ABSTRACT

TRAIFICANTE, AMANDA LYNN. Special Education Students’ Growth in Reading: A Meta-Analysis. (Under the direction of Dr. Scott Stage).

Students with disabilities tend to receive lower scores on measures of academic performance than their general education peers. Three theories have been used to explain this gap in academic performance: the developmental deficit hypothesis, the developmental lag hypothesis, and the Matthew effect. To determine the most supported theory of academic performance, the current researchers initially reviewed 214 studies to be used for a meta-analysis. Studies were included if the researchers of the studies investigated special education student growth over time in reading and mathematics. After applying inclusion and exclusion criteria, 11 studies remained for analysis of reading growth of students in special education versus their general education peers over time. Hierarchical linear modeling was used to analyze reading outcome data across four time points by special education eligibility, studies with above average samples of Black students, studies with above average samples of Hispanic students, studies with above average samples of students receiving free and/or reduced lunch, and weighted sample size. Significant differences for intercepts and slopes between each variable and its respective comparison group were not found. However, the outcome provides interesting descriptive information about trends for academic performance over time and indicates a need for more research analyzing unique samples of students.
Special Education Students’ Growth in Reading: A Meta-Analysis

by
Amanda Lynn Traficante

A thesis submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the degree of
Master of Science

Psychology

Raleigh, North Carolina

2018

APPROVED BY:

Scott Stage
Committee Chair

Kate Norwalk

Eui Kyung Kim
BIOGRAPHY

Amanda Traficante earned her B.S. in Psychology and her Addiction Specialist Certificate from Saint Vincent College in 2015. She is currently a graduate student in the School Psychology Ph.D. Program within the Department of Psychology at North Carolina State University. She first became interested in the field of School Psychology during her experience working as a counselor at the Summer Treatment Program at Florida International University in 2014.

Amanda began her research career during her undergraduate training at Saint Vincent College. Through these experiences, Amanda became passionate about conducting research and sought out multiple research opportunities. She presented her own work at the 86th Annual Meeting of the Eastern Psychological Association (2015). During graduate school, Amanda has served as a research assistant for multiple labs. Her work has been presented at the annual conference for the National Association of School Psychologists (2017). Current topics of interest include special education, differences for students in special education categories, teacher awareness of social dynamics, teacher best practices for students with disabilities, and parent trainings for children with disabilities.
ACKNOWLEDGMENTS

I would like to thank my advisor, Dr. Stage, for all of your hard work and patience throughout this process. Thank you for always having your door open to help out and taking the time to provide constructive feedback for my many drafts. I would like to thank my other Master’s Thesis committee members, Dr. Norwalk and Dr. Kim, for all of your hard work and time put into this process. In addition, I would like to thank Evadine Codd for her work collecting data for the reliability check.

Lastly, I would like to thank my family: my mom, Robin; my dad, Mike; and my sister, Becca. Your love and support have been an integral part of my achievements thus far. Thank you for always listening to me when I need it and celebrating with me when I reach a milestone. I know I can always depend on all of you even though we are miles away.
# TABLE OF CONTENTS

LIST OF TABLES ........................................................................................................... v
LIST OF FIGURES ........................................................................................................... vi
INTRODUCTION ............................................................................................................. 1

METHOD ........................................................................................................................ 7
  Sample ......................................................................................................................... 7
  Literature Search ...................................................................................................... 7
  Inclusion Criteria ....................................................................................................... 8
  Exclusion Criteria ...................................................................................................... 8
  Analytic Sample ....................................................................................................... 9

Data Collection and Measures .................................................................................. 10
  Independent Variables ............................................................................................. 10
  Outcome Variables .................................................................................................. 11
  Interrater Reliability ................................................................................................ 11
  Analytic Method ....................................................................................................... 11

RESULTS ....................................................................................................................... 13
  Sample Size ............................................................................................................... 14
  Special Education ..................................................................................................... 14
  Race ........................................................................................................................... 15
  Ethnicity ................................................................................................................... 15
  Free and/or Reduced Lunch ....................................................................................... 16

DISCUSSION .................................................................................................................. 16

REFERENCES .............................................................................................................. 22

TABLES .......................................................................................................................... 27
FIGURES ......................................................................................................................... 34
LIST OF FIGURES

Figure 1. Final estimation of fixed effects for each model of reading growth over time...........34
INTRODUCTION

According to the 2015 National Assessment of Educational Progress, 12% of students identified with a disability scored at or above proficient level on reading in Grade 4, compared to 40% of students without an identified disability in 2015 (U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2015). A similar trend exists in math testing for students with disabilities in Grade 4; 16% of students identified as students with disabilities scored at or above proficient level, but 44% of students without disabilities scored at or above proficient level (U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, 2015). This information is important because it shows that students with disabilities are performing below peers without disabilities; however, the data is limited in that it does not provide information as to how far below students with disabilities are performing. The data also does not indicate the level of improvement or growth each year of students with disabilities compared to students without disabilities. With the present gap in reading and math, researchers have attempted to observe and explain growth in reading and math for students with disabilities over time. Therefore, researchers hypothesized different theories of achievement gaps to explain the trends for children in special education compared to general education students. The purpose of this meta-analysis was to synthesize and analyze previous literature on reading and math growth over time for children in special education.

Researchers have used three theories to explain trends of reading growth for students with disabilities which include the developmental lag hypothesis (Fletcher, 1981), the developmental deficit hypothesis (Cromer, 1970), and the Matthew effect (Stanovich, 1986). The developmental lag hypothesis states that children who are initially behind in reading will eventually catch up given the appropriate level of instruction (Fletcher, 1981). The
developmental deficit hypothesis states that children who are initially behind in reading have the same rate of growth as regular achieving students but the gap between these groups remains stable over time (Cromer, 1970). The Matthew effect (i.e., the rich get richer; the poor get poorer) states that children who are initially poor in reading will have a slower rate of growth than regular achieving children (Stanovich, 1986).

