Spring 5-18-2019

Examining the Consequential Validity of Using an English Language Proficiency Assessment for Reclassification Decisions

Anthony Sparks
Southern Methodist University, asparks@smu.edu

Examining the Consequential Validity of Using an English Language Proficiency Assessment for Reclassification Decisions

A dissertation presented to the graduate faculty of

Southern Methodist University

in partial fulfillment of the requirements for the degree of

Doctor of Philosophy in Education

Anthony E. Sparks

April 2019

Dissertation Approval

This dissertation submitted by Anthony E. Sparks has been read and approved by the following faculty members of the Annette Caldwell Simmons School of Education and Human Development at Southern Methodist University. The final copy has been examined by the Dissertation Committee and the signatures which appear here verify the fact that any necessary changes have been incorporated and that the dissertation is now given the final approval with reference to content, form and mechanical accuracy.

The dissertation is therefore accepted in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

Advisor and Committee Chair: Doris Luft Baker, Ph.D.

Signature:____Date:_________________

Committee Member: Leanne R. Ketterlin-Geller, Ph.D.

Signature:____Date:_________________

Committee Member: Meredith P. Richards, Ph.D.

Signature:____Date:_________________

Committee Member: Deni L. Basaraba, Ph.D.

Signature:____Date:_________________

Dedication

To my partner, Greg and our fur baby above, Griffon

Acknowledgements

I am extremely grateful to my advisor, Dr. Doris Luft Baker for her guidance and wisdom throughout my entire doctoral program. I also thank Dr. Leanne Ketterlin Geller for supporting me and pushing me to be a more reflective researcher. I thank Dr. Meredith Richards for her insight and contagious energy she brings to research and teaching. I thank Dr. Deni Basaraba for being a role model of professionalism during tough research situations. I thank all of the faculty at Southern Methodist University for supporting me throughout my doctoral experience. I thank my family, both near and far, for their continued support over the years.

Abstract

The purpose of the current study is to develop a better understanding of reclassification decisions for English learners. Previous research has primarily viewed reclassification strictly through a policy lens without providing a framework for understanding these high-stakes decisions. However, I adopt a validity framework for reclassification decisions, specifically as a consequence of testing. The framework specifies that assessment uses, such as reclassification decisions, have consequences that need to be understood. Furthermore, I conduct analyses of both achievement data and data related to graduation and college readiness to fully understand the impact of reclassification decisions. The Duke Center for Child and Family Policy provided student-level data from North Carolina with 42,393 students. Outcomes of interest include English language arts performance, mathematics performance, graduation, ACT performance, and Advanced Placement (AP) enrollment. I adopted a coarsened exact matching technique to establish comparable groups of students and a difference in differences approach to assess the effects of reclassification on achievement outcomes. For outcomes related to college readiness and graduation, I utilize coarsened exact matching and regression techniques. I found positive or null effects of reclassification decisions on all outcomes, with some differential effects for subgroups of students for outcomes related to college readiness and graduation. I also find English learners that never reach the criteria for reclassification are limited in their access to AP courses and perform lower than the state average on ACT subtests. Limitations and areas for future research are discussed.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

The Every Student Succeeds Act of 2015 (ESSA, 2015) defines an English learner as an individual whose first language is not English and whose proficiency in English prevents full participation in English-only settings. These settings are not limited to the classroom but also include social contexts outside of school. In 2017, English learners represented 9.5% of the total public-school population (National Center for Education Statistics, 2017, Table 204.20). However, the percentage of English learners was much larger in some states than others. For example, in California, 21% of public-school students were identified as English learners. In Texas, 16.8% of public-school students were classified as English learners. In addition, some states have seen a dramatic increase in their English learner population over time. For instance, the English learner population in North Carolina grew from 44,165 in 2000 to 102,090 in 2015, representing a 131% increase (National Center for Education Statistics, 2017, Table 204.20).

Statement of the Problem

A major concern is that the performance of English learners in public schools has consistently lagged behind the population of public-school students. English learners perform lower on the National Assessment of Educational Progress (NAEP). In 2017, eighth-grade English learners performed, on average, 43 scale score points lower in reading compared to non-English learners ($p < .001$, National Center for Education Statistics, 2017b). Similarly, English learners performed, on average, 40 scaled-score points lower than non-English learners in math ($p < .001$, National Center for Education Statistics, 2017b). In summary, English learners, on average, perform lower than their non-English learner peers in reading and mathematics.

However**,** disaggregation of assessment and graduation data rarely occurs for students previously classified as English learners but have been reclassified and are no longer considered [or carry the label of] English learners. Therefore, the knowledge base about the performance of English learners after reclassification is small (Carlson & Knowles, 2016). Reclassification refers to the change in status that occurs when English learners reach specific criteria to no longer need additional English language support (Cimpian, Thompson, & Makowski, 2017; Robinson-Cimpian & Thompson, 2016). In most states, English language proficiency assessment performance is the sole basis for reclassification decisions (Linquanti & Cook, 2015). Given that these English language proficiency assessments are utilized to make high stakes decisions such as reclassification, it is essential to evaluate the appropriateness of criteria for reclassification because of the consequential validity these decisions have on the academic future of English learners.

Purpose of the Study

The purpose of this study is to investigate the consequential validity of test-based reclassification decisions of English learners. According to The Standards for Educational and Psychological Testing, consequences refer to, "the outcomes, intended and unintended, of using tests in particular ways in certain contexts and with certain populations" (American Educational Research Association [AERA], American Psychological Association [APA], & National Council on Measurement in Education [NCME], 2014, p. 217). Previous literature around reclassification has primarily studied either students' academic (e.g., academic performance in the year following reclassification) or outcomes related to graduation and college readiness (e.g., ACTs, AP course enrollment) but not both (Carlson & Knowles, 2016). To fully understand the impact of reclassification decisions from English language proficiency assessments, I argue for analyses of academic performance outcomes and outcomes related to graduation and college readiness. I argue for analyses of both due to the potential long-term impact reclassification decisions could have on English learners.

On the one hand, reclassified students without the necessary English language proficiency may experience declines in academic performance and may struggle in mainstream classrooms (Cimpian, Thompson, & Makowski, 2017). On the other hand, students who are reclassified a year or two after they have reached the appropriate English language proficiency may also miss the opportunity to enroll in advanced classes, which prepare them for post-secondary education (Carson & Knowles, 2016). Given that reclassification decisions in most states are based solely on performance on English language proficiency assessments, reclassification can be viewed as a decision based on a test. Thus, the consequential validity of reclassification decisions should be investigated further to understand better the effects of these decisions on reclassified students and students who remain in English language supports.

To better comprehend the consequences of the decision to reclassify English learners, the current study will focus on examining the consequences of reclassification decisions based on a language proficiency assessment in North Carolina. The reasons to focus on North Carolina is because first, the number of English learners in North Carolina has substantially increased in over the past 12 years (National Center for Education Statistics, 2017, Table 204.20), and second, because of the lack of research on the consequences of reclassification decisions in that state (see Tables 2 and 3).

English Language Proficiency Assessments

The validity and reliability behind English language proficiency assessments have been thoroughly studied and reported by test developers (e.g., WIDA, n.d.). According to the Standards (2014) validity is, "the degree to which accumulated evidence and theory supports the proposed interpretation(s) of test scores for a given use of a test" (p. 225). Validity is not a property of the test, but the degree in which evidence supports the interpretations and uses of the test. For example, reclassifying English learners out of English language support based on their performance on an English language proficiency test is an example of a decision based on a test score. Validity does not refer to the English language proficiency test itself but to the uses of the test to make educational decisions such as the reclassification of English learners. In the following section, I provide an overview of validity with an emphasis on the argument-based approach to validation (Cronbach, 1988; Kane, 2013; Messick, 1989).

Theoretical Framework

Messick (1975) developed the unified view of validity which considers previously disjoint validity "types" (i.e., construct, criterion, content) as complementary to each other. Under Messick's unified view of validity, all of these "types" of validity are categorized under construct validity because they all contribute to the score's meaning. For example, a strong relation between performance on an English language proficiency assessment and an English reading comprehension measure adds to the overall understanding of how the English language proficiency assessment and the English reading comprehension measure assess the same construct similarly. Thereby, providing more context for the meaning of the score and the construct the test is proposed to measure.

Cronbach (1988) proposed a “devil’s advocate” mindset when evaluating the interpretations and uses of test scores. In particular, he proposed that evaluators of test should be open to newer evidence that emerges that could contradict the current evidence in support of a particular test use or interpretation. This view challenged previously held notions of validation that insisted a single source of evidence was sufficient for evaluating the validity of a test use or interpretation.

Building upon Cronbach's (1988) and Messick's (1989) work, Kane (1992) developed the argument-based approach to validation. The argument-based approach involves stating what is being claimed by the measure and then an evaluation of those claims by the test developer. This approach to validation starts from a score on an assessment to a particular use of a measure. Between the score and use, multiple claims and assumptions are made to arrive at a specific score use. The network of claims and assumptions make up the interpretation and use argument (IUA) of the argument-based approach to validation (Kane, 1996). The validity argument consists of the evaluation of the claims made in the IUA. Sources of evidence to support the claims made are collected and evaluated for the adequacy of the claim within a particular use of an assessment. A test use or interpretation is conditionally valid only after the IUA has undergone, "a critical appraisal of its coherence and plausibility of its inferences and assumptions" (Kane, 2013, p. 453). However, the argument-based approach to validation is malleable when new evidence arises that could alter the meaning of the test score.

An example could be a score on an English language proficiency assessment and the use of measuring a student's English language proficiency to position students on a spectrum of English language proficiency (i.e., English language proficiency standards). One claim could be that the English language proficiency assessment accurately measures a student's current level of English language proficiency as determined by English language proficiency standards. A source of evidence to support this claim could be professional judgments on whether the test aligns with the standards. The evaluation of the evidence would critique the appropriateness of these professional judgments. However, reclassification decisions come after the specific use of the test, as a consequence of testing. For example, a score on an English language proficiency assessment leads to the interpretation of a student's current level of English proficiency. The interpretation is the basis for reclassification decisions.

The decision to reclassify English learners is appropriate when the academic performance of those reclassified is not impacted differentially compared to English learners still receiving English language supports (Cimpian et al., 2017; Robinson, 2011). However, to fully understand the consequences of reclassification decisions, multiple sources of evidence to support these criteria for reclassification purposes are needed (Kane, 2013). In the next section, I provide a more in-depth discussion around consequential validity as it relates to reclassification decisions.

Consequential aspects of reclassification. Haertel (2013) highlights both direct (intended) and indirect (unintended) uses and consequences of testing. The direct consequences of testing are those the test developers anticipated during test development. Indirect consequences are those the test developers did not intend. For example, developers of a math assessment may intend for users (teachers) to make mathematics instructional decisions (e.g., pedagogical changes) based on results of the test, but may not intend or anticipate for users such as policy makers and administrators to make policy and programmatic decisions (e.g., implementation of new teaching programs) from the same test results.

Typically, test developers explicitly document the intended uses of their assessment. For example, test developers for WIDA's (n.d.b) English language proficiency assessment (Accessing Comprehension and Communication in English State-to-State [ACCESS]) specify five potential uses of their assessment, which include: (1) measuring English language proficiency, (2) making decisions about entry/exit into programs (e.g., English learner programs), (3) informing classroom instruction, (4) monitoring annual progress towards English language proficiency, and (5) staffing decisions. Under the argument-based approach to validation (Kane, 2013), all proposed uses for an assessment require a rigorous evaluation of evidence to support their proposed use.

Unless specified as a use of the assessment by the test developer, reclassification decisions are an unintended use of a language proficiency assessment (Haertel, 2013; Lane, 2014). An additional unintended use of an English language proficiency assessment may include evaluations of teachers and schools. For example, a teacher may be evaluated based upon the progress their students made towards English proficiency during the school year. However, the test developer may not have specified their instrument as an evaluative tool for teacher performance. Therefore, leading to an unintended use of their assessment.

Intended uses of assessment have consequences. For example, a consequence of using the WIDA ACCESS to make reclassification decisions is that English language supports are no longer provided for the student, which may impact their performance on subsequent outcomes. Furthermore, measuring English language proficiency may help teachers improve instructional practices for English learners. In either example, the test developer supported using their assessment in that manner.

Unintended uses of assessments also have consequences. If the test developer did not intend on their assessment to be used for reclassification decisions, then the same impact on subsequent outcomes may be expected. Furthermore, a consequence of using an English language proficiency assessment for teacher evaluations may lead to higher turnover among teachers in school with higher populations of English learners. Evaluations of both the intended and unintended consequences of testing are needed (Haertel, 2013; Lane, 2014).

Sources of evidence to evaluate the use of an assessment for reclassification decisions could include performance on student outcomes. If reclassified and this decision negatively impacts their academic performance, then evidence exists against the decision to reclassify. However, significant positive effects of reclassification on subsequent outcomes could also indicate an invalid use of the assessment at the established criteria as it could indicate English learners were receiving unnecessary services, potentially preventing reclassified English learners from accessing advanced coursework. Therefore, to better understand the effects of reclassification, it is crucial to analyze the consequences of the use of the test on both, the immediate and the distal academic outcomes. In the next section, I discuss the historical underpinning of English language proficiency assessments to situate the current research better.

History of English Language Proficiency Assessments

Federal policies regarding English learners were non-existent before 1968 (Wright, 2005). English learners at that time were merely immersed in the mainstream classroom and were expected to perform without English language supports (e.g., sink or swim). However, in 1968, the Elementary and Secondary Education Act (ESEA) was amended to add Title VII or the Bilingual Education Act (BEA; Gándara, 2015). The BEA was the first time the federal government acknowledged the unique challenges of English learners in reaching the same educational goals as native-English speakers. The law itself was vague and did not provide much direct guidance on how the educational needs of English learners were to be served (Wright, 2005). Furthermore, BEA lacked accountability measures and mandates for districts to establish bilingual programming (e.g., tracking English language proficiency; Bunch 2011).

In 1974, the Supreme Court case *Lau v. Nichols* ruled English-only opportunities for English learners without additional language support was in fact, not allowing English learners equal access to education (Wright, 2005). The court decision made programming (e.g., bilingual programs, English as a second language) that permitted access to public education a right for English learners, protected by the Civil Rights Act of 1964 (Bunch, 2011). However, the use of assessments of English language proficiency for purposes of classification and reclassification were still not federally mandated.

The No Child Left Behind Act (NCLB; 2002) eliminated Title VII and replaced it with Title III, "Language Instruction for Limited English Proficient and Immigrant Students". Title III required programs for English learners to focus on only two requirements: teaching English and teaching the state content standards. In addition to the elimination of Title VII and the authorization of Title III, Title I was also updated. Title I mandated the annual testing of all students in grades 3 through 8 and once in high school to better monitor student progress (NCLB, 2002). Furthermore, states' accountability measures included English learners as a subgroup in states' accountability measures.

Due to NCLB's requirement to properly report outcomes of state assessments by subgroups (e.g., English learners, students with disabilities), the need to accurately identify students who are ready to be reclassified also grew. In addition to content area assessments, NCLB also required each state to develop English language proficiency standards and English language proficiency tests to monitor English learners' progress towards becoming English proficient (NCLB, 2002). Currently, the basis of these standards and assessments are on the four domains of language development (i.e., reading, writing, speaking, listening). States receiving funding for Title III must submit to the federal government annual reports highlighting the progress of English learners in reaching English proficiency.

Research related to the effects of test-based decisions, such as reclassification, is relatively new. Therefore, our understanding of how assessment-based decisions affect students' performance is limited (Robinson, 2011). In the next section, I provide an overview of the English language proficiency assessment utilized in North Carolina, and how the state uses the assessment in making reclassification decisions.

English Language Proficiency Assessment in North Carolina

NCLB required states to develop English language development standards and English language proficiency tests to monitor progress towards English proficiency (NCLB, 2002). The following English proficiency test is only one of the tests utilized in the U.S.

WIDA ACCESS. ACCESS is comprised of four different subtests that represent the various domains of language development (i.e., speaking, writing, reading, listening) for students in grades K-12 (WIDA, 2016). The test can be administered with paper-pencil or on the computer. The test format in kindergarten is different from the other grades. The kindergarten test has two parts, expository and narrative, while the remaining grade levels consist of four parts based on the language development domains. In grades 1-12, tests are administered across multiple grades and consist of three tiers (e.g., A, B, & C) of increasing difficulty.

Administration of the test for grades 1-12 depends on the particular domain tested and the tier in which the student is placed (WIDA, 2016). The speaking test must be administered individually, whereas tests of writing, reading, and listening are group administered. The online administration of the test is similar to the paper-pencil, except adaptive. Students must take the listening and reading portion of the test before being placed in tiers for the writing and speaking sections. The online administration is not available for kindergarten.

The test administrator directly scores the writing portion of the exam in kindergarten and the speaking portion for grades 1-12 based on rubrics provided by WIDA (WIDA, 2016). The test administrator must have completed training to administer the test before scoring. For the online version, a third-party scoring company scores the speaking portion of the exam in addition to the online and paper-pencil administration of the other three domain tests.

