *Journal of Policy Analysis and Management, 29*(3), 426-450. doi: 10.1002/pam.20504

Ladd, H.F., & Walsh, R.P. (2002). Implementing value-added measures of school effectiveness: getting the incentives right. *Economics of Education review, 21*(1), 1-17. doi: 10.1016/S0272-7757(00)00039-X

Lee, J., & Reeves, T. (2012). Revisiting the impact of NCLB high-stakes school accountability, capacity, and resources state NAEP 1990–2009 reading and math achievement gaps and trends. *Educational Evaluation and Policy Analysis, 34*(2), 209-231. doi: 10.3102/0162373711431604

Levenson, N. (2012). Boosting the quality and efficiency of special education. Washington, DC: Thomas B. Fordham Institute.

Linn, R.L. (2008). Toward a more effective definition of Adequate Yearly Progress. In G. L. Sunderman (Ed.), *Holding NCLB accountable: Achieving accountability, equity, & school reform* (pp. 27-42). Thousand Oaks, CA: Corwin Press.

Mathis, W.J. (2003). No Child Left Behind: What are the costs? Will we realize any benefits.

Mathis, W.J. (2004). No Child Left Behind Act: What Will It Cost States? *Spectrum: The Journal of State Government, 77*, 8-10.

McCaffrey, D.F., Sass, T.R., Lockwood, J.R., & Mihaly, K. (2009). The inter-temporal variability of teacher effect estimates. Nashville, TN: National Center on Performance Initiatives, Vanderbilt University.

McCall, M.S. (2005). Technical review of the Oregon Department of Education assessment system. Portland, OR: NWEA.

McLaughlin, M.J., & Thurlow, M.L. (2003). Educational accountability and students with disabilities: Issues and challenges. *Educational Policy, 17*(4), 431-451. doi: 10.1177/0895904803254962

McLoyd, V.C. (1998). Socioeconomic disadvantage and child development. *American Psychologist, 53*(2), 185-204. doi: 10.1037/0003-066X.53.2.185

Moores, D.F. (2005). The No Child Left Behind and the Individuals With Disabilities Education Acts: The uneven impact of partially funded federal mandates on education of deaf and hard of hearing children. *American Annals of the Deaf, 150*, 75-79. doi: 10.1353/aad.2005.0028

Nagle, K., Yunker, C., & Malmgren, K.W. (2006). Students with disabilities and accountability reform challenges identified at the state and local levels. *Journal of Disability Policy Studies, 17*, 28-39. doi: 10.1177/10442073060170010301

Nakamoto, J., Lindsey, K.A., & Manis, F.R. (2007). A longitudinal analysis of English language learners' word decoding and reading comprehension. *Reading and Writing, 20*(7), 691-719.

National Council of Teachers of Mathematics. (2006). Curriculum focal points for prekindergarten through grade 8 mathematics: A quest for cohereence. Reston, VA: Author.

No Child Left Behind Act, 20 U.S.C., Pub. L. No. 107-110 § 1424 *et seq.*, 6301 Stat. (2002).

Northwest Evaluation Association. (2005). Oregon statewide assessment system: Standard errors of measurement, reading and mathematics knowledge and skills tests. Portland, OR: NWEA.

Oregon Department of Education. (1999-2013). *ESD Direct Support Maintenance of Effort*. Retrieved from: http://www.ode.state.or.us/sfda/reports/r0096Select.asp

Oregon Department of Education. (2007a). Oregon Statewide Report Card: 2006-2007. Salem, OR: ODE.

Oregon Department of Education. (2007b). Oregon's statewide assessment system: Reliability and validity. Salem, OR: ODE.

Oregon Department of Education. (2012a). 2011-12 Adequate yearly progress (AYP) policy and technical manual. Salem, OR: ODE.

Oregon Department of Education. (2012b). Oregon's statewide assessment system. Salem, OR: ODE.

Oregon Department of Education. (2013). The LEA maintenance of effort (MOE) audit process. Retreived from http://www.ode.state.or.us/data/schoolanddistrict/funding/sped/moedescription.pdf

Oregon Department of Education. (2014). Frequently asked questions: Assessment reports. Retrieved September 21,2014, from http://www.ode.state.or.us/apps/faqs/index.aspx?=87 - 608

Oregon Legislative Revenue Office. (2004). K-12 and ESD school finance: State school fund distribution. Salem, OR.

Porter, A.C., Linn, R.L., & Trimble, C.S. (2005). The effects of state decisions about NCLB adequate yearly progress targets. *Educational Measurement: Issues and Practice, 24*, 32-39. doi: 10.1111/j.1745-3992.2005.00021.x

Provasnik, S., KewalRamani, A., Coleman, M.M., Gilbertson, L., Herring, W., & Xie, Q. (2007). Status of education in rural America (NCES 2007-040). Washington, DC: National Center for Education Statistics, Institute of Education Sciences, U.S. Department of Education.

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

Reardon, S.F., & Robinson, J.P. (2008). Patterns and trends in racial/ethnic and socieconomic academic achievement gaps. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of Research in Education Finance and Policy* (pp. 497-516). New York: Routledge.

Rentschler, T. (2005). No Child Left Behind: Admirable goals, disastrous outcomes. *Widener Law Review, 12*, 637-663.