Different researchers have found evidence for each of the different theories. For example, some researchers found evidence for the developmental lag hypothesis. Stanovich, Nathan, and Rossi (1986) found support for the developmental lag hypothesis for less skilled readers, or readers who are not disabled. However, the researchers in this study used a matched-reading level design to compare groups of readers at different grade levels. Thus, the study was not longitudinal and did not specifically factor in individual differences for changes in growth over time. McKinney and Feagans (1984) used longitudinal methods to test reading growth for children. They found evidence for the lag hypothesis using multivariate analyses of variance which measures the change in the mean average over data points by the within data point error variance. However, hierarchical linear modeling (HLM) may be more appropriate for an investigation of reading growth over time because HLM takes into account the linear change of growth across time points (Raudenbash, & Bryk, 2002).

Some researchers found evidence for the Matthew effect of reading growth. For example, Landman (2014) found evidence for a Matthew effect in reading when comparing students with and without Individual Education Plans (IEPs). Specifically, students with an IEP showed less growth in reading than students without an IEP.

Other researchers found evidence for the developmental deficit hypothesis. St. Clair, Durkin, Conti-Ramsden, and Pickles (2010) studied achievement trajectories for children from
age 7 to age 16 with Specific Language Impairments (SLI). The researchers found evidence for the deficit model of reading development, and they found no evidence for children with SLI catching up to their typically developing peers. St. Clair et al. (2010) found evidence for nonlinear growth trajectories; for example, children with SLI experienced more rapid growth earlier in their development and then slowed down in adolescence. Francis, Shaywitz, Stuebing, Shaywitz, and Fletcher (1996) studied a sample of reading impaired children, not reading impaired children, and low achieving children in Grades 1-9. The researchers found evidence supporting the developmental deficit theory for children with reading disabilities. They also found that students with reading disabilities did not develop the same reading skills as students without reading disabilities over time. Bussing and colleagues (2012) studied longitudinal trajectories of children grouped by ADHD status or by special education status. The researchers found support for the developmental deficit hypothesis for reading and mathematics achievement in that the children in the ADHD groups initially had lower scores but similar rates of growth as the nonspecial education group. Furthermore, Schulte, Stevens, Elliott, Tindal, and Nese (2016) studied achievement gaps in reading over a five-year period for 99,919 students in North Carolina, 14.4% of whom had disabilities. The researchers included 8 of the 13 different disability categories in their analyses: autism, emotional disturbance, hearing impairment, intellectual disability, learning disability in reading, learning disability in math or writing, other health impairment, and speech-language impairment. They found evidence for the developmental deficit theory for students with disabilities. Overall, the differences in reading scores between students with disabilities and general education students remained stable over time. The researchers did not find evidence of a Matthew effect or the developmental lag hypothesis. Additionally, Catts, Bridges, Little, and Tomblin (2008) studied children with
language impairments and children with typical language development. The researchers used latent growth curve analysis to examine the reading growth of children with language impairments. They found that children with language impairments initially performed below children with typical language development and that growth rates for children with language impairments did not differ from growth rates for typically developing children. The researchers also found higher initial growth for both groups of children that plateaued between the late elementary grades and the middle school/high school grades. Wei, Blackorby, and Schiller (2011) used hierarchical linear modeling to study growth in reading for students in 11 different disability categories (excluding blindness and deafness). These researchers also found evidence for the developmental deficit hypothesis when comparing students with disabilities to students with learning disabilities. Furthermore, Mattison, Hooper, and Glassberg (2002) used odds ratios to calculate differences in reading and intelligence scores for students with learning disabilities and behavioral disorders. Odds ratios were used by comparing the odds of learning disabilities at a follow-up point between students with and without learning disabilities. They found evidence for the deficit hypothesis for reading scores. Several other researchers (Swanson, Sáez, & Gerber, 2006; Wei, Yu, & Shaver, 2014) also found evidence for the developmental deficit hypothesis.

In addition to these theories of reading growth for students in special education, there are other student differences that may contribute to reading growth over time. For example, McCoach, O’Connell, Reis, and Levitt (2006) used growth modeling to investigate the differences in initial status and reading growth of students by gender, socioeconomic status, and race for kindergarten and first grade students. These researchers used piecewise growth curve models to investigate reading growth over time. For the kindergarten analysis, the researchers
found that students in families with lower socioeconomic status began kindergarten with lower reading scores than students in families with higher socioeconomic status. Additionally, McCoach et al. (2006) found that girls scored higher than boys on reading measures, Hispanic students scored lower on reading measures, and African American students did not have significantly different reading scores at the beginning of kindergarten. Growth analyses in first grade indicated that students from a high socioeconomic background had steeper slopes than students from lower socioeconomic backgrounds, and African American students had lower slopes than other students in the sample. These findings indicate that students from high socioeconomic backgrounds showed increased rates of reading skill than other students, and students who are African American showed a slower rate than other students in the sample during kindergarten. A number of explanations could be used to explain these differences in reading growth rates including exposure to reading material outside the school setting. These findings show individual differences in reading growth; therefore, it is important to consider these variables for analysis in the current study.