Students receive a score ranging from one to six in each domain with students at the lower end representing those of lower performance (WIDA, 2016). Composite scores are generated across domains to provide an overall indicator of English language proficiency. The test is criterion-referenced which means scores on the assessment linked to levels on the WIDA English Language Proficiency Standards. States can specify what level of English proficiency is adequate for reclassification based on the score from ACCESS, either from the composite score or a combination of the composite score and subtest scores (Linquanti & Cook, 2015).

Claims are made about assessments for which evidence is collected to support these claims. On claim made by WIDA (CAL, 2017) is that items on their assessment work appropriately together. WIDA's technical manual reports reliability estimates for the overall composite score. Reliability refers to, "the degree to which test scores for a group of test takers are consistent over repeated applications of a measurement procedure and hence are inferred to be dependable and consistent for an individual test taker" (AERA et al., 2014, pp. 222-223). Estimates for the overall composite score internal consistency reliability range between .932 and .974 (CAL, 2017).

Another claim made by developers of the WIDA assessment is that test takers are appropriately classified to proficiency levels defined by WIDA's English language proficiency standards (CAL, 2017). WIDA provides accuracy of classification for an undisclosed criterion assessment at the overall cut scores to support this claim (e.g., one vs. two, two vs. three). The accuracy of classification, defined as the proportion of correctly classified students in relation to a comparable English language proficiency assessment, across all cut scores for each grade level range from .870 to .992. High accuracy in classification is critical to consider because it provides an estimate of how well the test is distinguishing between students at different thresholds of language proficiency. For example, if a state adopts a threshold of five on the WIDA ACCESS for a student to no longer receive English language supports, then evidence needs to be provided that the test can accurately distinguish between students who scored above, below, or at the threshold.

Reclassification decisions with WIDA ACCESS depend on the state. Table 1 highlights the 35 states and the District of Columbia as well as the reclassification criteria summarized by Linquanti and Cook (2015). Nineteen states and the District of Columbia require only performance on WIDA ACCESS for reclassification while 15 states require additional criteria. Furthermore, the reclassification threshold on the WIDA ACCESS varies between states. Table 1 also highlights the lack of consistency between the number of criteria and the proficiency score required. The Office of English Language Acquisition (2015) reports that the minimum scores for reclassification in states that utilize WIDA ACCESS for reclassification decisions range from 4.5 to 5.0. For example, North Carolina requires a composite score of 4.8 (North Carolina Testing Program, 2017). North Carolina provides no documentation for this reclassification threshold, which highlights the need for understanding the appropriateness of these criteria for reclassification decisions. To gain a better perspective of the research conducted on reclassification, I review the related research in the next chapter.

Table 1

Summary of Reclassification Criteria by WIDA Consortium Member

Member	Number of Criteria	Proficiency Score	Member	Number of Criteria	Proficiency Score
Alabama	2	4.8	Montana	2	5.0
Alaska	1	5.0	Nevada	1	5.0
Colorado	2	5.0	New Hampshire	1	5.0
Delaware	2	5.0	New Jersey	3	4.5
DC	1	5.0	New Mexico	1	5.0
Florida	4	5.0	North Carolina	1	4.8
Georgia	1	5.0	North Dakota	1	4.8
Hawaii	1	4.8	Oklahoma	1	5.0
Idaho	3	5.0	Pennsylvania	3	5.0
Illinois	1	5.0	Rhode Island	3	4.5
Indiana	2	5.0	South Carolina	1	5.0
Kentucky	1	5.0	South Dakota	1	4.7
Maine	1	6.0	Tennessee	1	5.0
Maryland	1	5.0	Utah	2	5.0
Massachusetts	3	5.0	Vermont	1	5.0
Michigan	3	5.0	Virginia	1	5.0
Minnesota	4	5.0	Wisconsin	4	6.0
Missouri	4	6.0	Wyoming	1	5.0

Note: DC is the District of Columbia.
Source: Linquanti & Cook (2015); WIDA (n.d.c.)

CHAPTER II: LITERATURE REVIEW

To select studies for review, I searched several electronic databases including ERIC, Academic Search Complete, and Google Scholar for peer-reviewed publications using the following keywords: *consequential validity, English language proficiency assessment, and reclassification*. However, the search results yielded no empirical studies. Furthermore, when the search was broadened to include keywords such as *English learner* in combination with *consequential validity*, no empirical research was found. Therefore, I decided to search for any studies that examined the effects of reclassification on English learners. I again searched several electronic databases including ERIC, Academic Search Complete, and Google Scholar for peer-reviewed publications using the following keywords: *English learner, English language learner, reclassification,* and *reassignment*. I also conducted "Snowball sampling" of these 22 publications to ensure a comprehensive search. Snowball sampling involves checking references of initially selected literature for empirical research to ensure inclusiveness of current literature. The initial search yielded 32 articles. Abstract analysis to provide content relevancy produced 25 studies for further review.

After identification and abstract analysis, I read studies to determine if they met inclusion criteria. The process for reporting systematic reviews and meta-analyses developed by Moher, Liberati, Tetzlaff, Atlman, and the PRISMA Group (2009) was adopted to increase the quality of the review. Examples of items to review included objectives, participant selection, and methods. In addition, studies had to (a) be empirical, quantitative, and published in peer-reviewed journals or from a reputable research agency (e.g., government reports), (b) focus on the relation between reclassification and student outcomes, (c) have been published between the years 2001 and 2017, and (d) have been conducted in the United States. The decision to limit research only to studies conducted in the United States was due to the diverse English learner population in the country. I selected studies that were published between 2001 and 2017 because of the substantial increase in the English learner population during these years, and the rise in awareness of the consequences of reclassifying English learners. Furthermore, 2002 was when NCLB (2002) was signed into law, which increased accountability measures for English learners in terms of monitoring and assessing English language proficiency.

Thus, the current chapter reviews research related to the effects of reclassification on English learners and analyzes the consequences of reclassification decisions within the study's context. Furthermore, I also focus on the methodologies utilized to study reclassification. The methods used to study reclassification can provide insight into how to examine reclassification in the context of North Carolina.

Results of Literature Search

In total, 8 of the original 32 studies (25%) met eligibility for inclusion in the current synthesis. Only seven studies analyzed the effects of reclassification on students' immediate academic outcomes (Ardasheva, Tretter, & Kinny, 2012; Cimpian et al., 2017; Hill et al., 2014; Kim & Herman, 2010; Kim & Herman, 2012; Robinson, 2011; Robinson-Cimpian & Thompson, 2016), while the eighth study analyzed the relation between reclassification on college readiness (e.g., ACT performance) and post-secondary enrollment (Carlson & Knowles, 2016).

In the current chapter, I review studies analyzing the effects of reclassification on English learner outcomes. In particular, I review studies in this chapter based on the strengths of the methodologies employed. Also, I consider the consequences of reclassification in each study to understand the consequential validity better. The studies are organized by outcome type, starting with achievement outcomes and then moving to the outcomes associated with graduation and college readiness. Furthermore, the studies within the section are organized by increasing methodological rigor. For example, studies that adopted multilevel modeling and multivariate regression are reviewed first followed by increasingly rigorous methods such as regression discontinuity. I chose this organizational scheme to highlight the methodological strengths and weaknesses in each study.

Studies of Achievement Outcomes

Previous research has studied achievement outcomes with various methods. In this section, I highlight the methods within their design type. I start with correlational designs and then move to quasi-experimental designs. Furthermore, I organize each sub-section by increasing methodological rigor.

Correlational Studies

Table 2 summarizes the studies of achievement outcomes reviewed, including English proficiency assessment utilized in the state and the number of reclassification criteria as specified by Linquanti and Cook (2015) or the study. Overall, the representation of known states is quite low with only three states studied. Furthermore, variability exists in the number of criteria represented, ranging from one to four criteria for reclassification decisions. Lastly, multiple methodologies have been employed to study the impacts of reclassification.

Hill et al. (2014) compared districts with different reclassification policies to better understand if districts with higher reclassification standards led to higher performance on the state English language arts assessment in California. Four cohorts of English learners consisting of grades two, four, seven, and eight were studied over six school years with sample sizes ranging from 120,205 to 192,991 English learners. Survey data were also collected from districts to gain a better perspective of each districts' reclassification standards. Data analyses with cohorts of students were performed utilizing multivariate regressions over time. Multivariate regression is similar to ordinary least squares regression except instead of just one dependent variable; there are multiple (Rencher & Christensen, 2012). Results indicate that districts with stricter reclassification policies (e.g., higher language proficiency thresholds) had slightly better reclassified English learner performance and on-grade progression compared to non-reclassified English learners. Furthermore, districts with stricter standards for achieving reclassification yielded more positive results for reclassified English learners compared to districts with lower reclassification standards. Therefore, the reclassification decisions of districts in California that adopt stricter criteria for reclassification have positive consequences on reclassified English learner performance due to increases in performance of reclassified English learners.

Table 2

Studies of Achievement Outcomes

Study	State	Number of Criteria	Assessment	Method
Ardasheva et al. (2012)	Kentucky	1 (Composite + Domain)	LAS/WIDA ACCESS	Hierarchical Linear Modeling
Cimpian et al. (2017)	NA	1	NA	Regression Discontinuity
Kim & Herman (2010, 2012)	NA	2	NA	Hierarchical Linear Modeling
Hill et al. (2014)	California	4	CELDT	Multivariate Regression
Robinson (2011)	California	4	CELDT	Regression Discontinuity, Instrumental Variable Estimation
Robinson-Cimpian & Thompson (2016)	California	4	CELDT	Regression Discontinuity, Instrumental Variable Estimation, Difference-in-Differences, Inverse Probability Weighting

Ardasheva et al. (2012) studied the impact of English proficiency level (i.e., English learner, former English learner, native English speakers) on reading and math performance of middle school students in Kentucky. The sample consisted of 558 English learners, 500 former English learners, and 17,470 native English speakers. Researchers conducted a two-level hierarchical model which included student-level variables such as age, gender, current English learner status, former English learner status, and SES. A hierarchical linear model is a type of multi-level model and is a modeling technique that accounts for the dependence between observations (e.g., times or subjects; Singer & Willet, 2003). Findings indicated former English learners outperformed English learners and native English speakers in reading ($p < .001$) and mathematics ($p < .001$) when controlling for student and school-level characteristics. These higher performances suggest reclassified English learners are not adversely affected by the reclassification process. A weakness of the study was the lack of school-level variables present, such as percent minority or other school demographic variables. The only school-level variable utilized in the model was the poverty rate. Overall, Ardasheva et al. (2012) findings suggest reclassification decisions in Kentucky have positive consequences due to the reclassification process not causing a decrease in the academic performance of reclassified English learners.

Kim and Herman (2010, 2012) researched the effects of reclassification on math and English language arts performance in an undisclosed state. Six years of longitudinal data from 45,006 English learners and non-English learners in grade three were utilized. The researchers adopted hierarchical linear modeling for analyses. Results from both studies indicated reclassified English learners in grades four, five, and six have growth rates higher than non-English learners in math and reading (as defined by the state math and reading assessments), implying reclassified English learners would eventually catch-up to their non-English learner peers. These studies, however, neglected to model individual growth parameters, which leaves a large portion of variability unexplained. Similar to the study in Kentucky (Ardasheva et al., 2012), reclassification decisions in the undisclosed state have positive consequences due to reclassification decisions not causing a decrease in academic performance among reclassified English learners. A weakness of the current studies and similarly with the study conducted by Ardasheva et al. (2012) was the lack of district-level variables in their analytic model such as the programs English learners were offered in their districts.

Quasi-Experimental Studies

Quasi-experiments are studies lacking random assignment of participants or units (Murnane & Willett, 2011). Unlike the studies mentioned previously, quasi-experimental designs utilize additional techniques to isolate the treatment effect, or in the case of reclassification, the impact of reclassification on academic outcomes. In a study conducted by Shadish, Clark, and Steiner (2008), the researchers conducted a study with random and non-random assignment of participants. The researchers found that with proper utilization of covariate information (e.g., age, gender), the quasi-experiments can approximate results from randomized experiments. Quasi-experiments can be useful when randomization is not feasible due to ethical or logistical constraints.

Cimpian et al. (2017) analyzed the effects of reclassification on English language arts and math performance in an undisclosed state. Longitudinal data of 65,243 English learners in grades three through eleven were analyzed over five years. The researchers adopted a regression discontinuity approach. Researchers found no significant differences between English learners below and above the threshold for reclassification on English language arts and math performance. A lack of significant differences between students above and below the reclassification threshold signal positive consequences of reclassification decisions. Compared to previously mentioned studies, the methodology in this study allowed researchers to establish comparable groups of students, which strengthened the validity of their results.

Robinson (2011) assessed the effects of reclassification on subsequent student English language arts performance in California. Data from a large urban district in California with a sample size of 22,827 English learners in grades three through eleven were utilized. Analyses consisted of a regression discontinuity with instrumental variable estimation. An instrumental variable is a variable unrelated to the dependent variable of interest, and whose relation between the independent variable is not confounded by a third, unexplained variable (Murnane & Willett, 2011). The instrumental variable used by the researcher was whether a student reached the previous year's cutoff for reclassification. In California, reclassification decisions are not solely based on the English language proficiency assessment but additional criteria such as teacher and parent input (Linquanti & Cook, 2015). The instrumental variable helped control for the selection bias introduced by the subjective input in reclassification decisions. Results indicated no significant differences in the performance of English learners and reclassified English learners on subsequent English language arts performance as measured by state subject tests in elementary and middle school. However, in high school, there was an adverse effect of being reclassified on subsequent English language arts performance. Unlike the study by Cimpian et al. (2017) that found no significant impact of reclassification on student performance, this study found adverse effects for high school students, which indicates the decision to reclassify students may not be appropriate for all grade levels. In summary, decisions to reclassify students appear to have positive consequences for English learners in elementary and middle school, but negative consequences for English learners in high school, when controlling for the selection bias specific to California's reclassification policy. A strength of the study was the adoption of instrumental variable estimation to control for selection bias such as subjective teacher input.

Robinson-Cimpian and Thompson (2016) investigated the relation between the adoption of a stricter reclassification policy in California and its effects on the performance of English learners in English language arts. In 2006-2007, state policymakers in California increased the reclassification criteria on the state language proficiency test, which made reclassification more difficult for English learners. Researchers utilized a sample consisting of 609,431 Latino English learners in grades 3-12 from the Los Angeles School District. In addition to utilizing a regression discontinuity and instrumental variable estimation, the study also combined difference in differences and inverse probability treatment weighting. The difference in differences approach controls for the effect of reclassification that would have occurred under the previous policy (Murnane & Willett, 2011). Inverse probability treatment weighting allowed the researcher to account for the imbalanced number of reclassified English learners compared to English learners. English language arts performance was measured by the state's assessment. Results indicated that a stricter policy in California for determining English proficiency did not change the performance on English language arts assessments of students in grades 3-8. Before the new policy, the decision to reclassify English learners resulted in negative consequences on English language arts performance for high school students but diminished after the implementation of the stricter policy. The adoption of a new policy had no consequences for students in grades 3-8 but did help students in high school by increasing their academic performance, which provides evidence of positive consequences of stricter reclassification decisions. This study adds to the previous research by extending results to include effects of policy change at the state level and how more rigorous criteria for reclassification can benefit older English learners. As will be highlighted in the next section, this study also analyzed graduation and college academic outcomes.

Studies of Graduation and College Readiness Outcomes

Both studies of graduation and college readiness outcomes utilized quasi-experimental methods to research the impact of reclassification decisions. Table 3 illustrates the two studies addressed the relation between reclassification on graduation rate or college readiness (e.g., ACT performance) of English learners (Carlson & Knowles, 2016; Robinson-Cimpian & Thompson, 2016). Robinson-Cimpian and Thompson (2016), as mentioned in the previous section, were also interested in whether a policy change in California regarding reclassification affected graduation rates among English learners. The researchers utilized the same sample, school district, and research design. There was no indication of changes in state graduation standards during the study. Findings suggest that modifications to a stricter reclassification policy lead to improvements in graduation rates among English learners. Therefore, the adoption of more stringent reclassification criteria resulted in additional positive consequences of reclassification for reclassified English learners.

Table 3

Studies of Outcomes related to Graduation and College Readiness

Study	State	Number of Criteria	Assessment	Method
Carlson & Knowles (2016)	Wisconsin	4	WIDA ACCESS	Regression Discontinuity
Robinson-Cimpian & Thompson (2016)	California	4	CELDT	Regression Discontinuity, Instrumental Variable Estimation, Difference-in-Differences, Inverse Probability Weighting

Carlson and Knowles (2016) were interested in whether students reclassified as former English learners and current English learners in Wisconsin scored similarly on the ACTs. The sample consisted of 2,733 students representing all students ever identified as an English learner between the 2006-2007 and 2012-2013 school years. The researchers adopted a regression discontinuity approach. Results indicated that students who were reclassified by tenth grade had higher scores on the ACT with some evidence of an increase in high school graduation. Thus, decisions to reclassify students in Wisconsin resulted in positive consequences for ACT performance and some evidence of positive consequences for graduation. Unlike the last study, which only looked at Spanish-speaking English learners, the current study sample included students who spoke Hmong as their native language given that in Wisconsin, the proportion of Hispanic English learners is approximately the same as the proportion of English learners who speak Hmong. Unlike the previous study that utilized an instrumental variable, an instrumental variable was not needed for the current study because reclassification decisions are based solely on the English language proficiency assessment in Wisconsin.