Rice, J.K., & Schwartz, A.E. (2008). Toward an understanding of productivity in education. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of Research in Education Finance and Policy* (pp. 131-145). New York: Routledge.

Rose, L.C. (2004). No Child Left Behind: The mathematics of guaranteed failure. *Educational Horizons, 82*(2), 121-130.

Rosenshine, B. (2012). Principles of instruction: Research-based strategies that all teachers should know. *American Educator, 36*, 12-19.

Rouse, C.E., Hannaway, J., Goldhaber, D., & Figlio, D. (2013). Feeling the Florida heat? How low-performing schools respond to voucher and accountability pressure. *American Economic Journal: Economic Policy, 5*(2), 251-281. doi: 10.1257/pol.5.2.251

Rumberger, R.W., & Gándara, P.C. (2008). Resource needs for educating linguistic minority students. In H. F. Ladd & E. B. Fiske (Eds.), *Handbook of Research in Education Finance and Policy* (pp. 591-611). New York: Routledge.

Saven, J.L., Anderson, D., Nese, J.F.T., Farley, D., & Tindal, G. (2015). Patterns of statewide test participation for students with significant cognitive disabilities. *The Journal of Special Education*. Advance Online Publication. doi: 10.1177/0022466915582213

Schulte, A.C. (2010). Assessing growth for students with disabilities on large scale assessments. In J. E. Ysseldyke (Ed.), *Applications of growth models to making decisions about states, schools, programs, and students*: Symposium conducted at the annual National Conference on Student Assessment, Detroit, MI.

Schulte, A.C., & Stevens, J.J. (2015). Once, sometimes, or always in special education: Do inclusion criteria change portrayals of mathematics growth and achievement gaps for students with disabilities? *Exceptional Children*. Advance Online Publication. doi: 10.1177/0014402914563695

Schulte, A.C., & Villwock, D.N. (2004). Using high-stakes tests to derive school-level measures of special education efficacy. *Exceptionality, 12*(2), 107-126.

Schulte, A.C., Villwock, D.N., Whichard, S.M., & Stallings, C.F. (2001). High stakes testing and expected progress standards for students with learning disabilities: A five-year study of one district. *School Psychology Review, 30*(4), 487.

Shin, J., Espin, C.A., Deno, S.L., & McConnell, S. (2004). Use of hierarchical linear modeling and curriculum-based measurement for assessing academic growth and instructional factors for students with learning difficulties. *Asia Pacific Education Review, 5*(2), 136-148.

Shin, T., Davison, M.L., Long, J.D., Chan, C.-K., & Heistad, D. (2013). Exploring gains in reading and mathematics achievement among regular and exceptional students using growth curve modeling. *Learning & Individual Differences, 23*, 92-100. doi: 10.1016/j.lindif.2012.10.002

Simon, R. (2004). Calculating participation rates: Flexibility when making adequate yearly progress determinations. Retrieved September 26, 2014, from http://www2.ed.gov/policy/elsec/guid/stateletters/prates.html

Sirin, S.R. (2005). Socioeconomic status and academic achievement: A meta-analytic review of research. *Review of Educational Research, 75*(3), 417-453. doi: 10.3102/00346543075003417

Stevens, J.J., Schulte, A.C., 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.

Strayhorn, T.L. (2009). Accessing and analyzing national databases. In T. Kowalski & T. J. Lasley (Eds.), *Handbook of Data-Based Decision Making in Education* (pp. 105-122). New York: Routledge.

Thurlow, M.L., Wiley, H.I., & Bielinski, J. (2003). Going public: What 2000-2001 reports tell us about the performance of students with disabilities. Minneapolis, MN: NCEO.

U. S. Department of Education [USED]. (2014). No educator left behind: One hundred percent proficiency. Retrieved September 26, 2014, from http://www.educationworld.com/a_issues/NELB/NELB077.shtml

U.S. Department of Education. (2005). Alternate achievement standards for students with the most significant cognitive disabilities: Non-regulatory guidance.

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*. Advance Online Publication. doi: 10.1177/0741932512448253.

Wenglinsky, H. (1997). How money matters: The effect of school district spending on academic achievement. *Sociology of Education, 70*(3), 221-237.

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

Zehler, A.M., Fleischman, H.L., Hopstock, P.J., Stephenson, T.G., Pendzick, M.L., & Sapru, S. (2003). Policy report: Summary of findings related to LEP and SPED-LEP students. Washington, DC: Submitted by Development Associates, Inc. to the U.S. Department of Education.

Zhou, L. (2010). Revenues and expenditures for public elementary and secondary education: School year 2007-08 (fiscal year 2008) (NCES 2010-326). U.S. Department of Education. Washington, DC: National Center for Education Statistics.

Zvoch, K., & Stevens, J.J. (2005). Sample exclusion and student attrition effects in the longitudinal study of middle school mathematics performance. *Educational Assessment, 10*(2), 105-123. doi: 10.1207/s15326977ea1002_2

Zvoch, K., & Stevens, J.J. (2006a). Longitudinal effects of school context and practice on middle school mathematics achievement. *The Journal of Educational Research, 99*(6), 347-357. doi: 10.3200/JOER.99.6.347-357

Zvoch, K., & Stevens, J.J. (2006b). Successive student cohorts and longitudinal growth Models: An investigation of elementary school mathematics performance. *Education Policy Analysis Archives, 14*, 1-25.