Previous researchers investigated mathematics trends for students with disabilities. Stevens, Schulte, Elliott, Nese, and Tindal (2015) measured math performance of students in Grades 3 to 7 using North Carolina End of Grade Assessments. The researchers found that students with disabilities initially scored lower than their peers on mathematics testing. Results revealed similar rates of growth over time for students with disabilities with the exception of students with mild intellectual disabilities; students with mild intellectual disabilities showed higher rates of growth over time than other groups included in the analysis (Stevens, et al., 2015).
Studies included in the current meta-analysis are longitudinal and used growth modeling for the analyses because hierarchical linear modeling takes into account the linear change of student achievement over time (Raudenbash, & Bryk, 2002). The final sample of studies (n = 7) which included math outcomes was too small for analysis; therefore, reading outcomes were further investigated and analyzed. Of the studies included in this meta-analysis, no studies provided evidence for the developmental lag hypothesis, four studies provided evidence for the developmental deficit hypothesis (Compton, Fuchs, Fuchs, Lambert, & Hamlett, 2012; Fleming, Cook, and Stone, 2002; Francis, et al., 1996; Schulte, et al., 2016) and two studies provided evidence for the Matthew Effect (Judge, & Bell, 2010; Landman, 2014) for reading growth over time. Four studies (Ewing-Cobbs, et al., 2004; Lane, Little, Menzies, Lambert, & Wehby, 2010; Lieu, Tye-Murray, & Qiang, 2012; Wei, Christiano, Yu, Wagner, & Spiker, 2015) did not use a comparison group of students without disabilities; thus, inferences about the most supported theory of reading growth could not be made.

Given the sometimes contradictory results of the previous studies reviewed, quantitative synthesis of these studies might yield a more concise answer to the question regarding reading growth for students with disabilities. Therefore, the goals of the present study were to synthesize and analyze research on student growth in special education and identify the best-supported theory of reading growth for students in special education. The research questions for the current study were: Which theory of academic reading gaps (i.e., the developmental lag hypothesis, the developmental deficit hypothesis, or the Matthew Effect) is best supported when studies are compared on the same metric? This question is important to add to our current understanding of how students in special education increase in their reading ability over time. A second research question was: What are the differences in student reading growth over time between students
found eligible for special education versus students not eligible for special education? These questions were answered by collecting studies that investigated reading and math growth over time for students in special education using growth modeling for analysis. After collecting the studies and recoding the data, they were analyzed using a two-level hierarchical linear growth curve model that analyzed the growth of academic skills of students in special education over time. Based on the literature discussed above, it was hypothesized that the most supported theory of academic growth in reading would be the developmental deficit hypothesis (i.e., students in special education would initially perform below same-aged peers, and growth rates would be similar between groups).

**METHOD**

**Sample**

**Literature search.** The original search for published and unpublished studies was conducted through EbscoHost (2016) databases including Academic Search Complete, ERIC, Medline, PsycArticles, and PsycInfo. The original search was restricted to the years 1980 to the present. Combinations of the following phrases were used as search terms: growth curves, Individual Education Plans (IEPs), students with disabilities, achievement, test scores, academic outcomes, children with disabilities, growth trajectories, academic achievement over time, longitudinal, Matthew’s effect, learning disabilities, ADHD, emotional disturbance, other health impairments, intellectual disabilities, autism, deaf-blindness, deafness, hearing impaired, multiple disabilities, orthopedic impairment, speech or language impairment, traumatic brain injury, and visual impairment. A total of 105 online searches were conducted through the Ebscohost databases described above with these combinations of words. The studies revealed by each search were screened for initial relevance through the title and abstract. For studies that met
the inclusion criteria discussed below from titles and abstracts, full text articles were retrieved from databases or interlibrary loan resources. Finally, reference lists of the studies that met criteria were scanned for other relevant studies.

**Inclusion criteria.** During these literature searches, each study title and abstract was reviewed to ensure that the study included students in schools in the United States of America, students with disabilities who qualified for special education services in elementary, middle, or high school and used growth modeling for the analysis. In addition, selection criteria included that the study used reading or math scores as an outcome measure and included at least three time points for reading or math scores (the time points needed to be across multiple years, not within a year). Finally, to control for potential confounding variables, studies were included if they were descriptive or observational as opposed to experimental (e.g., testing an intervention).

**Exclusion criteria.** After studies were reviewed for relevance and inclusion, exclusion criteria were applied through a more thorough reading of the research studies. Studies were excluded if the researchers did not separate the students with disabilities from students without disabilities in the analysis or the write-up of the study did not provide enough information about the estimates from growth modeling for the purposes of data collection \((n = 6)\). Studies measuring oral reading fluency to measure growth within a year (rather than across years) were excluded from analysis \((n = 6)\). A number of studies used the same data set for analysis; in these cases, only one study using the data set was included to ensure independence of observations across studies \((n = 11)\). If multiple studies used the same data set, the study was given priority to be used for analysis if it was published in a journal as opposed to a dissertation. Studies analyzing math growth over time were initially included in the data files; however, there were not enough
studies which reported math growth over time to include these results in the final analysis \((n = 7)\).

**Analytic sample.** The initial search revealed 214 published and unpublished studies. After inclusion and exclusion criteria were applied, 11 studies were left for analysis. Table 1 shows a summary of the studies included in the analysis. Studies had different numbers of time points across years; three studies had three time points, three studies had four time points, two studies had five time points, one study had six time points, one study had seven time points, and one study had nine time points. Studies included used reading comprehension and word reading as reading outcome measures. Within each study there were unique participant samples; the studies were divided into 34 unique samples of participants. Participant samples were determined from the individual studies based on how the researchers separated participants for their growth analysis. For example, some studies included a special education participant sample and a general education participant sample. If this were the case for a study, the study would include two samples by which information would be separated in the data sheets. Of the 34 participant samples, 8 groups were not special education samples and served as comparison samples and 26 samples were special education samples. The special education samples were not always identified through the school, so a separate variable was used to identify the samples of students who had an IEP \((n = 17)\). Information on types of disabilities was also included in the data collection: eight samples included students with learning disabilities, one sample included students with a speech/language impairment, five samples included students with autism, three samples included students with an emotional disturbance, two samples included students with a hearing impairment, one sample included students with an intellectual disability, one sample included students with a traumatic brain injury, and four samples included students with an other
health impairment (attention-deficit/hyperactivity disorder). Data were collected on race, ethnicity, and gender. Of the 64% of studies which reported race, ethnicity, and gender, 46% of the participants in the studies were White, 28% were Black, 17% were Hispanic, and 50% were male participants.