Carlson and Knowles (2016) also investigated the relation between reclassification and post-secondary enrollment. Findings suggested that reclassification by tenth grade increased post-secondary enrollment. However, as only one study in the current review examined the relation between reclassification and post-secondary enrollment, more research is needed to understand this relation in different contexts.

Summary of Literature Review

Three main findings emerged from the literature review. First, I found that there is no one way of researching the effects of reclassification and these methods depend on the specific criteria for reclassification. For example, when researchers investigated reclassification within the context of California, an instrumental variable estimation approach was necessary to control for selection bias (Robinson, 2011), whereas in Wisconsin (Carson & Knowles, 2016) an instrumental variable was not necessary as reclassification decisions are solely based upon the English language proficiency assessment.

Secondly, I found the impacts of reclassification decisions on English learners were not fully understood and more research is needed to understand its impact on other outcomes (e.g., advanced coursework enrollment). This lack of understanding may be attributed to the methodological limitations that impact the contributions of the current research. For example, types of programming available (e.g., English as a second language, bilingual) to English learners may attribute to success with reclassification decisions. Furthermore, literature related to rates of reclassification highlight differential reclassification rates by subgroups (Burke, Morita-Mullaney, & Singh, 2016; Greenberg Motamedi, Singh, & Thompson, 2016), but little is known about the impacts of reclassification decisions by subgroup characteristics (e.g., gender, students with disabilities).

Lastly, I found a better understanding of the relation between consequential validity and reclassification decisions. Most studies reviewed were only interested in students' subsequent achievement or distal outcomes such as graduation and ACT performance. I argue for analyses of both to fully understand the effects of reclassification decisions.

Research Questions

The current dissertation seeks to build upon previous research by analyzing the consequential validity of reclassification decisions in North Carolina. To better comprehend the consequences of the decision to reclassify English learners, I will address the following research question(s):

1. What are the consequences of reclassification decisions of English learners in North Carolina on (a) their English language Arts performance, and (b) their mathematics performance compared to the consequences of not being reclassified?
    i. Do the consequences of reclassification decisions on English learners' English language arts and mathematics performance vary by student characteristics (i.e. gender, ethnicity, disability status, economic status)?
2. What are the consequences of reclassification decisions of English learners in North Carolina on (a) their advanced placement course enrollment, (b) their ACTs, and (c) their high school graduation compared to the consequences of not being reclassified?
    i. Do the consequences of reclassification decisions on English learners' advanced course enrollment, their ACTs, and their high school graduation vary by student characteristics?

CHAPTER III: METHOD

To better understand the effects of reclassification on English learner outcomes and the appropriateness of reclassification decisions, I studied the outcomes of reclassification in a state context not previously studied, North Carolina. As mentioned previously, the English learner population in North Carolina has doubled from 2000 to 2015, which highlights the importance of understanding the impacts of reclassification decisions (National Center for Education Statistics, 2017, Table 219.46). In this chapter I describe the data, the population in North Carolina, the outcome variables, the analytic strategies for each outcome type, and the analytic models.

Data

I utilized student-level performance data, student-level demographic data, and school-level data from the state of North Carolina. I obtained data from the Duke Center for Child and Family Policy (n.d.), which houses most of North Carolina's public-school data. Data were available from as early as 1980 to as recent as 2017. I chose to use data after 2008-2009 because North Carolina adopted the WIDA ACCESS for measuring English language proficiency the summer of 2008 (North Carolina Department of Public Instruction, 2018). Multiple datasets from the Duke Center for Child and Family Policy were available, and the data provided depends on the need of the researcher. Examples of datasets include yearly performance on end-of-grade tests, demographic information, and student exceptionalities. These datasets aided in estimating the effects of reclassification on English learners in North Carolina. A unique identifier connects student data which is consistently used from year to year to help track student performance over time. Data on school, district, and teachers are also linked. An abbreviated list of available variables is available in Appendix B.

I requested data and was approved from the Duke Center for Child and Family Policy. Approval from Duke required institutional review board (IRB) approval from Southern Methodist University which included a data security plan. Data were stored and analyzed on a secure remote desktop and could only be accessed by connecting to the remote desktop. Copies of the IRB approval and the data security plan can be found in Appendix A.

Population Description

The datasets consisted of students from 157 local education agencies (e.g., school districts) and 304 schools. Multiple cohorts (i.e., grades) of students were available for the analyses to help meet the assumptions of the methods and to gain enough power to detect significant effects. Three years of consecutive performance data were required due to the analytical strategy I adopted. Therefore, the earliest grade-level analyzed was fifth grade.

Table 4 provides the total number of English learners and the number reclassified in North Carolina by the grade of reclassification. For example, of students who were in fifth grade in 2008-2009, 1,187 were reclassified in sixth grade (2009-2010), and 819 were reclassified in seventh grade (2010-2011). Similarly, Table 5 illustrates the subgroup classification of the students as well as the reclassification count by grade 12. These tables show disproportionate rates of reclassification in some instances. The rate of reclassification for White and Asian English learners were almost six to ten percentage points higher than Hispanic ($z = 4.69$, $p < .001$; $z = 10.8$, $p < .001$) and Black English learners ($z = 4.29$, $p < .001$; $z = 6.33$, $p <.001$). Females were reclassified at a higher rate than males ($z = 12.2$, $p <.001$). The rate of reclassification for students without a disability was twice the percentage of students with a disability ($z = 27.9$, $p <.001$), and the percentage of economically disadvantaged students reclassified was higher than the percentage of students not identified as economically disadvantaged ($z = 6.13$, $p < .001$).

Table 4

Counts of Reclassified English Learners by School Year (N = 42,393)

Grade in 2008-2009	Year	Grade	Count Reclassified
5th (*N* = 6,243)	2009-2010	6th	1,187
	2010-2011	7th	819
	2011-2012	8th	365
	2012-2013	9th	166
	2013-2014	10th	1,497
	2014-2015	11th	386
	2015-2016	12th	168
		Total	4,588
6th (*N* = 7,713)	2009-2010	7th	1,394
	2010-2011	8th	808
	2011-2012	9th	229
	2012-2013	10th	1,976
	2013-2014	11th	419
	2014-2015	12th	209
		Total	5,035
7th (*N* = 7,027)	2009-2010	8th	960
	2010-2011	9th	607
	2011-2012	10th	1,755
	2012-2013	11th	555
	2013-2014	12th	251
		Total	4,128
8th (*N* = 6,556)	2009-2010	9th	558
	2010-2011	10th	1,969
	2011-2012	11th	551
	2012-2013	12th	365
		Total	3,443
9th (*N* = 7,175)	2009-2010	10th	1,403
	2010-2011	11th	964
	2011-2012	12th	330
		Total	2,697
10th (*N* = 4,369)	2009-2010	11th	1,020
	2010-2011	12th	443
		Total	1,463
11th (*N* = 3,310)	2009-2010	12th	765

Table 5

Descriptive Statistics for English Learners and Counts Reclassified by Grade 12

Variable	Subgroup	English Learners (%)	Reclassified English Learners (%)
Ethnicity			
	American Indian	25 (0.06%)	6 (0.01%)
	Asian	2,726 (6.43%)	2,368 (5.59%)
	Black	812 (1.92%)	470 (1.11%)
	Hispanic	21,000 (49.54%)	13,193 (31.12%)
	Multi-Ethnic	248 (0.59%)	151 (0.36%)
	White	769 (1.81%)	625 (1.47%)
Gender			
	Male	14,314 (33.77%)	8,392 (19.80%)
	Female	11,266 (26.58%)	8,421 (19.86%)
Disability			
	Yes	3,633 (8.57%)	933 (2.20%)
	No	22,007 (51.91%)	15,820 (37.32%)
Economically Disadvantaged			
	Yes	20,107 (47.43%)	13,628 (32.15%)
	No	5,473 (12.91%)	3,185 (7.51%)

Variables

There are seven outcome variables of interest for the current study. Two variables relate to achievement and five variables relate to college readiness and graduation. I briefly describe each variable in the next sections.

Achievement Variables

Two achievement outcome variables are English language arts and mathematics performance as measured by the state's End of Grade tests. These tests are administered at the end of the year at each grade between grades three through eight and are summative assessments of what students have learned in both English language arts and mathematics in their respective grades (North Carolina Department of Public Instruction, n.d.). These tests are intended to measure English language art and mathematics content learned in relation to the state's curriculum (i.e., criterion referenced). Due to multiple cohorts (i.e., grades) of students included which reflect different grade-level assessments, I scaled both variables (z-scored) by year to have a grand mean of zero and a standard deviation of one to make comparisons with multiple assessments.

College Readiness and Graduation Variables

The third, fourth, and fifth outcome variables relate to students' ACT English, reading, and mathematics performance. I chose to utilize ACT performance over SAT performance because the state mandates all students take the ACTs (North Carolina Department of Public Schools, n.d.b). Scores on each test range from 1 to 36 (ACT, n.d.). The ACT English subtest measures students' understanding of English, with an emphasis on language skills (e.g., grammar). The ACT reading subtest measures reading comprehension indicative of a first-year college student, and the mathematics subtest measures the comprehensive mathematics skills students are expected to acquire by 12^{th} grade. The ACT writing subtest was not included due to the student's ability to self-select out of the writing portion of the test and the ACT science subtest was not included due to the lack of three consecutive years of academic performance data to match.

The sixth and seventh outcome variables are both dichotomous variables. The graduation variable takes the value of 1 if a student graduated in four years and the value of 0 if a student did not graduate in four years. Similarly, the AP enrollment variable takes the value of 1 if a student enrolled in an AP course and 0 if the student did not enroll in an AP course. I chose to dichotomize the variable due to the large number of students who did not take an AP course.

Analytic Strategy

In this section, I describe the analytic strategy for the different types of outcomes separately because each type of outcome variable necessitated a slightly different analysis. In particular, I describe the difference in differences approach for the achievement outcomes and the regression approach for the graduation and college readiness outcomes.

Analytic Strategy for Achievement Outcomes

In previous studies, performance data on the English language proficiency assessment were readily available, which is a necessity to assess the effects of the forcing variable (i.e., reclassification threshold) in regression discontinuity research (Carlson & Knowles, 2016; Robinson-Cimpian & Thompson, 2016). However, in North Carolina performance data on the English language proficiency assessment were not available and other methodological considerations were needed. For the current study, I adopted a difference in differences approach to establish the treatment effect of reclassification. Weights were added to the regression model to account for the imbalance in the number of students reclassified compared to students not reclassified. The significance of the difference in differences term provides insight into the effects of reclassification decisions in North Carolina.

Analytic Strategy for Outcomes Related to College Readiness and Graduation

To better understand the consequences of reclassification on graduation and college readiness outcomes, I adopted different methodologies than the previously mentioned outcomes due to the lack of baseline data on these outcomes. Unlike the change of English learner status from one year to the next that was analyzed in the first research question, students may have received the "treatment" of reclassification between the grade of analysis for the academic performance outcomes and the grade levels in the current section (e.g., eleven or twelve). For example, if reclassification is studied between grades six and twelve, then students can be reclassified in any grade between grade six and the grade students takes the ACT, or the grade students graduate. Therefore, instead of a difference in differences approach utilized in the first research questions, I adopted weighted regression analyses to predict these outcomes.

Regression analyses allowed me to predict the likelihood of graduating/AP enrollment or predict performance on the ACT. I included weights to account for the imbalance between the number of reclassified English learners and the number of English learners never reclassified. I am restricted to three grade levels in the analyses for ACT performance because the ACT for grade eleven was not mandated in North Carolina until 2013 (North Carolina Testing Program, 2017). For graduation and the AP course enrollment outcomes, I utilized all seven cohorts. Table 6 summarizes the *N* by outcome variable.

Table 6

Available N by Outcome Variable

Variable	Number of Cohorts	*N*
English Language Arts/Mathematics	3	20,983
ACTs	3	20,983
Graduation/AP Enrollment	7	42,393

Matching Strategy

A difference in differences approach assumes parallel trajectories (i.e., performance over time), which did not apply to the current data in its current state. As expected, large variability exists within the English learner group, and not all English learners performed similarly to those who were reclassified. Therefore, I adopted coarsened exact matching to establish comparable comparison groups before I conducted analyses. I discuss the matching technique on each outcome type in the following sections.

Matching Strategy for Achievement Outcomes

I adopted Coarsened exact matching (CEM) to match reclassified students with a comparable group of English learners. I chose CEM to rectify numerous limitations associated with traditional techniques (Iacus et al., 2011). First, CEM does not rely on a cycle of balance checks that other matching techniques require (e.g., propensity score matching). Second, CEM treats missing data as a value to match upon. For example, if two students performed similarly on two out of the three continuous matching variables but were missing on the third, the two students would match. Missing data on any variable ranged from 0% to 4.9%, which makes matching via propensity score analysis difficult due to its reliance on complete data (Murnane & Willett, 2011). To conduct matching via propensity scores with missing data, researchers must impute the data before conducting the matching technique. Coarsened exact matching circumvents the need to impute data values. Lastly, adjustments to allowable imbalance on any of the matching variables does not affect the imbalance of other matching variables, which is a limitation of other techniques that generate a composite (i.e., propensity scores). Figure 1 provides an illustration of the CEM approach.

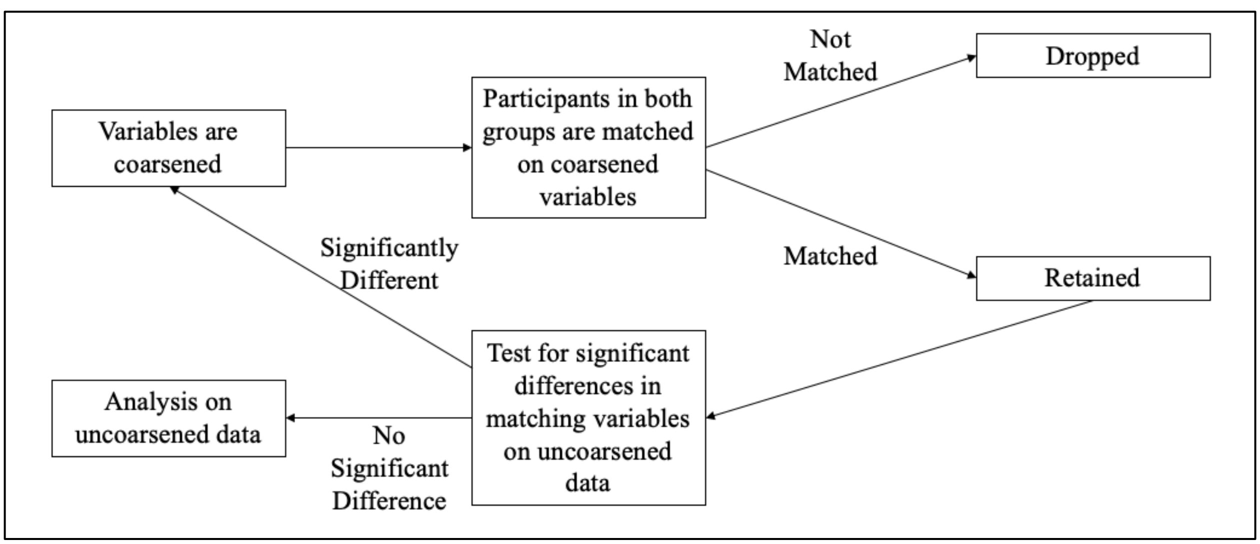

Figure 1. Process of Coarsened Exact Matching

The R statistical packages CEM (Iacus, King, & Porro, 2018) and PLM (Croissant & Millo, 2008) were utilized. First, the CEMSPACE function in the CEM package conducts an iterative process of coarsening the continuous matching variables and testing the difference between the treatment and control groups. The coarsening process is a balance of sample size and equivalence of groups. If one coarsens too much, the sample size will be bigger, but the difference between groups at baseline will be significant. In contrast, if insufficient coarsening is conducted, the difference between groups at baseline are insignificant, but the sample size will not be sufficient to detect significant results.

The coarsening process provides intervals to group similar students based on continuous variables and exact matches for categorical variables. If executed correctly, coarsened exact matching will provide comparable treatment and control groups with similar baseline achievement on the matching variables. Matching occurred on three consecutive years of end of grade test data standardized to a mean of zero and standard deviation of one. I assessed the baseline differences of the groups before and after matching.

CEM can match in multiple ways depending on the nature of the data provided (Iacus, King, & Porro, 2009). For example, CEM can execute a one-to-one match which restricts each subject in the control group to one subject in the treatment group. This technique may be useful if the two populations of interest are of equal size. However, if imbalance between the two populations are suspected, the CEM algorithm can match a single subject in the smaller group to multiple individuals in the larger group, or a one-to-many match. The CEM algorithm also provides weights to include in subsequent analyses to account for the imbalance between sample sizes. I adopt a one-to-many match in the current study due to the imbalance in the number of students reclassified compared to students who remain in English learner services.