**Data Collection and Measures**

After the original search described above, information was coded into Excel spreadsheets. This information included study title, study authors, year of publication, data set, number of time points, outcome measures used, analytic sample size, study population, disability type, ethnicity and race, grade level, and study design. The studies were assigned a study number, and participant samples used for analysis within each study were assigned a unique sample number. More information on data collection and operational definitions of variables is described below.

**Independent variables.** Independent variables were coded by individual study group as 0 (no) or 1 (yes). The variables included elementary school, middle or high school, has IEP (i.e., researchers used school information to define students in special education), and qualified for a special education category (i.e., researchers defined students with disabilities through psychoeducational testing, or researchers used school information to determine students who were in special education). Race, ethnicity, gender, and free and/or reduced lunch status were coded based on the proportion of the study sample for each group. Specific disability categories used for each study are included in Table 1. The independent variables were included in a Level 2 file to allow for hierarchical linear modeling (further description is provided below in the design section).
**Outcome variables.** Reading and math outcome score estimates were collected and included in a Level 1 file. However, only reading scores were included in the final analysis due to the lack of studies reporting math growth over time. Study outcome measures were reading comprehension, word reading, or a reading cluster score (i.e., a combination score of word reading and reading comprehension). The studies included in the analysis used growth modeling; therefore, intercepts and slopes from the original studies were reported in tables. For the purposes of the current analysis, the primary researcher calculated estimates for the number of time points included in each study from the linear intercepts and slopes unless the estimates from growth modeling were already presented in the article. In that case, the provided estimates would have been collected for use in the analysis. If only intercepts and slopes were provided, the researcher added or subtracted the slope to the intercept depending on the time point for which the intercept was centered. The outcome measures were reported in different metrics across the studies. For example, some studies used growth scores, some studies used standard scores with a mean of 100 and standard deviation of 15, and some studies used T-scores with a mean of 50 and a standard deviation of 10. Measures were transformed to T-scores with a mean of 50 and standard deviation of 10 in order to make comparisons across studies.

**Interrater reliability.** Interrater reliability was calculated for the 11 studies used in the analysis. Specifically, a first-year graduate student was provided with the directions to collect data from the research studies for the Level 1 and Level 2 files. Results of separate intraclass correlation analyses revealed reliability coefficients of .99 for the Level 1 data file of reading scores. Interrater reliability for variables included in the Level 2 file ranged from 0.66 to 1.00.

**Analytic Method.** The current study used a two-level hierarchical linear modeling (HLM) procedure (Raudenbush & Bryk, 2002) using HLM-6. The first level included the
reading T-scores with the intercept centered at the first time point. Four time points were used for analysis; therefore, estimates were required for three studies which only included three time points. Empirical Bayes Estimates were used to account for missing data in this Level 1 data file (Raudenbush, & Bryk, 2002). The null model statistical equation was $Y = B_0 + B_1 \times \text{Time} + R$, for which $B_0$ is the reading score intercept, $B_1$ is the slope, and $R$ is the error term. In the equation, time is a linear constant with fixed data points (i.e., 0, 1, 2, and 3). Results of this model indicated the fixed intercept and slope for all studies included in the analysis. A significance test indicated if the T-score intercept was different from 0 and if the T-score slope (rate of change between time points) was different from 0. In addition, the final estimation of variance components tested the amount of random variability in the null model.

The Level 2 data file included the following independent variables: sample size, special education eligibility, race, ethnicity, and free and reduced lunch status. Each variable was analyzed in a separate model due to the small sample size. A model of sample size was included in the analysis to consider the weighting of each study by the relative proportion of students as a moderating effect. The special education variable was dichotomously coded 0 (i.e., students who did not qualify for special education services) or 1 (students who did qualify for special education services). Because differences in intercept and slopes for race and ethnicity have been found in previous studies, information about race and ethnicity were collected for analysis. To investigate the effect of race and ethnicity, a technique used previously by Fabiano, et al. (2008) was employed in the present study by dichotomizing variables based on census data. Specifically, studies with a higher proportion of Black students than the U.S. census data (i.e., 15%) were coded as 1, and if there was a lower proportion of students who were Black than census data, the study was coded with a 0 (U.S. Census, 2010). To analyze the effect of ethnicity,
if there was a higher proportion of students who were Hispanic than 10% of the sample, the study was coded as 1. To analyze the effect of free and reduced lunch status, if there was a higher proportion of students receiving free and reduced lunch than 37%, the study was coded as 1. Due to the low sample of participants identified as Hispanic or as qualifying for free and reduced lunch in the current sample of studies, lower percentages than census data were used but included in the model to represent these demographic characteristics in relationship to student growth in academics based on their relative distribution for the student sample characteristics. Results of the intercepts and slopes of these models are presented below.

**RESULTS**

To allow for comparison across studies, data were transformed into T-scores so the metric of analysis was comparable; thus, the intercept and slope coefficients presented below will be in the form of T-scores. Table 2 shows means and standard deviations for the independent variables prior to the HLM analysis.

First, the researchers ran a null model to analyze whether there were significant differences in intercept from 0 and differences in change over time from 0. Results from the null model indicated that the fixed initial intercept estimate of reading outcomes was 45.77 ($SE = 2.36$). The fixed slope (i.e., rate of change between time points presented as a T-score) was 1.11 ($SE = .60$) indicating little reading growth over time relative to the normative sample.

Variability was also tested within the initial model to understand the amount of variance unaccounted for that could be tested with independent variables. The initial model revealed significant random variability for the intercept ($p < .001$), indicating a significant amount of variability that could be further analyzed using independent variables that may better explain the variation at the initial intercept. The initial model did not reveal significant variability for slope
(\(p = .095\)). Testing of the random effects variability within the model revealed more variability in intercept (\(p < .001\)) but not slope (\(p > .500\)) across studies than accounted for in the fixed effect model. This suggests that the fixed effect was the best estimate of the slope because there was no additional random variability to account for in the study of growth curves. See Table 3 for estimations of variance components.