Results of matching for achievement outcomes. The match rate for the English language arts and mathematics outcomes were 90.5% and 91.8% for reclassified English learners matched to English learners, respectively. Tables 7 and 8 provide an overview of the sample after matching for academic outcomes.

Table 7

Description of Sample after Matching on English Language Arts Performance

Variable	Subgroup	English Learners	Reclassified English Learners
Ethnicity			
	Hispanic	4,925 (71.7%)	1,941 (28.3%)
	Non-Hispanic	962 (60.8%)	617 (39.1%)
Gender			
	Male	3,073 (72.4%)	1,169 (27.6%)
	Female	2,814 (67.0%)	1,389 (33.0%)
Disability			
	Yes	384 (91.4%)	36 (8.6%)
	No	5,503 (68.6%)	2,522 (31.4%)
Economically Disadvantaged	Yes	4,981 (71.6%)	1,979 (28.4%)
	No	906 (61.0%)	579 (39.0%)

Table 8

Description of Sample after Matching on Mathematics Performance

Variable	Subgroup	English Learners	Reclassified English Learners
Ethnicity			
	Hispanic	5,717 (74.3%)	1,973 (25.7%)
	Non-Hispanic	1,126 (64.4%)	622 (35.6%)
Gender			
	Male	3,718 (76.1%)	1,167 (23.9%)
	Female	3,125 (68.6%)	1,428 (31.4%)
Disability			
	Yes	543 (93.9%)	35 (6.1%)
	No	6,300 (71.1%)	2,560 (28.9%)
Economically Disadvantaged	Yes	5,827 (74.3%)	2,014 (25.7%)
	No	1,016 (63.6%)	581 (36.4%)

Matching Strategy for Outcomes Related to College Readiness and Graduation

Similar to the academic outcomes, I matched on three consecutive years of test data standardized to a mean of zero and standard deviation of one. The matching variable depends on the outcome of interest. I matched on three years of consecutive English language arts performance for the ACT reading and English subtests, and three years of consecutive mathematics performance for the ACT mathematics subtest. Furthermore, I average students' English language arts and math performance and match on this averaged variable over three years for the graduation and AP enrollment outcomes. Again, I assessed for the quality of the match to ensure baseline differences diminished after matching. However, instead of a one-to-many match with reclassified English learners as the smaller group, the group of English learners became the smaller group due to more students being reclassified over time.

Results of matching for outcomes related to college readiness and graduation. The match rate for ACT Reading and English was 95.1% for the English learner group on baseline English language arts performance. I report the match rate for the English learner group for these outcomes due to the nature of group mobility over time. It is expected that more students are reclassified over time. Therefore, there are more students to match in the reclassified English learner group as compared to the English learner group.

Similarly, the match rate for the ACT mathematics outcomes was 87.9% for English learners matched to reclassified English learners on baseline mathematics performance. Finally, the match rate for the graduation and AP enrollment outcomes was 98.0% for English learners matched to reclassified English learners on baseline mathematics performance. I adopted mathematics performance as the matching variable for graduation and AP enrollment because initial analyses suggested mathematics performance as a stronger predictor of these outcomes.

Tables 9-11 provide the breakdown of both groups by student characteristics for the ACT outcomes and the graduation/AP enrollment outcomes, respectively.

Table 9

Description of Sample after Matching for ACT English and Reading Performance

Variable	Subgroup	English Learners	Reclassified English Learners
Ethnicity			
	Hispanic	1,336 (21.2%)	4,952 (78.8%)
	Non-Hispanic	232 (18.3%)	1,034 (81.7%)
Gender			
	Males	903 (23.6%)	2,924 (76.4%)
	Females	665 (17.8%)	3,062 (82.2%)
Disability			
	Yes	551 (60.6%)	358 (39.4%)
	No	117 (2.04%)	5,628 (98.0%)
Economically Disadvantaged	Yes	1,309 (21.7%)	4,733 (78.3%)
	No	259 (17.1%)	1,253 (82.9%)

Table 10

Description of Sample after Matching for ACT Mathematics Performance

Variable	Subgroup	English Learners	Reclassified English Learners
Ethnicity			
	Hispanic	1,233 (24.0%)	3,897 (76.0%)
	Non-Hispanic	216 (22.7%)	737 (77.3%)
Gender			
	Males	843 (28.1%)	2,161 (71.9%)
	Females	606 (19.7%)	2,473 (80.3%)
Disability			
	Yes	511 (65.7%)	267 (34.3%)
	No	938 (17.7%)	4,367 (82.3%)
Economically Disadvantaged	Yes	1,209 (25.0%)	3,631 (75.0%)
	No	240 (19.3%)	1,003 (80.7%)

Table 11

Description of Sample after Matching for Graduation and AP Course Enrollment

Variable	Subgroup	English Learners	Reclassified English Learners
Ethnicity			
	Hispanic	4,216 (25.6%)	12,263 (74.4%)
	Non-Hispanic	1,233 (27.8%)	3,196 (72.2%)
Gender			
	Males	2,955 (28.0%)	7,660 (72.0%)
	Females	2,494 (24.2%)	7,799 (75.8%)
Disability			
	Yes	1,102 (61.9%)	678 (38.1%)
	No	4,347 (22.7%)	14,781 (77.3%)
Economically Disadvantaged	Yes	4,266 (27.4%)	11,298 (72.6%)
	No	1,183 (22.1%)	4,161 (77.9%)

Balance Checks

Matching techniques, such as propensity score matching, depend on iterative balance. However, this is not a requirement of coarsened exact matching due to the matching being exact for categorical variables and within specified intervals for continuous variables (Iacus et al., 2009). However, I still report the baseline differences pre- and post-matching to highlight the comparability between the groups.

Figure 2 provides an overview of the baseline differences before and after matching on the three consecutive years of academic performance. As illustrated by the figure, significant baseline differences existed before matching. After matching, the significant baseline differences diminished. Therefore, the matching technique was effective in establishing comparable groups at baseline (i.e., three consecutive years of performance).

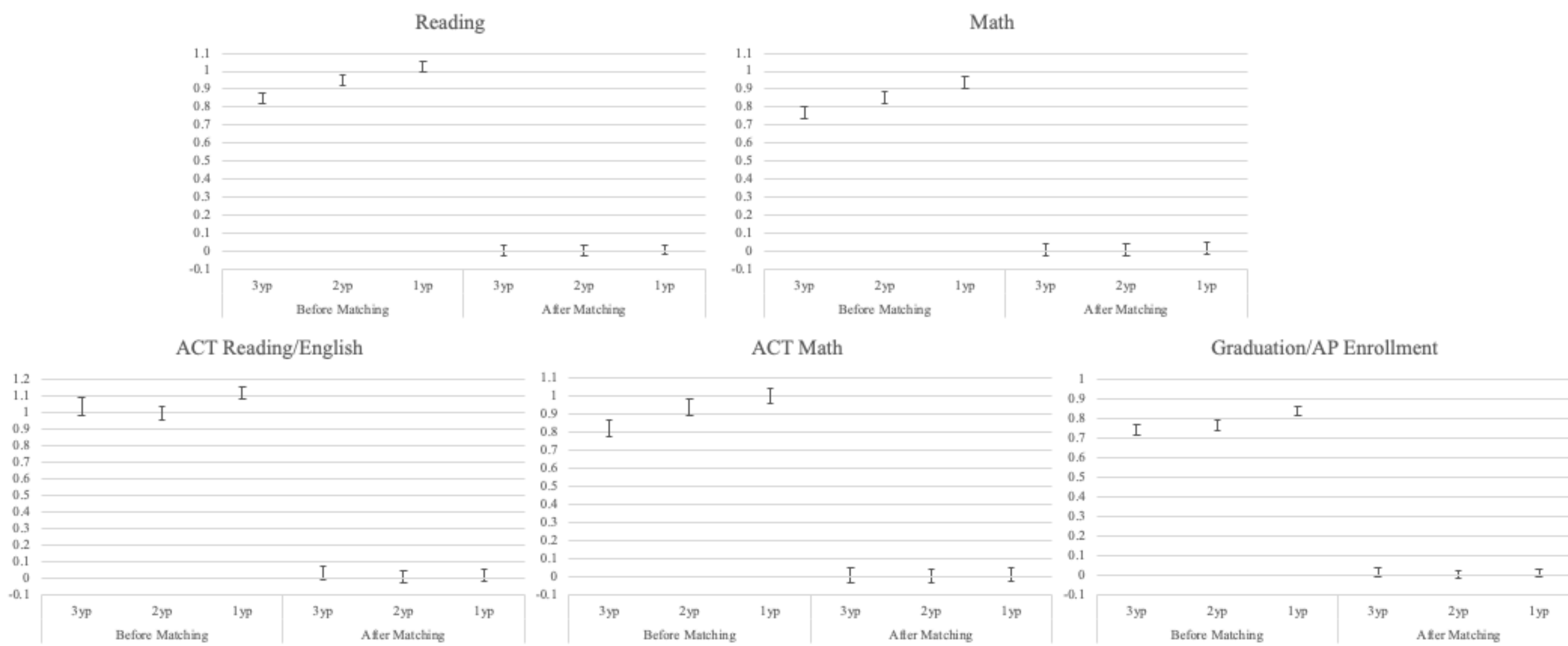

Figure 2. Differences between reclassified English learners and English learners before and after matching

Analytic Models

For academic outcomes, I modeled the difference in differences approach with multiple regression models. Similarly, I modeled the multiple reclassification time points for the outcomes related to college readiness and graduation with multiple regression techniques (i.e., multiple linear regression, multiple logistic regression). In this section, I explain the models I used by outcome type.

Analytic Models for Achievement Outcomes

I used the following equation to assess the effect of reclassification on the English language arts performance outcome:

$$y_{it} = \alpha_i + \beta_1(YEAR\_POST_t) + \beta_2(YEAR\_POST_t \times RECLASS_i) + \gamma_i + \delta_i + \boldsymbol{C} + \varepsilon_i$$

where α is a fixed effect for each student. Fixed effects help eliminate variation in non-time varying sources of observed and unobserved heterogeneity among students (Clark et al., 2010). I used student-level fixed effects instead of student-level covariates because controlling for observed student-level characteristics still leaves a source of unobserved heterogeneity among students. YEAR_POST is an indicator variable that represents the second year (0=2009, 1=2010) and the corresponding beta represents the change in performance from the first year to the second. I do not estimate the effects after more than one year of reclassification due to the difficulty in accounting for students who are reclassified after the first year. RECLASS is a categorical variable that represents whether a student was reclassified (0 = no, 1 = yes) and the corresponding beta represents the difference in performance between reclassified English learners and English learners. Lastly, the interaction represents the adjusted effect of reclassification on English Language arts performance.

The significance of the interaction term addresses the treatment effect of interest by providing an estimate of the reclassification effect. Furthermore, the regression model does not estimate the main effect of reclassification because reclassification does not occur without the change in year. I included the fixed effects for school (γ) and district (δ) to account for the shared variability among school and districts. I conducted the same analysis for the math performance outcome. Appendix C provides an overview of the regression equations.

I included interactions terms with subgroup characteristic variables (**C**) in the analytical model to better understand what characteristics moderate the effects of reclassification decisions. For example, to test whether gender may moderate the impact of reclassification on student outcomes, the interaction term *(*YEAR_POST *x* RECLASS *x* GENDER*)* was included in the model and examined. Previous research has found females are reclassified at a faster rate than males (Greenberg Motamedi et al., 2016). The significance of the interaction term would further these results and address the consequences of reclassification decisions for different subgroups of students.

Analytic Models for Outcomes Related to College Readiness and Graduation

To represent the multiple time points in which English learners could be reclassified for outcomes related to college readiness and graduation, I created new indicator variables representing the year students were reclassified. For example, if students were reclassified in tenth grade, the indicator iTENTH will equal one and zero for students who were not reclassified in tenth. The corresponding beta coefficient represents the difference in performance on the ACT for students reclassified in grade ten as compared to students who were never reclassified. I also included interaction terms with demographic variables to assess whether the effects of reclassification year on ACT performance depends on certain demographics. I utilized the following equation to evaluate the impact of reclassification year on ACT performance:

$$y_i = \beta_0 + \beta_1(iSIXTH_i) + \cdots + \beta_6(iTWELFTH_i) + \beta_3(iSIXTH_i \times GENDER_i) + \cdots + \gamma_i + \varepsilon_i$$

For example, the dummy variable *iSeventh* provides an estimate of the impact of reclassification in seventh grade on ACT performance. The interaction term, *iSeventhxGender,* provides an estimate of whether the effect of reclassification in seventh grade on ACT performance depends on gender. I also analyzed ethnicity, disability status, and economic status. The reference group is English learners never reclassified in this time period. All grade-level estimates compared students reclassified immediately before grades six through eleven to students never reclassified.

Significant main effects of reclassification year provide some indication of how English learners reclassified immediately before grades six through eleven compare to English learners never reclassified, which also contributes to my understanding of the consequences of reclassification decisions. For example, if students reclassified in tenth grade, on average, perform worse on the ACTs than students who were never reclassified, then the consequences of reclassification decisions should be considered. Furthermore, significant interaction effects provide insight into subgroup performance on the ACT by reclassification year and can also provide evidence for the consequences of reclassification decisions by subgroups. For example, if female English learners are negatively impacted by reclassification as compared to male English learners, then the decision to reclassify is not appropriate for all students. Similar to the academic performance outcomes, school-level (γ) and district level (δ) fixed effects were included in the model to control for differences attributed to school enrollment and district-level enrollment.

For categorical outcomes, such as Advanced Placement course enrollment and graduation status, appropriate adjustments were made to the regression to reflect the use of categorical data as a dependent variable. If the outcome is binary (i.e., 0 or 1), such as graduation status, then a logistic regression would be most appropriate (Tabachnick & Fidell, 2013). Interpretation of the beta coefficient changes from the linear effect on the dependent variable to the multiplicative increase in the odds of the outcome variable. Due to model convergence issues, only the school-level fixed effects were included in the model to control for the shared variability attributed to school enrollment. In the next chapter, the results of these analyses are presented.

CHAPTER IV: RESULTS

In the current chapter, I provide the results of the analyses discussed in the previous chapter. I begin with student achievement outcomes and then move to outcomes related to graduation and college readiness.

Achievement Outcomes

Table 12 provides the difference in differences estimates of reclassification on subsequent English language arts and mathematics performance, when controlling for individual, school-level, and district-level fixed effects. I find a significant positive effect of reclassification on English language arts and mathematics performance, which means reclassified English learners performed significantly better in English language arts and mathematics than English learners not reclassified. Furthermore, as highlighted in Table 13, the interactions between the difference in differences estimates and student characteristics were all non-significant (i.e., ethnicity, disability status, economic disadvantage, gender), which indicates reclassification decisions so not vary by student characteristics. I also tested the models with each subgroup individually and found similar results.

Table 12

Difference in Differences Estimates for English Language Arts and Mathematics

Variable	β	SE	N
ELA (SD)			
Reclass	.08	*(.030)* **	8,372
Math (SD)			
Reclass	.15	*(.031)* ***	9,353

Note. Controlling for individual, school, and district fixed effects.

Table 13

Difference in Differences Estimates for English Language Arts/Math by Subgroup Characteristic

Variable	*ELA (SDs)*	*Math (SDs)*
Ethnicity (Non-Hisp)	.028 *(.026)*	-.019 *(.027)*
Gender (Male)	.005 *(.020)*	-.036 *(.021)*
SWD (Yes)	-.074 *(.087)*	-.062 *(.091)*
EDS (Yes)	-.0003 *(.028)*	-.016 *(.029)*
N	8,372	9,353

Note. Controlling for individual, school, and district fixed effects; SWD = Students with disabilities; EDS = Economically disadvantaged status.

Graduation and College Readiness

In addition to achievement outcomes, I investigated five outcomes to evaluate the consequential validity of reclassification decisions on outcomes related to college readiness and graduation: ACT reading performance, ACT English performance, ACT math performance, graduation status, and AP course enrollment. I investigated these outcomes to better understand how reclassification decisions affect students' long-term outcomes. Investigations into all of these outcome variables provide increased evidence into the appropriateness of reclassification decisions. I present the results in this section by grade students were immediately reclassified before. For example, the effect for grade six is the estimate of reclassification at the end of fifth grade, when the determination is made that English language supports will no longer be provided in sixth grade. I first provide the results for ACTs, followed by the results for graduation and AP enrollment.

Results of ACT Outcomes

A weighted least squares regression was conducted with the year of reclassification as a predictor and controlling for school fixed effects for each subtest. Table 14 provides the estimates of these regressions. For the ACT English and reading subtests, reclassification before grades six through eleven was related to higher performance on these subtests compared to students who were never reclassified. Similarly, reclassification before grades six through ten was related to higher performance on the ACT math subtests compared to students who were never reclassified. Students reclassified before grade eleven were not significantly different from students never reclassified in their ACT math performance. Results from full model analyses are included in Appendix D.