**Sample Size**

The following models were used to test the effect of sample size:

- **Level-1 Model:** \( Y = B_0 + B_1(\text{TIME}) + R \)
- **Level-2 Model:** \( B_0 = G_{00} + G_{01}(\text{SAMPLESIZE}) + U_0 \)
  \( B_1 = G_{10} + G_{11}(\text{SAMPLESIZE}) + U_1 \)

The level 1 model tested the effect of time (\(\text{TIME}\)) for all samples. The level 2 model tested the effect of sample size (\(\text{SAMPLESIZE}\)) on the intercept (\(B_0\)) and slope (\(B_1\)). Results of the models are shown in Tables 4 and 5. The model for sample size revealed no significant differences between studies with small and large sample sizes, suggesting that sample size of the studies used did not significantly influence the effects within the current sample.

**Special Education**

The following models were used to test the effect of students with special education eligibility:

- **Level-1 Model:** \( Y = B_0 + B_1(\text{TIMER}) + R \)
- **Level-2 Model:** \( B_0 = G_{00} + G_{01}(\text{SPED}) + U_0 \)
  \( B_1 = G_{10} + G_{11}(\text{SPED}) + U_1 \)

The level 1 model tested the effect of time (\(\text{TIME}\)) for all samples. The level 2 model tested the effect of students with special education eligibility (\(\text{SPED}\)) on the intercept (\(B_0\)) and
slope (B1). Results of the models are shown in Tables 4 and 5. The model testing the effect of students eligible for special education revealed no significant differences between a fixed intercept for the special education sample and a fixed intercept for the comparison sample. No significant differences were found between the fixed slope for the special education sample and the fixed slope for the comparison sample.

**Race**

The following models were used to test the effect of participant samples with Black students greater than census data:

**Level-1 Model:** \[ Y = B0 + B1 \text{(TIMER)} + R \]

**Level-2 Model:** \[ B0 = G00 + G01 \text{(RACE)} + U0 \]

\[ B1 = G10 + G11 \text{(RACE)} + U1 \]

The level 1 model tested the effect of time (TIME) for all samples. The level 2 model tested the effect of samples with more Black students (RACE) on the intercept (B0) and slope (B1). Results of the models are shown in Tables 4 and 5. The model testing the effect of participant samples with Black students greater than census data revealed no significant differences between a fixed intercept and slope for the samples with more Black students and a fixed intercept and slope for the comparison sample.

**Ethnicity**

The following models were used to test the effect of samples with more Hispanic students than 10%:

**Level-1 Model:** \[ Y = B0 + B1 \text{(TIMER)} + R \]

**Level-2 Model:** \[ B0 = G00 + G01 \text{(ETHNICITY)} + U0 \]

\[ B1 = G10 + G11 \text{(ETHNICITY)} + U1 \]
The level 1 model tested the effect of time (TIME) for all samples. The level 2 model tested the effect of samples with more Hispanic students (ETHNICITY) on the intercept (B0) and slope (B1). Results of the models are shown in Tables 4 and 5. The fixed intercept for samples with more Hispanic students than 10% was not significantly different than the overall fixed intercept. The fixed slope for samples with more Hispanic students was not significantly different than the total fixed study estimate.

**Free and/or Reduced Lunch Status**

The following models were used to test the effect samples with more free and reduced lunch status students than 37%:

- Level-1 Model: \( Y = B_0 + B_1 \times \text{TIME} + R \)
- Level-2 Model: \( B_0 = G_{00} + G_{01} \times \text{FRL} + U_0 \)
  \[
  B_1 = G_{10} + G_{11} \times \text{FRL} + U_1
  \]

The level 1 model tested the effect of time (TIME) for all samples. The level 2 model tested the effect of samples with more free and reduced lunch status students (FRL) on the intercept (B0) and slope (B1). Results of the models are shown in Tables 4 and 5. The model testing the effect of free and reduced lunch status revealed no significant differences between a fixed intercept and slope for the samples with more free and reduced lunch status students and a fixed intercept and slope for the comparison sample.

**DISCUSSION**

The purpose of this meta-analysis was to better understand the synthesized data of reading growth for students in special education and to determine the theory that best represents this growth for special education students compared to typically developing students: the developmental lag hypothesis, the developmental deficit hypothesis, or the Matthew Effect. The
developmental lag hypothesis states that students who are initially behind in reading will eventually catch up to their peers when they receive appropriate reading instruction (Fletcher, 1981). The developmental deficit hypothesis states that students who are initially behind in reading will have the same rate of reading growth as their peers but will not increase at a rate to catch their peers (Cromer, 1970). The Matthew Effect states that students who are initially behind in reading will have a slower rate of reading growth than their peers and will over time fall further and further behind (Stanovich, 1986). It was hypothesized that the most supported theory of reading growth over time would be the developmental deficit hypothesis.