Table 14

Weighted Least Squares Regression Estimates of ACTs subtests by Grade of Reclassification

	ACT English $N = 6{,}316$			ACT Reading $N = 6{,}316$			ACT Math $N = 4{,}999$		
	β	SE	N	β	SE	N	β	SE	N
Intercept	11.8	*(0.711)* ***	1,107	13.9	*(0.767)* ***	1,107	15.5	*(0.641)* ***	1,018
Grade 6 (G6)	4.37	*(0.222)* ***	309	4.14	*(0.239)* ***	309	1.35	*(0.163)* ***	249
Grade 7 (G7)	4.05	*(0.170)* ***	677	3.72	*(0.184)* ***	677	1.50	*(0.125)* ***	516
Grade 8 (G8)	3.55	*(0.173)* ***	634	3.34	*(0.187)* ***	634	1.53	*(0.125)* ***	511
Grade 9 (G9)	3.27	*(0.221)* ***	312	3.06	*(0.239)* ***	312	1.37	*(0.165)* ***	243
Grade 10 (G10)	1.75	*(0.127)* ***	2,557	1.63	*(0.137)* ***	2,557	0.73	*(0.089)* ***	1916
Grade 11 (G11)	0.44	*(0.164)* **	720	0.75	*(0.177)* ***	720	0.18	*(0.120)*	546

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Note. Controlling for school and district fixed effects.

Table 15

Weighted Least Squares Regression Estimates of ACTs by Grade of Reclassification and Ethnicity

	ACT English $N = 6{,}316$			ACT Reading $N = 6{,}316$			ACT Math $N = 4{,}999$		
	β	SE	N	β	SE	N	β	SE	N
Intercept	11.3	*(0.705)* ***	917	13.5	*(0.765)* ***	917	15.3	*(0.630)* ***	841
Grade 6 (G6)	4.21	*(0.243)* ***	248	4.16	*(0.264)* ***	248	1.26	*(0.176)* ***	202
Grade 7 (G7)	3.95	*(0.186)* ***	549	3.72	*(0.202)* ***	549	1.53	*(0.134)* ***	433
Grade 8 (G8)	3.27	*(0.189)* ***	505	3.24	*(0.206)* ***	505	1.35	*(0.133)* ***	427
Grade 9 (G9)	2.97	*(0.244)* ***	247	2.81	*(0.265)* ***	247	1.19	*(0.180)* ***	194
Grade 10 (G10)	1.75	*(0.139)* ***	2,081	1.61	*(0.150)* ***	2,081	0.64	*(0.096)* ***	1,574
Grade 11 (G11)	.543	*(0.177)* **	476	0.93	*(0.193)* ***	476	0.23	*(0.128)*	461
Ethnicity (Non-Hisp)	1.06	*(0.289)* ***	190	1.03	*(0.314)* **	190	0.82	*(0.196)* ***	177
G6 x Ethnicity (Non-Hisp)	1.03	*(0.561)*	61	0.05	*(0.609)*	61	0.67	*(0.411)*	47
G7 x Ethnicity (Non-Hisp)	0.85	*(0.437)*	128	0.29	*(0.474)*	128	0.06	*(0.331)*	83
G8 x Ethnicity (Non-Hisp)	1.53	*(0.441)* ***	129	0.62	*(0.478)*	129	1.29	*(0.329)* ***	84
G9 x Ethnicity (Non-Hisp)	1.54	*(0.549)* **	65	1.31	*(0.596)* *	65	0.90	*(0.407)* *	49
G10 x Ethnicity (Non-Hisp)	0.07	*(0.334)*	476	0.21	*(0.362)*	476	0.55	*(0.232)* *	342
G11 x Ethnicity (Non-Hisp)	-0.49	*(0.451)*	112	-1.01	*(0.489)* *	112	-0.23	*(0.327)*	85

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Note. Controlling for school and district fixed effects.

Table 16

Weighted Least Squares Regression Estimates of ACTs by Grade of Reclassification and Gender

	ACT English $N = 6,316$			ACT Reading $N = 6,316$			ACT Math $N = 4,999$		
	β	SE	N	β	SE	N	β	SE	N
Intercept	12.0	*(0.724)* ***	481	14.1	*(0.780)* ***	481	15.3	*(0.646)* ***	435
Grade 6 (G6)	4.12	*(0.296)* ***	197	4.30	*(0.319)* ***	197	1.26	*(0.217)* ***	158
Grade 7 (G7)	4.19	*(0.248)* ***	368	3.93	*(0.267)* ***	368	1.70	*(0.179)* ***	295
Grade 8 (G8)	3.18	*(0.250)* ***	357	3.09	*(0.269)* ***	357	1.58	*(0.178)* ***	306
Grade 9 (G9)	2.86	*(0.304)* ***	185	3.01	*(0.328)* ***	185	1.47	*(0.222)* ***	148
Grade 10 (G10)	1.55	*(0.198)* ***	1,190	1.62	*(0.213)* ***	1,190	0.77	*(0.139)* ***	943
Grade 11 (G11)	0.23	*(0.247)*	359	0.45	*(0.266)*	359	0.33	*(0.182)*	267
Gender (Males)	-0.45	*(0.219)* *	626	-0.40	*(0.236)*	626	0.40	*(0.147)* **	583
G6 x Gender (Males)	0.37	*(0.455)*	112	-0.74	*(0.490)*	112	0.53	*(0.333)*	91
G7 x Gender (Males)	-0.47	*(0.340)*	309	-0.59	*(0.366)*	309	-0.29	*(0.249)*	221
G8 x Gender (Males)	0.48	*(0.349)*	277	0.42	*(0.377)*	277	0.11	*(0.254)*	205
G9 x Gender (Males)	0.77	*(0.447)*	127	-0.08	*(0.481)*	127	0.02	*(0.334)*	95
G10 x Gender (Males)	0.30	*(0.257)*	1,367	-0.04	*(0.277)*	1,367	0.03	*(0.179)*	973
G11 x Gender (Males)	0.31	*(0.334)*	361	0.52	*(0.360)*	361	-0.19	*(0.244)*	279

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Note. Controlling for school and district fixed effects.

Table 17

Weighted Least Squares Regression Estimates of ACTs by Grade of Reclassification and Disability Status

	ACT English $N = 6{,}316$			ACT Reading $N = 6{,}316$			ACT Math $N = 4{,}999$		
	β	SE	N	β	SE	N	β	SE	N
Intercept	11.9	*(0.713)* ***	784	14.0	(0.770) ***	784	15.7	*(0.640)* ***	720
Grade 6 (G6)	4.30	*(0.228)* ***	304	4.14	(0.246) ***	304	1.25	*(0.168)* ***	245
Grade 7 (G7)	3.96	*(0.178)* ***	668	3.68	(0.192) ***	668	1.35	*(0.132)* ***	508
Grade 8 (G8)	3.48	*(0.181)* ***	615	3.31	(0.195) ***	615	1.42	*(0.132)* ***	497
Grade 9 (G9)	3.21	*(0.228)* ***	305	3.01	(0.247) ***	305	1.24	*(0.170)* ***	238
Grade 10 (G10)	1.74	*(0.137)* ***	2,395	1.66	(0.148) ***	2,395	0.62	*(0.099)* ***	1,798
Grade 11 (G11)	0.40	*(0.177)* *	638	0.69	(0.191) ***	638	0.05	*(0.131)*	482
Disability (Yes)	-0.47	*(0.292)*	323	-0.18	(0.316)	323	-0.64	*(0.164)* ***	298
G6 x Disability (Yes)	-0.12	*(1.55)*	5	-2.49	(1.67)	5	-3.06	*(1.14)* **	4
G7 x Disability (Yes)	0.61	*(1.18)*	9	0.79	(1.27)	9	-0.28	*(0.841)*	8
G8 x Disability (Yes)	-0.01	*(0.841)*	19	0.07	(0.908)	19	-1.34	*(0.634)* *	14
G9 x Disability (Yes)	-0.52	*(1.34)*	7	1.02	(1.44)	7	-0.62	*(1.04)*	5
G10 x Disability (Yes)	-0.73	*(0.399)*	162	-0.69	(0.431)	162	-.107	*(0.270)*	118
G11 x Disability (Yes)	0.15	*(0.491)*	82	0.50	(0.530)	82	0.43	*(0.344)*	64

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Note. Controlling for school and district fixed effects.

Table 18

Weighted Least Squares Regression Estimates of ACTs by Grade of Reclassification and Economic Disadvantaged Status (EDS)

	ACT English $N = 6,316$			ACT Reading $N = 6,316$			ACT Math $N = 4,999$		
	β	SE	N	β	SE	N	β	SE	N
Intercept	11.9	*(0.747)* ***	195	13.3	*(0.806)* ***	195	15.7	*(0.656)* ***	178
Grade 6 (G6)	4.47	*(0.436)* ***	90	5.03	*(0.470)* ***	90	1.63	*(0.299)* ***	81
Grade 7 (G7)	4.40	*(0.366)* ***	168	4.84	*(0.394)* ***	168	1.51	*(0.253)* ***	134
Grade 8 (G8)	4.17	*(0.382)* ***	139	4.50	*(0.412)* ***	139	1.73	*(0.260)* ***	122
Grade 9 (G9)	4.03	*(0.468)* ***	74	3.95	*(0.505)* ***	74	1.29	*(0.354)* ***	51
Grade 10 (G10)	1.83	*(0.293)* ***	524	2.39	*(0.316)* ***	524	0.66	*(0.192)* ***	407
Grade 11 (G11)	0.25	*(0.384)*	137	0.71	*(0.415)*	137	-0.08	*(0.270)*	107
EDS (Yes)	-0.13	*(0.278)*	912	0.82	*(0.300)* ***	912	-0.19	*(0.175)*	840
G6 x EDS (Yes)	-0.14	*(0.505)*	219	-1.13	*(0.545)* *	219	-0.43	*(0.355)*	168
G7 x EDS (Yes)	-0.46	*(0.411)*	509	-1.39	*(0.443)* **	509	-0.03	*(0.288)*	382
G8 x EDS (Yes)	-0.78	*(0.425)*	495	-1.44	*(0.459)* **	495	-0.27	*(0.296)*	389
G9 x EDS (Yes)	-1.01	*(0.530)*	238	-1.11	*(0.572)*	238	0.11	*(0.399)*	192
G10 x EDS (Yes)	-0.11	*(0.323)*	2,033	-0.92	*(0.349)* **	2,033	0.09	*(0.216)*	1,509
G11 x EDS (Yes)	0.23	*(0.426)*	583	0.06	*(0.459)*	583	0.33	*(0.303)*	439

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Note. Controlling for school and district fixed effects.

Table 15 highlights the interaction results between year of reclassification and ethnicity on each ACT subtest. Non-Hispanic students reclassified before grades eight and nine performed significantly better on the ACT English subtest compared to Hispanic students reclassified at the same time. For the ACT reading subtest, non-Hispanic students reclassified before grade nine performed significantly better than Hispanic students reclassified at the same time. However, non-Hispanic students reclassified before grade eleven performed significantly worse on the ACT reading subtest than Hispanic students reclassified at the same time. Lastly, non-Hispanic students reclassified before grades eight, nine, and ten performed significantly better on the ACT math subtest compared to Hispanic students reclassified at the same time.

A significant interaction was also detected between grade of reclassification and disability status. Table 17 provides the weighted least squares regression estimates of this model. Students identified with a disability reclassified before grades six and eight performed significantly lower than students not identified with a disability reclassified at the same time.

Lastly, significant interactions were detected between year of reclassification and economically disadvantaged status on the ACT reading subtest. Table 18 provides the weighted least squares estimates. Economically disadvantaged students reclassified before grades six, seven, eight, and ten performed significantly lower on the ACT reading subtest compared to non-economically disadvantaged students reclassified at the same time. Gender did not significantly interact with year of reclassification (see Table 16).

Results of Graduation and AP Enrollment Outcomes

Graduation. A weighted logistic regression model with school fixed effects was utilized to model the log likelihood of graduation by grade of reclassification and student demographic variables. Tables 19-23 provide the results of the aggregated and disaggregated analyses.

Overall, reclassification before grades seven, eight, and ten related to significantly higher odds of graduating compared to students never reclassified. For example, students reclassified before grade seven were $e^{1.14} = 3.13$ times the odds to graduate compared to never reclassified English learners. The non-significant results for grades six and nine may be attributed to transition years in which students transition from one school to the next (e.g., elementary to middle school), which may attenuate the impact of reclassification decisions at these grade levels.

Analyses suggested certain grades significantly interacted with ethnicity. Non-Hispanic students reclassified before grade eight were significantly less likely to graduate compared to Hispanic students reclassified at the same time. Furthermore, one grade level significantly interacted with economically disadvantaged status. Students identified as economically disadvantaged and reclassified before grade ten were significantly less likely to graduate compared to non-economically disadvantaged students reclassified at the same time. We discuss these findings in the next section.

AP enrollment. Similar to the graduation outcome variable, a weighted logistic regression model with school-level fixed effects was adopted to model the likelihood of a student enrolling in an Advanced Placement course. As illustrated in Table 19, students reclassified before grades six through twelve had a significantly higher likelihood of enrolling in Advanced Placement courses than never reclassified English learners. Furthermore, reclassification year interacted with ethnicity and economic status, which are illustrated in Tables 20 and 23, respectively. Non-Hispanic students reclassified before grade twelve were more likely to take AP courses than Hispanic students reclassified at the same time. Furthermore, economic disadvantaged students reclassified before grades seven and nine were significantly more likely to enroll in AP courses than non-economically disadvantaged students reclassified at the same time. However, English learners never reclassified had low odds of enrolling in AP courses ($e^{-2.14} = .118$).

Table 19

Weighted Logistic Regression Estimates for Graduation and AP Enrollment by Grade of Reclassification

	Graduation $N = 20{,}703$				AP Enrollment $N = 20{,}703$			
	β	e^{β}	SE	N	β	e^{β}	SE	N
Intercept	2.47	11.8	*(0.485)* ***	5,449	-2.14	0.12	*(0.298)* ***	5,449
Grade 6 (G6)	0.28	1.32	*(0.189)*	723	1.37	3.94	*(0.106)* ***	723
Grade 7 (G7)	1.14	3.13	*(0.203)* ***	1,359	1.52	4.57	*(0.085)* ***	1,359
Grade 8 (G8)	0.51	1.67	*(0.163)* **	1,356	1.39	4.01	*(0.086)* ***	1,356
Grade 9 (G9)	0.18	1.19	*(0.164)*	950	1.33	3.78	*(0.097)* ***	950
Grade 10 (G10)	0.36	1.44	*(0.093)* ***	6,307	0.62	1.85	*(0.066)* ***	6,307
Grade 11 (G11)	0.03	1.03	*(0.109)*	2,760	0.39	1.48	*(0.081)* ***	2,760
Grade 12 (G12)	0.13	1.13	*(0.125)*	2,004	0.50	1.66	*(0.088)* ***	2,004

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Note. Controlling for school fixed effects.

Table 20

Weighted Logistic Regression Estimates for Graduation and AP Enrollment by Grade of Reclassification and Ethnicity

	Graduation *N* = 20,703				AP Enrollment *N* = 20,703			
	β	e^{β}	SE	N	β	e^{β}	SE	N
Intercept	2.38	10.8	*(0.487)* ***	4,216	-2.49	0.08	*(0.308)* ***	4,216
Grade 6 (G6)	0.36	1.43	*(0.208)*	581	1.52	4.57	*(0.126)* ***	581
Grade 7 (G7)	1.09	2.97	*(0.214)* ***	1,089	1.65	5.21	*(0.104)* ***	1,089
Grade 8 (G8)	0.70	2.01	*(0.189)* ***	1,091	1.46	4.31	*(0.106)* ***	1,091
Grade 9 (G9)	0.18	1.20	*(0.183)*	732	1.41	4.10	*(0.119)* ***	732
Grade 10 (G10)	0.38	1.46	*(0.103)* ***	5,107	0.70	2.00	*(0.084)* ***	5,107
Grade 11 (G11)	0.10	1.11	*(0.123)*	2,162	0.35	1.43	*(0.106)* ***	2,162
Grade 12 (G12)	0.07	1.08	*(0.137)*	1,501	0.31	1.36	*(0.121)* *	1,501
Ethnicity (Non-Hisp)	0.35	1.42	*(0.159)*	1,233	0.90	2.45	*(0.109)* ***	1,233
G6 x Ethnicity (Non-Hisp)	-0.31	0.74	*(0.484)*	142	-0.21	0.82	*(0.242)*	142
G7 x Ethnicity (Non-Hisp)	0.82	2.26	*(0.768)*	270	-0.13	0.88	*(0.189)*	270
G8 x Ethnicity (Non-Hisp)	-0.80	0.45	*(0.371)* *	265	0.01	1.01	*(0.191)*	265
G9 x Ethnicity (Non-Hisp)	0.04	1.04	*(0.415)*	218	-0.06	0.94	*(0.210)*	218
G10 x Ethnicity (Non-Hisp)	0.06	1.06	*(0.247)*	1,200	-0.04	0.96	*(0.137)*	1,200
G11 x Ethnicity (Non-Hisp)	-0.27	0.76	*(0.268)*	598	0.22	1.24	*(0.169)*	598
G12 x Ethnicity (Non-Hisp)	0.40	1.49	*(0.349)*	503	0.48	1.62	*(0.185)* **	503

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Note. Controlling for school fixed effects.