Past researchers in the field have not found evidence that students in special education are catching up to their general education peers (Compton, et al., 2012; Fleming, et al., 2002; Francis, et al., 1996; Schulte, et al., 2016). On the contrary, researchers found evidence that students in special education tend to have similar rates of reading growth as their general education peers; however, special education students tend to have lower initial reading scores than their general education peers (Compton, et al., 2012; Fleming, et al., 2002; Francis, et al., 1996; Schulte, et al., 2016). The current results do not support significant differences in initial reading status or growth for students in special education compared to students in general education. However, of the 11 studies analyzed for this meta-analysis, only 4 of the studies supported the developmental deficit hypothesis (Compton, et al., 2012; Fleming, et al., 2002; Francis, et al., 1996; Schulte, et al., 2016), 2 studies supported the Matthew effect, and no studies supported the developmental lag hypothesis (Judge, & Bell, 2010; Landman, 2014). Four studies used in the analysis did not use a comparison group; thus, conclusions could not be drawn as to which theory of reading growth was most supported. Explanations for these results will be discussed further below.
Results from the null model revealed variability for initial intercept but not slope. Specifically, there were differences in the intercept for the studies included in the analysis, but there were not differences in change over time for the studies included in the analysis. However, this model indicated random variation, meaning that the model showed a significant level of variability about the intercept that could ostensibly be accounted for by other independent variables. The small sample did not allow the current researchers to include some variables that may explain more of the variation (i.e., specific disability category). In addition, reading scores within the analyzed studies were presented using different metrics; thus, the scores were transformed into T-scores with a mean of 50 and standard deviation of 10 to allow for comparison across studies. Thus, the finding of little growth over time for the general sample indicates that there was little reading growth relative to others.

Furthermore, results revealed no significant differences in initial level of reading score or rate of change over time for models of sample size, special education, eligibility, relative sample proportion by race, relative sample proportion by ethnicity, or the relative sample proportion by free and reduced lunch status. The lack of significance may be due to the lack of power in the analysis (i.e., only 11 studies were included in the final analysis of 10 linear growth curves). Accurate conclusions cannot be drawn for which of these theories best explains reading growth for students in special education. However, the results are still interesting and useful for descriptive purposes.

Although not significant, one interesting finding from the current results is that the special education sample revealed smaller increases over time (e.g., Intercept T-score = 49.76 and final T-score = 50.93) than the average comparison group and other comparative samples (see Figure 1), suggesting a possible trend towards the Matthew effect for special education.
students. The special education sample’s reading scores remained consistently flat in the current analysis meaning there was little improvement over time. Thus, special education students were not changing in their reading growth over time relative to their general education peers; however, one role of special education is to promote academic growth for students who need specially designed instruction. Furthermore, in an effort to improve outcomes for children in special education, the Office of Special Education and Rehabilitative Services (2014) requires results-driven accountability, but current results showed that the gap in reading was widening for students in special education. One explanation for this result could be due to the difference between empirically based interventions and common practice for special education. Weisz, Donenberg, Han, and Kauneckis (1995) studied the differences in effects for psychotherapy across experimental studies and clinical practice. The researchers attributed the better effects of psychotherapy in experimental studies to the use of more specific, focused, and structured treatment. The studies analyzed in the current report were not experimental; thus, it is possible that special education in practice led to less growth than expected.

Additionally, the Hispanic sample showed a relatively low intercept at the initial data point but grew an estimated 10 T-score points (Intercept T-score = 43 and final T-score = 53). This reflects a large effect of a standard deviation. Although this result is not significant, probably due to a lack of statistical power and random variability in the estimates, it is still interesting in that it may indicate that students in schools with higher proportions of Hispanic students are initially performing below students with lower proportions of Hispanic students. However, the Hispanic sample showed a higher rate of growth than other samples, indicating that the Hispanic sample caught up to peers for reading performance. This finding is similar to previous research findings indicating that students who receive English as a Second Language
(ESL) services have lower initial reading scores than comparison groups and periods of higher reading slopes (Al Otaiba et al., 2009).

Furthermore, the Black sample showed an average intercept at the initial data point and grew an estimated 4 T-score points (Intercept T-score = 49 and final T-score = 53). The free and reduced lunch sample also showed an average intercept at the initial data point and grew an estimated 5 T-score points (Intercept T-score = 48 and final T-score = 53). Although not significant, the Black sample and free and reduced lunch sample grew by approximately half of a standard deviation, a finding that indicates that these samples are making growth in reading over time.

Previous research found statistically significant differences in outcomes between the variables chosen for the current analysis (i.e., special education, race, ethnicity, and free and reduced lunch status) for intercept and/or slopes (e.g., Catts et al., 2008; Francis, et al., 1996; McCoach et al., 2006; Schulte, et al., 2016). The nonsignificant differences for intercept and slope between samples could indicate that the studies used in the meta-analysis had small effects for these differences in reading scores. Moreover, an examination of the studies used for analysis generally reveals that studies found small effect sizes for differences (e.g., Francis, et al., 1996; Schulte, et al., 2016). Thus, the findings from individual studies may not be robust enough to be generalized when summarized in a meta-analysis. In addition, the limited power due to a small sample size could also contribute to the lack of significant findings.

Although the present study provides descriptive information about special education students’ reading growth; the study has some limitations. First of all, the data used in the current study was limited to the data provided by previous researchers. Thus, some information was missing from analysis (e.g., free and reduced lunch status). More information about the samples
of the studies would have made for more interesting results and a better estimate of the trends for all participant samples. In addition, the small sample size did not allow for a powerful analysis. Future researchers should continue to use growth modeling techniques to determine which theory of reading growth for students in special education is most supported. A larger sample size also would have allowed for growth analysis of separate disability categories. The small sample size also indicates one important implication of the present study. There is currently a lack of research in the field employing growth modeling techniques to analyze special education students’ growth over time. The search for articles revealed that researchers were using the same data sets (e.g., the Early Childhood Longitudinal Study, Kindergarten Cohort) across studies to analyze the same samples of students. Unique samples should be used in the future so that generalizations to the population can be made.

Researchers in the past have investigated special education students’ growth over time, but there is limited research employing growth modeling techniques for analysis which account for individual differences between students. More specifically, researchers should use growth modeling to analyze students’ academic scores to determine their rates of growth over time compared to other samples and to account for the linear change of growth over time (Raudenbush, & Bryk, 2002). Researchers should also use comparison groups when analyzing academic growth for students in special education. A comparison group would allow researchers to understand special education student performance compared to students in general education. Thus, with a comparison group, researchers can make better conclusions as to whether or not special education students are catching up to their general education peers academically.
REFERENCES


*Judge, S., & Bell, S. M. (2010). Reading achievement trajectories for students with learning disabilities during the elementary school years. Reading and Writing Quarterly, 27, 153-178.