Table 21

Weighted Logistic Regression Estimates for Graduation and AP Enrollment by Grade of Reclassification and Gender

	Graduation *N* = 20,703				AP Enrollment *N* = 20,703			
	β	e^{β}	SE	N	β	e^{β}	SE	N
Intercept	2.60	13.5	*(0.493)* ***	2,494	-2.03	0.13	*(0.304)* ***	2,494
Grade 6 (G6)	0.15	1.16	*(0.256)*	412	1.31	3.71	*(0.146)* ***	412
Grade 7 (G7)	1.02	2.77	*(0.274)* ***	744	1.55	4.71	*(0.119)* ***	744
Grade 8 (G8)	0.45	1.56	*(0.222)* *	775	1.35	3.86	*(0.119)* ***	775
Grade 9 (G9)	-0.0002	1.00	*(0.216)*	560	1.27	3.56	*(0.132)* ***	560
Grade 10 (G10)	0.39	1.47	*(0.146)* **	3,021	0.62	1.86	*(0.096)* ***	3,021
Grade 11 (G11)	0.06	1.06	*(0.167)*	1,339	0.47	1.61	*(0.116)* ***	1,339
Grade 12 (G12)	0.24	1.27	*(0.199)*	948	0.54	1.72	*(0.126)* ***	948
Gender (Males)	-0.22	.803	*(0.133)*	2,955	-0.17	0.84	*(0.105)*	2,955
G6 x Gender (Males)	0.24	1.26	*(0.379)*	311	0.08	1.08	*(0.124)*	311
G7 x Gender (Males)	0.21	1.24	*(0.411)*	615	-0.10	0.91	*(0.170)*	615
G8 x Gender (Males)	0.06	1.07	*(0.329)*	581	0.04	1.04	*(0.173)*	581
G9 x Gender (Males)	0.35	1.42	*(0.339)*	390	0.08	1.09	*(0.195)*	390
G10 x Gender (Males)	-0.06	0.94	*(0.188)*	3,286	-0.03	0.97	*(0.131)*	3,286
G11 x Gender (Males)	-0.08	0.92	*(0.221)*	1,421	-0.21	0.81	*(0.164)*	1,421
G12 x Gender (Males)	-0.21	0.81	*(0.255)*	1,056	-0.10	0.91	*(0.176)*	1,056

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Note. Controlling for school fixed effects.

Table 22

Weighted Logistic Regression Estimates for Graduation and AP Enrollment by Grade of Reclassification and Disability Status

	Graduation *N* = 20,703				AP Enrollment *N* = 20,703			
	β	e^{β}	SE	N	β	e^{β}	SE	N
Intercept	2.44	11.47	*(0.485)* ***	4,347	-2.09	0.12	*(0.299)* ***	4,347
Grade 6 (G6)	0.30	1.36	*(0.190)*	716	1.33	3.78	*(0.107)* ***	716
Grade 7 (G7)	1.16	3.19	*(0.205)* ***	1,342	1.48	4.39	*(0.086)* ***	1,342
Grade 8 (G8)	0.52	1.69	*(0.164)* **	1,330	1.36	3.90	*(0.088)* ***	1,330
Grade 9 (G9)	0.20	1.22	*(0.166)*	936	1.29	3.63	*(0.098)* ***	936
Grade 10 (G10)	0.38	1.47	*(0.097)* ***	6,033	0.60	1.82	*(0.067)* ***	6,033
Grade 11 (G11)	0.05	1.06	*(0.113)*	2,573	0.38	1.46	*(0.083)* ***	2,573
Grade 12 (G12)	0.16	1.17	*(0.130)*	1,851	0.52	1.67	*(0.090)* ***	1,851
Disability (Yes)	0.28	1.32	*(0.220)*	1,102	-0.65	0.52	*(0.214)* **	1,102
G6 x Disability (Yes)	13.9	1.5×10^{5}	*(1444)*	7	-15.1	2.8×10^{-7}	*(890)*	7
G7 x Disability (Yes)	13.2	5.4×10^{5}	*(935)*	17	-0.32	0.73	*(0.805)*	17
G8 x Disability (Yes)	13.8	9.8×10^{5}	*(810)*	26	-1.45	0.24	*(1.05)*	26
G9 x Disability (Yes)	14.1	1.3×10^{6}	*(1061)*	14	-0.64	0.53	*(1.07)*	14
G10 x Disability (Yes)	-0.02	0.99	*(0.424)*	274	-0.47	0.63	*(0.352)*	274
G11 x Disability (Yes)	-0.17	0.84	*(0.427)*	187	-0.36	0.70	*(0.428)*	187
G12 x Disability (Yes)	-0.36	0.70	*(0.464)*	153	-0.72	0.49	*(0.481)*	153

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Note. Controlling for school fixed effects.

Table 23

Weighted Logistic Regression Estimates for Graduation and AP Enrollment by Grade of Reclassification and Economically Disadvantaged Status (EDS)

	Graduation *N* = 20,703				AP Enrollment *N* = 20,703			
	β	e^{β}	SE	N	β	e^{β}	SE	N
Intercept	2.26	9.58	*(0.497)* ***	1,183	-1.85	0.16	*(0.308)* ***	1,183
Grade 6 (G6)	0.41	1.50	*(0.332)*	283	1.24	3.46	*(0.177)* ***	283
Grade 7 (G7)	1.76	5.81	*(0.403)* ***	484	1.23	3.42	*(0.141)* ***	484
Grade 8 (G8)	1.07	2.92	*(0.346)* **	408	1.28	3.60	*(0.148)* ***	408
Grade 9 (G9)	0.23	1.25	*(0.298)*	254	0.68	1.96	*(0.190)* ***	254
Grade 10 (G10)	0.66	1.94	*(0.179)* ***	1,648	0.54	1.71	*(0.113)* ***	1,648
Grade 11 (G11)	0.07	1.08	*(0.208)*	656	0.53	1.70	*(0.140)* ***	656
Grade 12 (G12)	0.49	1.64	*(0.266)*	473	0.58	0.15	*(0.153)* ***	473
EDS (Yes)	0.28	1.32	*(0.142)*	4,266	-0.52	0.59	*(0.110)* ***	4,266
G6 x EDS (Yes)	-0.16	0.85	*(0.401)*	485	0.17	1.18	*(0.221)*	485
G7 x EDS (Yes)	-0.87	0.42	*(0.466)*	875	0.42	1.53	*(0.176)* *	875
G8 x EDS (Yes)	-0.75	0.47	*(0.392)*	948	0.15	1.16	*(0.181)*	948
G9 x EDS (Yes)	-0.06	0.95	*(0.356)*	696	0.92	2.51	*(0.220)* ***	696
G10 x EDS (Yes)	-0.41	0.68	*(0.208)* *	4,659	0.12	1.13	*(0.138)*	4,659
G11 x EDS (Yes)	-0.06	0.94	*(0.244)*	2,104	-0.19	0.82	*(0.172)*	2,104
G12 x EDS (Yes)	-0.49	0.61	*(0.301)*	1,531	-0.08	0.92	*(0.187)*	1,531

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Note. Controlling for school fixed effects.

CHAPTER V: DISCUSSION

The purpose of this study was to understand the consequences of reclassification decisions on students' achievement outcomes and outcomes related to college readiness and graduation. I obtained student-level data from 42,393 English learners in grades five through eleven. I assessed the consequences of reclassification by first matching reclassified English learners with comparable non-reclassified English learners. I then evaluated the effects of reclassification on achievement outcomes with a difference in differences approach, and the impact on outcomes related to college readiness and graduation with linear and logistic regressions. In the current study, three main findings emerge: (a) multiple sources of evidence are useful when evaluating the consequences of reclassification decisions; (b) reclassification decisions may have negative impacts on students who remain English learners; and (c) subgroup analyses provide insight into the differential impact of reclassification decisions by student characteristics.

I divided this chapter into three sections. The first section summarizes and interprets the findings from the study, focusing on how the results further expand my understanding of the consequences of reclassification decisions. In the second section, I address implications for policymakers, researchers, test developers, and practitioners. Lastly, I discuss the limitations of the research and suggestions for future research.

Findings

Multiple Sources of Evidence

In the current study, I found positive or null effects of reclassification decisions on achievement outcomes. In particular, I found reclassification decisions did not adversely affect the performance of reclassified students relative to the performance of students who remain classified as English learners. Previous reclassification studies found null or adverse effects of reclassification on students' academic outcomes (Cimpian et al., 2017; Robinson, 2011). Robinson (2011) argues for the desirability of null effects in reclassification research. In particular, the researcher argues that reclassified students should perform similarly to students who continue to receive English language supports. Furthermore, Robinson (2011) posits that positive effects may indicate students were reclassified too late, while negative effects may mean students were reclassified too soon. However, positive effects do not necessarily indicate students were reclassified too late. Analyses of additional outcomes in which target a number of years after reclassification can provide evidence of whether students were reclassified too late. Therefore, I argue in this study that multiple sources of evidence are needed to justify whether reclassification decisions are appropriate.

In the current study, I also found reclassification decisions did not lead to adverse effects on outcomes related to college readiness and graduation for reclassified students. Evidence came from analyses on students' ACT performance, AP enrollment, and on-time graduation. The use of multiple sources of evidence helped evaluate the consequences of reclassification decisions. Multiple sources of evidence for reclassification decisions look beyond student performance after one year and contributes to a better understanding of the consequences of reclassification decisions overall. Therefore, I argue that positive effects on subsequent academic outcomes are appropriate if evidence from college readiness and graduation outcomes are also supportive of reclassification decisions.

For outcomes related to college readiness and graduation, my findings are similar to those of Carlson and Knowles (2016). Their study found students reclassified by grade ten performed better on their ACTs compared to students reclassified after grade ten. I find in North Carolina that even reclassification immediately before grade eleven can lead to significant improvements on two out of the three ACT subtests studied. Different criteria for reclassification may explain these slight differences in results. Wisconsin adopts a WIDA ACCESS score requirement of 6.0 and also requires three additional criteria for reclassification (Linquanti & Cook, 2016; WIDA, n.d.c.), while other states require a lower threshold for reclassification and fewer criteria. For North Carolina, I found results in support of reclassification decisions with only one criterion. Thereby, calling into question the necessity for more than one criterion for reclassification.

Absence of Reclassification

A contribution of the current study is the addition of an outcome variable previously not analyzed in reclassification literature, AP course enrollment, which indicates whether a student enrolled in an AP course. I found significant increases in the odds of reclassified English learners enrolling in an AP course compared to English learners never reclassified. In particular, I find that the probability an English learner enrolled in an AP course was only 10.5%, while English learners reclassified before grade seven had a 35% chance of enrollment. Previous research has only investigated the relation between reclassification and related outcomes such as graduation, ACT performance, and post-secondary enrollment (Carlson & Knowles, 2016; Robinson-Cimpian & Thompson, 2016).

I initially included the AP course enrollment variable to better understand the effects of reclassification decisions on reclassified students. In particular, I was interested in whether reclassification by certain grade levels impacted the odds a student would take an AP course. I considered AP enrollment as an indicator of access to coursework that prepares students for post-secondary coursework. AP enrollment also provided additional insight into the opportunity for advanced coursework for students who were not reclassified by grade twelve. Although reclassification before grades six through twelve related to higher odds of enrolling in an AP course as compared to students who were never reclassified, the low probability of English learners enrolling in AP courses gives concern to the lack of opportunity English learners have in accessing higher-level coursework.

ACT performance can also provide additional insights into the absence of reclassification decisions for English learners. In 2013, the average performance of all students in North Carolina on the ACT English, reading, and mathematics subtest was 17.1, 19.6, and 18.8, respectively (National Center for Education Statistics, 2017, Table 226.60). However, English learners who I matched on similar academic trajectories performed lower in these same ACT subtests. In particular, matched English learners' average performance on the ACT English, reading, and mathematics subtests was 11.8, 13.9, and 15.5, respectively. The nature of coursework for English learners may explain these differences in performance. English learners may not have been exposed to the rigorous curriculum to prepare them for assessments that measure their college readiness.

Differential Impacts of Reclassification Decisions

This study found no differential impact of reclassification decisions on subsequent achievement outcomes by subgroups (i.e., gender, ethnicity, disability status, economic status). However, even after controlling for school-level fixed effects, the study found differential impacts of reclassification decisions on outcomes related to college readiness and graduation. Table 24 highlights the differential impacts detected by grade level. Previous research has focused on the effects of reclassification decisions on the aggregated group of English learners. However, these studies assume subgroup characteristics do not differentially impact the effects of reclassification decisions. I included these subgroup analyses to understand whether reclassification decisions were equitable for all students, regardless of subgroup characteristics.

Table 24

Summary of Grades with Differential Impact of Subgroup Status on Graduation and College Readiness Outcomes

	Gender	Ethnicity	SWD	EDS
ACT English	-	8th, 9th	-	-
ACT Reading	-	9th, 11th	-	6th, 7th,8th,10th
ACT Math	-	8th,9th,10th	6th, 8th	-
Graduation	-	8th	-	10th
AP Enrollment	-	12th	-	7th, 9th

Note. SWD = Students with Disability; EDS = Economically Disadvantaged Status.

The differential impacts detected on outcomes related to college readiness and graduation in the current study involved ethnicity, disability status, and economically disadvantaged status. However, uncertainty surrounds the detection of differential impacts at certain grades and not others. For example, why did ethnicity moderate the relation between reclassification in grades six and eight for the ACT English and not grade seven? In some instances, the direction of the differential impact was counter-intuitive to the expected trend. For example, the interaction effect for grades seven and nine was positive for AP enrollment, indicating differential impact in favor of economically disadvantaged students. Furthermore, was sample size a contributing factor to not detecting additional significant effects by subgroups? For example, the AP enrollment by disability status analysis had sample sizes as little as seven (see Table 22). These differential impacts shed light on the need to make equitable reclassification decisions for all students, regardless of student characteristics. It also highlights the need for more research into why the differential impacts are emerging. In the next section, I discuss the implications of these finding for policymakers, researchers, test developers, and practitioners.

Implications

Implications for Policymakers

The current study found reclassification decisions for English learners in North Carolina were mostly appropriate for students who were reclassified. Furthermore, the current test-based criteria for reclassification appear appropriate within the context of North Carolina. For policymakers, these results should provide relief about how English learners perform after reclassification.

These results may insinuate that reclassification itself provided the positive results and states should lower the threshold to reclassify more students. However, I caution against lowering the threshold to reclassify more students without the evidence to support that a lower threshold does not adversely affect reclassified students. No study to date has investigated the effects of lowering the threshold for reclassification decisions. Robinson-Cimpian and Thompson (2016) investigated the adoption of stricter criteria in California and found that changing the criteria can have a positive effect on student outcomes. If policymakers want to consider changes to their criteria for reclassification, then studies should be conducted on the potential adverse effects of these changes on students' outcomes.

This study also provides policymakers additional context for reclassification decisions based solely on an objective English language proficiency measure. Only two previously conducted studies were in state contexts where the sole determining factor for reclassification was the English language proficiency assessment (Ardasheva et al., 2012; Cimpian et al., 2017). The Standards (2014) argue that high-stakes decisions, such as reclassification, should consider more than one criterion due to their potential negative consequences on student outcomes. However, most states only make reclassification decisions from the English language proficiency assessment (Linquanti & Cook, 2015). Furthermore, variability exists in the application of assessments for reclassification decisions. For example, states that adopt the WIDA ACCESS for English language proficiency testing adopt scores between 4.5 and 6.0 for reclassification decisions (see Table 1; WIDA, n.d.c.). With such variability in state implementation of reclassification criteria, policymakers should evaluate reclassification decisions within the context of their state to understand if their criteria are appropriate for reclassification.

Implications for Researchers

I situated the current research under the argument-based validity framework to understand the consequences of using WIDA ACCESS scores to make reclassification decisions. The argument-based approach to validation (Cronbach, 1988; Kane, 1992; Messick, 1989) highlights the need to understand the consequences of intended and unintended test uses. The findings in the current study contribute a deeper understanding of the consequences of using an English language proficiency assessment for reclassification decisions. However, North Carolina is not the only state where reclassification decisions are based solely on an English language proficiency assessment. Twenty-nine states and the District of Columbia make reclassification decisions exclusively based upon performance on the English language proficiency assessment (Linquanti & Cook, 2015). Therefore, researchers could adopt a similar framework when investigating reclassification decisions in these states.

The Standards (2014) highlight the need to understand the consequences of intended and unintended test uses with multiple sources of evidence but provide little guidance on the kinds of evidence to collect. Some researchers have proposed program evaluation as a theory of action to understand these consequences (Cizek, 2016; Lane, 2014). Cizek (2016) has even gone further to differentiate between validation of an intended score inference and justifying test score uses. The current study provides an example of the evaluation of multiple sources (i.e., multiple student outcomes) to evaluate the appropriateness of reclassification decisions (i.e., score use).

Researchers that adopt a consequential validity framework to investigate reclassification decisions provide additional context of the effects of these decisions outside of the context of just a particular policy, which can inform a broader range of stakeholders (e.g., test developers). For example, the results from the current study help inform not only the policymakers who adopted the criteria for reclassification in North Carolina but also the test developers whose tests are used to make reclassification decisions. In the next section, I discuss a few of the implications for test developers of English language proficiency assessments.

Implications for Test Developers

ESSA (2015) recently increased the amount of monitoring required of not only the progress English learners make towards obtaining English language proficiency but also increased the monitoring of students after reclassification. These increased levels of monitoring should make test developers take pause due to the heavy reliance of reclassification decisions based on their assessments. If states determine that their criteria for reclassification was inappropriate and lead to adverse effects on student outcomes, states may question whether the assessment is appropriate for reclassification decisions and not the criteria set by states.

Test developers need to provide clear guidance defining English language proficiency on their respective assessments or provide guidance for how states can investigate the appropriateness of reclassification decisions with their assessments. A critical aspect of the argument-based approach to validity is understanding the consequences, intended or unintended, of assessment uses (Kane, 2013). Test developers that ignore their assessments' role in high-stakes decisions risk misuse among test users, which may raise questions over the validity of the intended use of the test.

Implications for Practitioners

Implications also extend to educator practices. Teachers should consider how reclassification decisions affect English learners, not just reclassified English learners. For example, the current study found English learners had little access to advanced courses and were not as prepared for the ACTs. The low probability of enrollment could be attributed to the required coursework to support English language development, while the low performance on the ACTs may be reflective of the lack of exposure to rigorous content to prepare English learners for the ACTs. In either case, English learners were not exposed to sufficiently rigorous content, and educators can assist by improving English learners' likelihood of reclassification.

To improve English learners' likelihood of reclassification, practitioners need to understand the complex nuances of English language development. For example, the stages of English language development and the influence of proficiency in their native language can help educators improve instructional practices for English learners (Lightbown & Spada, 2013). Furthermore, educators need to understand their students' current level of English language proficiency and how to better instruct English learners at particular levels of English proficiency. This type of individualized instruction may expedite the time it takes English learners to reach proficiency and help English learners access more rigorous coursework.

Limitations

Research on existing data can raise numerous constraints, including generalizability and causality (Murnane & Willett, 2008). The current study only utilized data from North Carolina, which limits generalizability to other states. Furthermore, matching students based on previous academic performance dropped students who were too dissimilar. For example, a student who was reclassified and historically performed well in mathematics was likely dropped from the math analyses because there was not a student in the English learner group of similar mathematical abilities. Table 25 provides an overview of the count and percentage of the population that was dropped by outcome and subgroup characteristic. In some instances, the percentage dropped was as high as 83.4% and as low as 38.3%. Therefore, generalizability should be cautioned and not applied to all students within the English learner and reclassified English learner groups.

Table 25

Count (%) Dropped by Outcome and Subgroup Characteristic

	ELA	Math	ACT Eng/Read	ACT Math	Graduation/AP Enrollment
Ethnicity					
Hispanic	10,478 (60.4%)	9,654 (55.7%)	11,056 (63.7%)	12,214 (70.4%)	17,714 (51.8%)
Non-Hispanic	2,060 (56.6%)	1,891 (52.0%)	2,373 (65.2%)	2,686 (73.8%)	3,771 (46.0%)
Gender					
Female	5,591 (57.1%)	6,304 (56.3%)	6,067 (61.9%)	6,715 (68.6%)	9,394 (47.7%)
Male	6,947 (62.1%)	5,241 (53.5%)	7,362 (65.8%)	8,185 (73.2%)	12,091 (53.3%)
Disability					
Yes	2,105 (83.4%)	1,947 (77.1%)	1,616 (64.0%)	1,747 (69.2%)	2,786 (61.0%)
No	10,433 (56.5%)	9,598 (52.0%)	12,713 (68.9%)	13,153 (71.3%)	18,699 (49.4%)
Economically Disadvantaged					
Yes	10,623 (60.4%)	9,742 (55.4%)	11,541 (65.6%)	12,743 (72.5%)	18,171 (53.9%)
No	1,915 (56.3%)	1,803 (53.0%)	1,888 (55.5%)	2,157 (63.4%)	3,314 (38.3%)

Causality is an additional limitation of the current study. Every attempt was made to isolate the effects of reclassification. For example, coarsened exact matching was adopted with achievement outcomes to establish comparable treatment (reclassified) and control (English learner) groups. Furthermore, a difference in differences approach was adopted to control for the effect that would have occurred had students reclassified remained classified as English learners. Lastly, I included student-, school-, and district-level fixed effects where possible to control for these sources of shared variability. Previous studies (Ardasheva et al., 2012; Kim & Herman, 2010, 2012; Hill et al., 2014) included random effects for district characteristics. I chose fixed effects instead of random effects in the current study due to strong assumption that random effects are uncorrelated with the independent variables of interest (Murnane & Willett, 2011). Fixed effects do not require that assumption but do not provide insight into nature of observed district covariates, such as district SES. Furthermore, I recognize that additional variables not controlled for within the models, such as school programming (e.g., English as a second language, bilingual), could influence the outcome variables and therefore, limits my ability to attribute the results to reclassification decisions solely.

Insufficient power is also a limitation of the current study. Power is the probability of detecting significant effects (Murnane & Willett, 2011). Power increases with higher sample sizes and in some analyses by subgroup, the sample was as low as four students. This low sample representation may be a limitation of the matching method. Certain subgroups of students historically perform lower (e.g., students with disabilities). Therefore, the matching method dropped students from these subgroups because matches could not be found.

Future Research

Future studies should investigate variables not included in the current study. Examples of these variables include attendance data, discipline data, and assessed outcomes other than English language arts and mathematics (e.g., science performance). Analyses of these variables can provide additional evidence in support of reclassification decisions, especially when analyzing the appropriateness of reclassification decisions for grades without standardized assessments (K-2).

Studies should also be conducted to investigate the low probability of English learners enrolling in advanced coursework (i.e., AP courses). These investigations should consider course enrollment policies for English learners and whether students are enrolled in classes to support English language development or if other factors such as the English learner stigma curtails access to advanced coursework (Dabach, 2014).

Future research should also address anomalies in the rates of reclassification found in the current study. For example, the study found a higher proportion of English learners reclassified before tenth grade as compared to any other grade level. State policies did not provide context for these higher proportions. Qualitative research on district reclassification policies may provide context for this anomaly. For example, interviews with school and district personnel could contribute additional insight into whether teachers are encouraged to adopt strategies to exit students out of services before grade ten (e.g., teaching to the test).

Furthermore, additional research in different state context is warranted. Reclassification research has been limited to only three known states (some states in previous studies were kept anonymous). Increases in the number of studies will not only provide evidence to evaluate the consequences of state's reclassification decisions but also increase the knowledge base on how to evaluate these decisions and better situate reclassification decisions within a particular theoretical framework.

Conclusion

As mandated by federal law, every state must annually assess the progress English learners make towards English proficiency. Once English learners reach specific criteria to be reclassified, English language supports are removed. The purpose of this study was to understand the consequences of reclassification decisions in North Carolina. In particular, this study sought to understand the consequences of reclassification decisions when these decisions are solely based upon an English language proficiency assessment. Reclassification decisions could adversely affect outcomes for English learners if they are inappropriately applied. In particular, English learners reclassified too soon may lack the English proficiency necessary to perform adequately without English language supports. Conversely, English learners reclassified too late may miss opportunities to participate in higher-level coursework. Therefore, it is essential to understand the impacts of these decisions.

This study analyzed multiple outcomes to assess the impacts of reclassification decisions critically. Furthermore, the study adopted a matching technique to match reclassified English learners with comparable English learners. Then, regression analyses were conducted to compare English learners with reclassified English learners.

Results from this study suggest reclassification decisions are appropriate for reclassified English learners in North Carolina. In particular, reclassification decisions do not adversely affect reclassified English learners. However, this study also found English learners who never reached the criteria for reclassification perform lower on their ACTs and have a low probability of enrolling in AP courses. Furthermore, this study found differential impacts of reclassification decisions by subgroup characteristics. More research is needed to understand the differential impacts of reclassification decisions and the impacts of never reaching the criteria for reclassification. The decision to reclassify students out of English learner supports applies to every state. Therefore, every state should investigate the appropriateness of their reclassification decisions to ensure English learners are not adversely affected by these decisions.

References

ACT (n.d.). The ACT Test: US Students. Retrieved from http://www.act.org/content/act/en/products-and-services/the-act.html

American Educational Research Association (AERA), American Psychological Association (APA), & National Council on Measurement in Education (NCME). (2014). *Standards for educational and psychological testing.* Washington, DC: American Psychological Association.

Ardasheva, Y., Tretter, T. R., & Kinny, M. (2012). English language learners and academic achievement: Revisiting the Threshold Hypothesis. *Language Learning, 62*(3), 769-812.

Bunch, M. B. (2011). Testing English language learners under No Child Left Behind. *Language Testing, 28*(3), 323-341.

Burke, A. P., Morita-Mullaney, T., Singh, M. (2016). Indiana emergent bilingual student time to reclassification: A survival analysis. *American Educational Research Journal, 53*(5), 1310-1342.

Carlson, D., & Knowles, J. E. (2016). The effect of English language learner reclassification on student ACT scores, high school graduation, and postsecondary enrollment: Regression discontinuity evidence from Wisconsin. *Journal of Policy Analysis and Management, 35*(3), 559-586.

Center for Applied Linguistics (2017). *Annual technical report for ACCESS for ELLs® English Language Proficiency Test, Series 400, 2015–2016 administration* (WIDA Consortium Annual Technical Report No. 12B). Retrieved from https://www.wida.us/get.aspx?id=2136

Cimpian, J. R., Thompson, K. D., & Makowski, M. B. (2017). Evaluating English learner reclassification policy effects across districts. *American Education Research Journal, 54*(1S), 255S-278S.

Cizek, G. J. (2016). Validating test score meaning and defending test score use: Different aims, different methods. *Assessment in Education: Principles, Policy & Practice, 23*(2), 212-225.

Clark, P., Crawford, C., Steele, F., & Vignoles, A. (2010). The choice between fixed and random effects models: Some considerations for educational research. Institute of Education DoQSS Working Paper No. 10-10.

Croissant, Y., & Millo, G. (2008). Panel data econometrics in R: The plm package. *Journal of Statistical Software, 27*(2), 1-43.

Cronbach, L. J. (1988). Five perspectives on validity argument. In H. Wainer & H. Braun (Eds.), *Test validity* (pp. 3-17). Hillsdale, NJ: Erlbaum.

Cummins, J. (1980). The cross-lingual dimensions of language proficiency: Implications for bilingual education and the optimal age issue. *Applied Linguistics, 2*(2), 175-187.

Dabach, D. B. (2014). "I am Not a Shelter!": Stigma and social boundaries in teachers' accounts of students' experience in separate "Sheltered" English learner classrooms. *Journal of Education for Students Placed at Risk, 19*(2). 98-124.

Duke Center for Child and Family Policy. (n.d.). North Carolina Education Research Data Center. Retrieved from https://childandfamilypolicy.duke.edu/research/nc-education-data-center/

ESSA (2015). Every Student Succeeds Art of 2015, Pub. L. No. 114-95 §114 Stat. 1177 (2015-2016).

Gándara, P. (2015). Charting the relationship of English learners and the ESEA: One step forward, two steps back. *The Russel Sage Foundation Journal of Social Sciences, 1*(3), 112-128.

Gándara, P. & Rumberger, R. W. (2009). Immigration, language, and education: How does language policy structure opportunity? *Teachers College Record, 111*(3), 750-782.

Greenberg Motamedi, J., Singh, M., & Thompson, K. D. (2016). *English learner student characteristics and time to reclassification: An example from Washington State* (REL 2016-128). Washington, DC: U.S. Department of Education, Institute of Education Science, National Center for Education Evaluation and Regional Assistance, Regional Educational Laboratory Northwest. Retrieved from http://ies.ed.gov/ncee/edlabs

Haertel, E. (2013). Getting the help we need. *Journal of Educational Measurement, 50*(1), 84-90.

Hill, L. E., Weston, M., & Hayes, M. (2014). *Reclassification of English learner students in California.* San Francisco: Public Policy Institute of California.

Ho, D. E., Kosuke, I., King, G., & Stuart, E. A. (2011). MatchIt: Nonparametric preprocessing for parametric causal inference. Journal of Statistical Software, *42*(8), 1-28.

Iacus, S. M., King, G., & Porro, G. (2009). cem: Software for coarsened exact matching. *Journal of Statistical Software, 30*(9), 1-27.

Iacus, S. M., King, G., Porro, G. (2011). Causal inference without balance checking: Coarsened exact matching. *Political Analysis, 20*(1), 1-24.

Iacus, S. M., King, G., & Porro, G. (2018). CEM: Coarsened exact matching. R package version 1.1.19. https://CRAN.R-project.org/package=cem

Kane, M. (1992). An argument-based approach to validation. *Psychological Bulletin, 112*(3), 527-535.

Kane, M. (2013). The argument-based approach to validation. *School Psychology Review, 42*(4), 448-457.

Kim, J., & Herman, J. L. (2010). *When to exit ELL students: monitoring subsequent success and failure in mainstream classrooms after ELLs' reclassification.* (CRESST Report 779). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Kim, J., & Herman, J. L. (2012). *Understanding patterns and precursors of ELL success subsequent to reclassification* (CRESST Report 818). Los Angeles, CA: University of California, National Center for Research on Evaluation, Standards, and Student Testing (CRESST).

Lane, S. (2014). Validity evidence based on testing consequences. *Psicothema, 26*(1), 127-135.

Lightbown, P. M., & Spada, N. (2013). *How languages are learned* (4th edition). Oxford: Oxford university press.

Linquanti, R., & Cook, H. G. (2015). *The reclassification dilemma: Guidance from a national working session on policies and practices for exiting students from English learner status.* Washington, DC: Council of Chief State School Officers.

Massachusetts Department of Elementary and Secondary Education (2017). *ACCESS for ELLs.* Retrieved from http://www.doe.mass.edu/mcas/access/

Messick, S. (1975). The standard problem: Meaning and values in measurement and evaluation, *American Psychologist, 30*, 955-966.

Messick, S. (1989). Meaning and values in test validation: The science and ethics of assessment. *Educational Researcher, 18*(2), 5-11.

Murnane, R. J., & Willett, J. B. (2011). *Methods matter: Improving causal inference in educational and social science research.* New York, NY: Oxford University Press.

No Child Left Behind Act of 2001. P.L. 107-110, 20 U.S.C. § 6319 (2002).

North Carolina Department of Public Instruction. (2018). WIDA ACCESS for ELLs 2.0: North Carolina policy and procedure guidance for annual testing 2018-19. Retrieved from http://www.ncpublicschools.org/docs/accountability/policyoperations/lep/accessgde2018-19.pdf

North Carolina Department of Public Instruction (n.d.). North Carolina End-of-Grade Tests at grades 3-8. Retrieved from http://www.ncpublicschools.org/accountability/testing/eog/

North Carolina Department of Public Instruction (n.d.b). Accountability services. Retrieved from http://www.ncpublicschools.org/accountability/

Rencher, A. C., & Christensen, W. F. (2012). *Methods of multivariate analysis.* Hoboken, NJ: John Wiley & Sons, Inc.

Robinson, J. P. (2011). Evaluating criteria for English learner reclassification: A causal-effects approach using a binding score regression discontinuity design with instrumental variables. *Educational Evaluation and Policy Analysis, 33*(3), 267-292.

Robinson-Cimpian, J. P., & Thompson, K. D. (2016). The effects of changing test-based policies for reclassifying English learners. *Journal of Policy Analysis and Management, 35*(2), 279-305.

Singer, J. D., & Willett, J. B. (2003). *Applied longitudinal data analysis: Modeling change and event occurrences.* New York, NY: Oxford University Press.

Shadish, W. R., Clark, M. H., & Steiner, P. M. (2008). Can nonrandomized experiments yield accurate answers? A randomized experiment comparing random and nonrandom assignments. *Journal of the American Statistical Association, 103*(484), 1334-1356.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (2nd ed.). Allyn & Bacon/Pearson Education.