Stanovich, K. E. (1986). Matthew effects in reading: Some consequences of individual
Stanovich, K. E., Nathan, R. G., & Vala-Rossi, M. (1986). Developmental changes in the
cognitive correlates of reading ability and the developmental lag hypothesis. *Reading
in mathematics achievement of students with and without disabilities on a statewide
children at risk or not at risk for reading disabilities. *Journal of Educational Psychology,
98*, 247-264.
U.S. Department of Education, Institute of Education Sciences, National Center for Education
Statistics (2015). *Reading proficiency levels by disability status and grade, 2015* [Data
file].
United States Department of Education, Office of Special Education and Rehabilitative Services
(2014). RDA: Results driven accountability. Retrieved from
https://www2.ed.gov/about/offices/list/osers/osep/rda/index.html
psychotherapy outcomes in experiments versus clinics: Why the disparity? *Journal of
Abnormal Child Psychology, 23*, 83-106.
Wei, X., Blackorby, J., & Schiller, E. (2011). Growth in reading achievement of students with
disabilities, ages 7 to 17. *Council for Exceptional Children, 78*, 89-106.

<table>
<thead>
<tr>
<th>Authors and Date</th>
<th>Experimental Design</th>
<th>Independent Variables</th>
<th>Time Points</th>
<th>Outcome Measure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Compton, Fuchs, Fuchs, Lambert, &amp; Hamlett (2012)</td>
<td>Growth modeling using linear regression models</td>
<td>Students with learning disabilities and students without learning disabilities</td>
<td>4</td>
<td>Word Reading</td>
</tr>
<tr>
<td>Ewing-Cobbs, Barnes, Fletcher, Levin, Swank, &amp; Song (2004)</td>
<td>Individual growth curve analyses</td>
<td>Students with traumatic brain injury</td>
<td>6</td>
<td>Word Reading</td>
</tr>
<tr>
<td>Fleming, Cook, &amp; Stone (2002)</td>
<td>Two-level hierarchical linear model</td>
<td>Students with learning disabilities and students without learning disabilities</td>
<td>4</td>
<td>Combined Reading Score</td>
</tr>
<tr>
<td>Francis, Shaywitz, Stuebing, Shaywitz, &amp; Fletcher (1996)</td>
<td>Quadratic growth curve modeling</td>
<td>Students without an impairment in reading, students performing below peers in reading, and students performing below peers with an IQ discrepancy</td>
<td>9</td>
<td>Combined Reading Score</td>
</tr>
<tr>
<td>Judge, &amp; Bell (2010)</td>
<td>Two-level hierarchical linear model</td>
<td>Three groups of students with learning disabilities (early emerging, emerging, and late emerging) and students without learning disabilities</td>
<td>5</td>
<td>Combined Reading Score</td>
</tr>
<tr>
<td>Study</td>
<td>Methodology</td>
<td>Group Description</td>
<td>Score</td>
<td>Subject Area</td>
</tr>
<tr>
<td>-----------------------------</td>
<td>--------------------------------------</td>
<td>-----------------------------------------------------------------------------------</td>
<td>-------</td>
<td>---------------------</td>
</tr>
<tr>
<td>Landman (2014)</td>
<td>Hierarchical linear modeling</td>
<td>Students with an IEP and students without an IEP</td>
<td>4</td>
<td>Combined Reading Score</td>
</tr>
<tr>
<td>Lane, Little, Menzies, Lambert, &amp; Wehby (2010)</td>
<td>Hierarchical linear modeling</td>
<td>Students with emotional and behavioral disorders in rural and suburban areas</td>
<td>3</td>
<td>Combined Reading Score</td>
</tr>
<tr>
<td>Lieu., Tye-Murray, &amp; Qiang (2012)</td>
<td>Multilevel random regression modeling</td>
<td>Students with unilateral hearing loss</td>
<td>3</td>
<td>Combined Reading Score</td>
</tr>
<tr>
<td>Massetti, Lahey, Pelham, Loney, Ehrhardt, Lee, &amp; Kipp (2008)</td>
<td>Longitudinal linear regression in general estimating equations using an autoregressive correlation structure</td>
<td>Students without ADHD, students with ADHD Inattentive Type, students with ADHD Combined type, and students with ADHD Hyperactive-Impulsive type</td>
<td>7</td>
<td>Word Reading</td>
</tr>
<tr>
<td>Schulte, Stevens, Elliott, Tindal, &amp; Nese, (2016)</td>
<td>Two-level hierarchical linear models</td>
<td>Students without disabilities, students with autism, students with emotional disturbance, students with a hearing impairment, students with an intellectual disability, students with a learning disability in reading, students with a learning disability in math or writing, students with an other health impairment, and students with a speech-language impairment</td>
<td>5</td>
<td>Reading Comprehension</td>
</tr>
</tbody>
</table>
Table 1 (continued).

| Wei, Christiano, Yu, Wagner, & Spiker (2015) | Two-level hierarchical linear models | Higher-achieving students with autism, students with hyperlexia, students with hypercalculia, and students with autism who are lower-achieving | 3 | Word Reading |

*Note. The column labeled *Time Point* indicates the number of measurement points for a particular study. Typically, time points were one year apart. *Combined Reading Score* is a score made up of multiple domains of reading such as both word reading and reading comprehension.*
Table 2. Means and Standard Deviations for Independent Variables.