U.S. Department of Education, National Center for Education Statistics. (2017). Table 204.20: English language learner (ELL) students enrolled in public elementary and secondary schools, by state: Selected years, fall 2000 through fall 2015. In U.S. Department of Education, National Center for Education Statistics (Ed.), *Digest of Education Statistics* (2017 ed.). Retrieved from https://nces.ed.gov/programs/digest/current_tables.asp

U.S. Department of Education, National Center for Education Statistics. (2017). Table 219.46: Public high school 4-year adjusted cohort graduation rate (ACGR), by selected student characteristics and state: 2010-2011 through 2015-16. In U.S. Department of Education, National Center for Education Statistics (Ed.), *Digest of Education Statistics* (2017 ed.). Retrieved from https://nces.ed.gov/programs/digest/current_tables.asp

U.S. Department of Education, National Center for Education Statistics. (2017). Table 226.60: Average ACT Score and percentage of graduates taking the ACT, by state: 2013 and 2017. In U.S. Department of Education, National Center for Education Statistics (Ed.), *Digest of Education Statistics* (2017 ed.). Retrieved from https://nces.ed.gov/programs/digest/d17/tables/dt17_226.60.asp?current=yes

U.S. Department of Education, Institute of Education Sciences, National Center for Education Statistics, National Assessment of Educational Progress (NAEP). (2017b). NAEP data explorer. Retrieved from https://www.nationsreportcard.gov/ndecore/landing

U.S. Department of Education, Office of English Language Acquisition (OELA). (2015). Dual language program: Current state policies and practices. Washington, D.C. Retrieved from http://www2.ed.gov/about/offices/list/oela/resources.html

World-Class Instructional Design and Assessment [WIDA]. (2012). *The WIDA standards framework and its theoretical foundations.* Retrieved from https://www.wida.us/standards/eld.aspx#2012

WIDA. (2016). *Assessing Comprehension and Communication in English state to state for English Language Learners*. Retrieved from https://www.wida.us/assessment/access%202.0/.../2016TestAdministratorManual.pdf

WIDA. (n.d.). *W-APT.* Retrieved from https://www.wida.us/assessment/w-apt/

WIDA. (n.d.b). *ACCESS for ELLs scores and reports.* Retrieved from https://wida.wisc.edu/assess/access/scores-reports

WIDA. (n.d.c). *Consortium members.* Retrieved from https://www.wida.us/membership/states/index.aspx

Appendix A: IRB Documentation and Security Plan

From: IRB Committee

To: Anthony Sparks

Date: March 8, 2018

Re: IRB Expedited New submission approval; Protocol # H17-138-SPAA - Bilingual Programming and Reclassification

Dear Mr. Sparks,

The IRB Committee completed review of your application and granted approval of your protocol on 12/05/2017. This approval is valid until **12/05/2018**. If work will continue beyond this date, it is the responsibility of the principal investigator to submit an annual review of progress (CFR 21 §56.109(f)). Failure to gain approval of this annual review prior to the expiration date could result in suspension of the work covered under this protocol. This suspension of work would include halting all subject enrollment, collecting data, and/or analyzing previously collected, identified data.

Any proposed changes in the protocol should be submitted to the IRB as an amendment prior to initiation (CFR 21 §56.108 (a)(3); §56.108 (a)(4)). Please be advised that as the principal investigator, you are required to report unanticipated adverse events to the Office of Research Administration within 24 hours of the occurrence or upon acknowledgement of the occurrence (CFR 21 § 56.108 (b)(1)).

All investigators and key personnel identified in the protocol must have documented IRB CITI or NIH Training on file with this office. The certification will expire in 3 years, so please plan your renewal accordingly. For NIH training only, please include a copy of your certificate with your submission.

Southern Methodist University's Office of Research and Graduate Studies appreciates your continued commitment to the protection of human subjects in research. Should you have questions, or need to report completion of study procedures, please contact the Office of Research compliance at 214-768-2033 or at researchcompliance@smu.edu.

Thank You,

Austin Baldwin
IRB Chair

Office of Research and Graduate Studies
Southern Methodist University PO Box 750302 Dallas TX 75275-0240 Office: 214-768-2033 Fax: 214-768-1079

SMU Data Security Plan

Policy Requirements

The project researchers will connect to a NCERDC data folder through a secure file server housed in a secure datacenter on the Southern Methodist University campus. All data will be viewed and modified on the server over an encrypted network connection.

ALL storage and analysis of NCERDC data will take place exclusively on the secure server. Data may not be downloaded to local workstations, or to any external devices, including laptops. Desktop and laptop workstations may be used only for remote access to the secure server.

Portable storage devices, including laptops, will not be used for downloading or storing data.

NCERDC data will NOT be shared with any other institution or any investigator not currently listed in the data use agreement. This restriction applies to source data as well as all derived data files. Project investigators, including the PI, do not have discretion to modify access to the NCERDC data. Any changes in access to the data on the secure server require explicit prior approval by the NCERDC.

All data security protections apply to the original NCERDC data, derived files, and temporary analysis files.

Technical Details

LOCATION

The computing platform is located at the SMU Primary Managed Data Center. Physical Access is provided to IT Technical Staff only through multiple levels of ID-secured and monitored locked access doors with video surveillance recording.

COMPUTING PLATFORM

SMU Systems Infrastructure provides platform services from a shared virtualization platform, with access controls to enforce resource separation. Users connect to the server from authorized campus managed desktop clients with enforced security controls. User authenticate with their Active Directory-based campus username and password.

SECURITY SYSTEMS

SMU Systems Infrastructure enforces secure transport protocols and secure firewalled VLANs for datacenter network subsystems. User connections to datacenter servers pass through firewalls, require strong encryption protocols, and are only allowed for authorized University IP addresses. User accounts are policy-managed, with enforced complexity, age, rotation, and other identity management best practices. Desktop clients feature security and system management agents which further secure the endpoint. Access to data on the server is strictly controlled via user ACLs based on the users' identity in AD. No unencrypted copies of the system data are permitted for backups or any other purpose. Only the designated researchers and IT system administrators will have access to the folder with the NCERDC data.

TIMELINE FOR DATA USE

These data would be under active analysis through June 30, 2019, but would be stored for up to five years. The data will be destroyed by December 31, 2021.

Appendix B: List of Variables

Student Level

- ACT
  - Composite Score
  - English Score
  - Grade
  - Math Score
  - Science Score
- Advanced Placement
  - AP Course Code
  - AP Course Grade
- Demographics/Absences
  - District
  - AIG
  - Days Absent
  - Grade
  - Gender
  - English learner status
  - Ethnicity
  - 504 ID
  - Accommodation
- EOC/EOG Tests
  - Achievement Score
  - Percentile
  - Score
  - Scale Score
  - Ethnicity
- Graduates
  - Graduation Classification
  - Plans after high school

Appendix C: Analytic Models

Research Question 1

$$ELA_{it} = \alpha_i + \beta_1(YEAR\_POST_t) + \beta_2(YEAR\_POST_t \times RECLASS_i) + \gamma_i + \delta_i + \boldsymbol{C} + \varepsilon_i$$

$$MATH_{it} = \alpha_i + \beta_1(YEAR\_POST_t) + \beta_2(YEAR\_POST_t \times RECLASS_i) + \gamma_i + \delta_i + \boldsymbol{C} + \varepsilon_i$$

Research Question 2

$$ACT\_English_i = \beta_0 + \beta_1(iSIXTH_i) + \cdots + \beta_6(iTWELFTH_i) + \beta_3(iSIXTH_i \times GENDER_i) + \cdots + \gamma_i + \delta_i + \varepsilon_i$$

$$ACT\_Read_i = \beta_0 + \beta_1(iSIXTH_i) + \cdots + \beta_6(iTWELFTH_i) + \beta_3(iSIXTH_i \times GENDER_i) + \cdots + \gamma_i + \delta_i + \varepsilon_i$$

$$ACT\_Math_i = \beta_0 + \beta_1(iSIXTH_i) + \cdots + \beta_6(iTWELFTH_i) + \beta_3(iSIXTH_i \times GENDER_i) + \cdots + \gamma_i + \delta_i + \varepsilon_i$$

$$logit(Grad_i) = \beta_0 + \beta_1(iSIXTH_i) + \cdots + \beta_6(iTWELFTH_i) + \beta_3(iSIXTH_i \times GENDER_i) + \cdots + \gamma_i + \boldsymbol{C} + \varepsilon_i$$

$$logit(AP_i) = \beta_0 + \beta_1(iSIXTH_i) + \cdots + \beta_6(iTWELFTH_i) + \beta_3(iSIXTH_i \times GENDER_i) + \cdots + \gamma_i + \boldsymbol{C} + \varepsilon_i$$

Appendix D: Full Regression Results for Outcomes Related to College Readiness and Graduation

Full Weighted Least Squares Regression Results with Select ACT Subtests

	ACT English		ACT Reading		ACT Math	
	β	SE	β	SE	β	SE
Intercept	11.8	*(0.758)* ***	13.0	*(0.821)* ***	15.4	*(0.651)* ***
Grade 6	3.83	*(0.509)* ***	5.27	*(0.552)* ***	1.20	*(0.352)* ***
Grade 7	4.16	*(0.435)* ***	4.98	*(0.471)* ***	1.57	*(0.301)* ***
Grade 8	3.17	*(0.446)* ***	4.05	*(0.484)* ***	1.23	*(0.306)* ***
Grade 9	3.23	*(0.526)* ***	3.52	*(0.569)* ***	0.99	*(0.398)* *
Grade 10	1.58	*(0.348)* ***	2.39	*(0.377)* ***	0.36	*(0.231)*
Grade 11	0.12	*(0.444)*	0.63	*(0.481)*	-0.05	*(0.317)*
Ethnicity (Non-Hisp)	1.10	*(0.293)* ***	1.24	*(0.317)* ***	0.77	*(0.196)* ***
Gender (Males)	-0.52	*(0.216)* *	-0.43	*(0.234)*	0.38	*(0.144)* **
SWD (Yes)	-0.42	*(0.288)*	-0.11	*(0.313)*	-0.65	*(0.161)* ***
EDS (Yes)	-0.06	*(0.278)*	0.94	*(0.301)* **	-0.17	*(0.172)*
Grade 6						
Ethnicity (Non-Hisp)	1.01	*(0.571)*	-0.17	*(0.618)*	0.54	*(0.429)*
Gender (Male)	0.29	*(0.449)*	-0.79	*(0.487)*	0.44	*(0.334)*
SWD (Yes)	0.23	*(1.53)*	-2.16	*(1.66)*	-3.53	*(1.13)* **
EDS (Yes)	0.09	*(0.507)*	-1.13	*(0.550)* *	-0.14	*(0.356)*
Grade 7						
Ethnicity (Non-Hisp)	0.750	*(0.446)*	-0.02	*(0.483)*	0.02	*(0.335)*
Gender (Male)	-0.41	*(0.336)*	-0.58	*(0.364)*	-0.28	*(0.244)*
SWD (Yes)	0.90	*(1.17)*	0.91	*(1.26)*	-0.36	*(0.835)*
EDS (Yes)	-0.22	*(0.414)*	-1.29	*(0.449)* **	-0.003	*(0.288)*
Grade 8						
Ethnicity (Non-Hisp)	1.37	*(0.453)* **	0.33	*(0.491)*	1.20	*(0.338)* ***
Gender (Male)	0.58	*(0.346)*	0.36	*(0.375)*	0.15	*(0.250)*
SWD (Yes)	0.14	*(0.834)*	0.10	*(0.904)*	-1.13	*(0.628)*
EDS (Yes)	-0.36	*(0.432)*	-1.25	*(0.468)* **	0.07	*(0.299)*
Grade 9						
Ethnicity (Non-Hisp)	1.34	*(0.557)* *	1.19	*(0.603)* *	0.89	*(0.410)* *
Gender (Male)	0.75	*(0.444)*	-0.20	*(0.482)*	-0.09	*(0.328)*
SWD (Yes)	-0.13	*(1.32)*	1.33	*(1.43)*	-0.20	*(1.03)*
EDS (Yes)	-0.90	*(0.529)*	-0.93	*(0.573)*	0.24	*(0.395)*
Grade 10						
Ethnicity (Non-Hisp)	0.02	*(0.339)*	0.02	*(0.368)*	0.59	*(0.233)* *
Gender (Male)	0.33	*(0.254)*	-0.06	*(0.275)*	0.02	*(0.175)*
SWD (Yes)	-0.72	*(0.394)*	-0.68	*(0.427)*	-0.15	*(0.264)*
EDS (Yes)	-0.04	*(0.324)*	-0.88	*(0.351)* *	0.25	*(0.213)*
Grade 11						
Ethnicity (Non-Hisp)	-0.47	*(0.457)*	-1.06	*(0.495)* *	-0.17	*(0.328)*
Gender (Male)	0.34	*(0.330)*	0.54	*(0.358)*	-0.19	*(0.239)*

SWD (Yes)	0.15	*(0.484)*	0.54	*(0.525)*	0.49	*(0.337)*
EDS (Yes)	0.20	*(0.426)*	-0.07	*(0.461)*	0.360	*(0.300)*

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$
Note. Controlling for school and district fixed effects.

Weighted Logistic Interaction Results of Reclassification Grade by Subgroup Characteristics

	Graduation			AP Enrollment		
	β	SE	e^{β}	β	SE	e^{β}
Intercept	2.24	*(0.507)* ***	9.39	-2.05	(0.327) ***	0.13
Grade 6	0.42	*(0.391)*	1.52	1.28	(0.225) ***	3.60
Grade 7	1.64	*(0.458)* ***	5.16	1.33	(0.188) ***	3.78
Grade 8	1.34	*(0.403)* ***	3.82	1.24	(0.192) ***	3.46
Grade 9	0.05	*(0.351)*	1.05	0.54	(0.238) *	1.71
Grade 10	0.73	*(0.220)* ***	2.08	0.58	(0.153) ***	1.78
Grade 11	0.21	*(0.258)*	1.23	0.53	(0.189) **	1.69
Grade 12	0.56	*(0.323)*	1.75	0.38	(0.214)	1.46
Ethnicity (Non-Hisp)	0.42	*(0.160)* **	1.53	0.81	(0.112) ***	2.24
Gender (Male)	-0.25	*(0.133)*	0.78	-1.84	(0.107)	0.16
SWD (Yes)	0.37	*(0.221)*	1.44	-5.17	(0.216) *	0.01
EDS (Yes)	0.33	*(0.143)* *	1.39	-0.43	(0.113) ***	0.65
Grade 6						
Ethnicity (Non-Hisp)	-0.34	*(0.498)*	0.71	-0.11	*(0.248)*	0.89
Gender (Males)	0.26	*(0.382)*	1.30	0.07	*(0.216)*	1.24
SWD (Yes)	13.8	*(1446)*	$9.8x10^{5}$	-15.1	*(873.5)*	$2.8x10^{-7}$
EDS (Yes)	-0.22	*(0.409)*	0.81	0.17	*(0.226)*	1.18
Grade 7						
Ethnicity (Non-Hisp)	0.70	*(0.777)*	2.01	-0.01	*(0.195)*	0.99
Gender (Males)	0.17	*(0.413)*	1.18	-0.14	*(0.173)*	0.87
SWD (Yes)	13.1	*(927.7)*	$4.9x10^{5}$	-0.35	*(0.810)*	0.71
EDS (Yes)	-0.86	*(0.472)*	0.43	0.41	*(0.182)* *	1.51
Grade 8						
Ethnicity (Non-Hisp)	-0.94	*(0.376)* *	0.39	0.09	*(0.196)*	1.09
Gender (Males)	0.08	*(0.330)*	1.08	0.02	*(0.176)*	1.02
SWD (Yes)	13.7	*(807.3)*	$8.9x10^{5}$	-1.50	*(1.07)*	0.22
EDS (Yes)	-0.89	*(0.395)* *	0.41	0.20	*(0.188)*	1.22
Grade 9						
Ethnicity (Non-Hisp)	0.05	*(0.426)*	1.05	0.18	*(0.218)*	1.19
Gender (Males)	0.35	*(0.341)*	1.42	0.01	*(0.199)*	1.01
SWD (Yes)	14.1	*(1063)*	$1.3x10^{6}$	-0.59	*(1.08)*	0.55
EDS (Yes)	-0.02	*(0.364)*	0.98	1.02	*(0.231)* ***	2.77
Grade 10						
Ethnicity (Non-Hisp)	-0.01	*(0.250)*	0.99	0.004	*(0.141)*	1.00
Gender (Males)	-0.05	*(0.189)*	0.95	-0.06	*(0.133)*	0.95
SWD (Yes)	-0.05	*(0.425)*	0.95	-0.53	*(0.354)*	0.59
EDS (Yes)	-0.40	*(0.211)*	0.67	0.15	*(0.142)*	1.16

Grade 11							
Ethnicity (Non-Hisp)	-0.28	*(0.273)*	0.75	0.21	*(0.174)*		1.23
Gender (Males)	-0.06	*(0.221)*	0.94	-0.23	*(0.167)*		0.80
SWD (Yes)	-0.23	*(0.431)*	0.79	-0.37	*(0.433)*		0.69
EDS (Yes)	-0.10	*(0.248)*	0.91	-0.11	*(0.178)*		0.90
Grade 12							
Ethnicity (Non-Hisp)	0.33	*(0.353)*	1.40	0.48	*(0.185)*	**	1.62
Gender (Males)	-0.21	*(0.257)*	0.81	-0.13	*(0.182)*		0.88
SWD (Yes)	-0.30	*(0.467)*	0.74	-0.67	*(0.496)*		0.51
EDS (Yes)	-0.44	*(0.306)*	0.64	0.001	*(0.195)*		1.01

*** $p < 0.001$, ** $p < 0.01$, * $p < 0.05$

Note. Controlling for school fixed effects.