<table>
<thead>
<tr>
<th>Time Point</th>
<th>Comparison Mean (SD)</th>
<th>Special Education Mean (SD)</th>
<th>Black Mean (SD)</th>
<th>Hispanic Mean (SD)</th>
<th>Free and/or Reduced Lunch Mean (SD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>48.94 (12.46)</td>
<td>45.89 (7.05)</td>
<td>49.99 (4.51)</td>
<td>45.49 (6.62)</td>
<td>49.75 (7.42)</td>
</tr>
<tr>
<td>2</td>
<td>47.71 (14.41)</td>
<td>45.89 (7.95)</td>
<td>50.09 (4.60)</td>
<td>44.41 (6.94)</td>
<td>49.50 (7.07)</td>
</tr>
<tr>
<td>3</td>
<td>47.51 (18.23)</td>
<td>47.00 (9.22)</td>
<td>50.85 (3.84)</td>
<td>48.65 (6.74)</td>
<td>49.45 (6.43)</td>
</tr>
<tr>
<td>4</td>
<td>50.43 (15.62)</td>
<td>49.63 (8.73)</td>
<td>51.33 (4.85)</td>
<td>52.96 (4.16)</td>
<td>49.25 (6.01)</td>
</tr>
</tbody>
</table>

*Note.* Means are presented in T-scores.
### Table 3. Final Estimation of Variance Components.

<table>
<thead>
<tr>
<th></th>
<th>Null Model</th>
<th>Sample Size</th>
<th>Special Education</th>
<th>Black</th>
<th>Hispanic</th>
<th>Free and/or Reduced Lunch</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient (SD)</td>
<td>Coefficient (SD)</td>
<td>Coefficient (SD)</td>
<td>Coefficient (SD)</td>
<td>Coefficient (SD)</td>
<td>Coefficient (SD)</td>
</tr>
<tr>
<td><strong>Intercept</strong></td>
<td>49.23 (7.02) ***</td>
<td>49.51 (7.04) ***</td>
<td>51.66 (7.19) ***</td>
<td>38.55 (6.21) ***</td>
<td>56.05 (7.49) ***</td>
<td>56.11 (7.49) ***</td>
</tr>
<tr>
<td><strong>Slope</strong></td>
<td>1.03 (1.01)</td>
<td>1.07 (1.03)</td>
<td>1.15 (1.07)</td>
<td>1.19 (1.09)</td>
<td>1.08 (1.04)</td>
<td>0.93 (0.96)</td>
</tr>
</tbody>
</table>

*Note.* Intercepts and slopes are presented in T-scores. SD is the standard deviation. *p < .05, **p < .01, ***p < .001.*
Table 4. Final Estimation of Fixed Effects.

<table>
<thead>
<tr>
<th></th>
<th>Null Model</th>
<th>Sample Size</th>
<th>Special Education</th>
<th>Black</th>
<th>Hispanic</th>
<th>Free and/or Reduced Lunch</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient (SE)</td>
<td>Coefficient (SE)</td>
<td>Coefficient (SE)</td>
<td>Coefficient (SE)</td>
<td>Coefficient (SE)</td>
<td>Coefficient (SE)</td>
</tr>
<tr>
<td>Intercept</td>
<td>45.77 (2.36) ***</td>
<td>41.62 (2.74)</td>
<td>49.76 (5.54)</td>
<td>49.24 (4.31)</td>
<td>42.67 (8.35)</td>
<td>47.70 (6.25)</td>
</tr>
<tr>
<td>Slope</td>
<td>1.10 (0.60)</td>
<td>1.53 (0.70)</td>
<td>0.39 (1.69)</td>
<td>1.32 (1.24)</td>
<td>3.24 (1.86)</td>
<td>1.73 (1.32)</td>
</tr>
</tbody>
</table>

*Note.* Intercepts and slopes are presented in T-scores. SE is the standard error. *p < .05, **p < .01, ***p < .001.
Table 5. Final Estimations of Fixed Effects Coefficients.

<table>
<thead>
<tr>
<th></th>
<th>Null Model</th>
<th>Sample Size</th>
<th>Special Education</th>
<th>Black</th>
<th>Hispanic</th>
<th>Free and/or Reduced Lunch</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Coefficient (SE)</td>
<td>Coefficient (SE)</td>
<td>Coefficient (SE)</td>
<td>Coefficient (SE)</td>
<td>Coefficient (SE)</td>
<td>Coefficient (SE)</td>
</tr>
<tr>
<td>Level 1 Intercept</td>
<td>45.77 (2.36)***</td>
<td>45.76 (2.46)***</td>
<td>44.46 (2.92)***</td>
<td>41.66 (3.17)***</td>
<td>46.34 (2.63)***</td>
<td>45.57 (2.83)***</td>
</tr>
<tr>
<td>Level 2 Intercept</td>
<td>-4.53 (2.74)</td>
<td>5.30 (5.54)</td>
<td>7.58 (4.31)</td>
<td>-3.67 (8.35)</td>
<td>2.13 (6.25)</td>
<td></td>
</tr>
<tr>
<td>Level 1 Slope</td>
<td>1.10 (0.60)</td>
<td>0.72 (0.63)</td>
<td>1.08 (0.69)</td>
<td>0.88 (0.93)</td>
<td>0.77 (0.64)</td>
<td>0.84 (0.73)</td>
</tr>
<tr>
<td>Level 2 Slope</td>
<td>0.81 (0.70)</td>
<td>-0.69 (1.69)</td>
<td>0.45 (1.24)</td>
<td>2.47 (1.86)</td>
<td>0.89 (1.32)</td>
<td></td>
</tr>
</tbody>
</table>

Note. Intercepts and slopes are presented in T-scores. SE is the standard error. The Level 2 intercept and slope must be added to the Level 1 intercept and slope to determine the coefficient for each sample effect. * p < .05, ** p < .01, *** p < .001.
Figure 1. Final estimation of fixed effects for each model of reading growth over time presented in T-scores with a mean of 50 and standard deviation of 10. *Average* represents the null model.