Duke University

Undergraduate Honors Thesis

Rural and At-Risk: Predictors of and Methods to Address High School Dropout for Rising Ninth Graders in Rural North Carolina

Author:
Emily Hadley

Supervisors:
Dr. Helen Ladd
Dr. Jerome Reiter

A thesis submitted in fulfillment of the requirements for Honors

in the

Sanford School of Public Policy

April 2015

Declaration of Authorship

I, Emily Hadley, declare that this thesis titled, 'Rural and At-Risk: Predictors of and Methods to Address High School Dropout for Rising Ninth Graders in Rural North Carolina' and the work presented in it are my own. I confirm that:

- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- I have upheld the Duke Community Standard.

Signed: ______________________________

Date: ______________________________

DUKE UNIVERSITY

Abstract

Dr. Helen Ladd, Dr. Jerome Reiter
Sanford School of Public Policy

Undergraduate Honors Thesis

Rural and At-Risk: Predictors of and Methods to Address High School Dropout for Rising Ninth Graders in Rural North Carolina

by Emily HADLEY

This study identifies the individually significant predictors of high school dropout for rising ninth graders in rural North Carolina and suggests methods that policymakers could implement in a ninth grade dropout prevention program. A multivariate, multilevel logistic model with k-means clustering was used to identify significant predictors for the 53,996 students initially enrolled in third grade at any North Carolina school in 2002-2003, who remained enrolled in a North Carolina school through their expected eighth grade year of 2007-2008, and who were enrolled in a school classified as rural in their expected eighth grade year. The strongest predictors of high school dropout in rank order are low math test scores, many absences, economic disadvantage, low parent education levels, retention, and male gender. Race, learning disability, LEP status, and reading test scores are not good predictors of dropout. Evidence-based methods are then suggested for a ninth grade intervention program that would address low math achievement, poor attendance, retention, or male gender risk factors in rural North Carolina high schools.

Acknowledgements

I would like to thank Professor Helen Ladd from the Sanford School of Public Policy for her extraordinary support and guidance. I would like to thank Professor Jerome Reiter from the Department of Statistical Sciences for his statistical advice and enthusiasm. I would like to thank the North Carolina Education Research Data Center for providing the data and Professor Mayer from the Sanford School of Public Policy for his encouragement. Finally, I would like to thank my family and friends for their critiques, patience, and kindness.

Contents

Declaration of Authorship i

Abstract ii

Acknowledgements iii

1 Introduction 1
- 1.1 Motivation for Research 1
- 1.2 Organization of Paper 1

2 Research Questions 2

3 Background 3
- 3.1 Motivation for Preventing High School Dropout 3
- 3.2 Who is Considered a Dropout? 4
- 3.3 Factors in High School Dropout 4
  - 3.3.1 Demographic Factors: Gender and Race 5
  - 3.3.2 Home Factors: Economic Disadvantage and Parent Education Level 5
  - 3.3.3 Behavioral Factor: Days Absent 6
  - 3.3.4 Achievement Factors: Test Scores 6
  - 3.3.5 Learning Disabilities and Limited English Proficiency 6
  - 3.3.6 Retention 7
- 3.4 The Rural Landscape in North Carolina 7
  - 3.4.1 Defining Rural 8
- 3.5 Unique Assets of Rural Communities 8
- 3.6 Unique Challenges Faced by Rural Communities and Schools 8

4 Data 10
- 4.1 Data Overview 10
  - 4.1.1 Student Data 10
  - 4.1.2 Demographic Variables: Gender and Race 11
  - 4.1.3 Dropout, Economic Disadvantage, LEP, and Retention 11
  - 4.1.4 Days Absent 12
  - 4.1.5 Test Score Data 12
  - 4.1.6 School Data 14
- 4.2 Missing Data 14

4.3 The Limitation of “Ever Identified” 15

5 Methodology **16**
5.1 Logistic Regression Model 16
5.2 Clustering 16
5.3 Multilevel Model with Random Effects 18
5.4 Change-in-Deviance Test 19

6 Results **20**
6.1 Selecting the Model with the Best Fit 20
6.1.1 Testing Marginal Associations through Univariate Models 20
6.1.2 Initial Multivariate Logistic Regression Model 21
6.1.3 Adding Individual Coefficients to the Initial Model 23
6.1.4 Developing the Final Model 24
6.1.5 The Final Model 25

7 Using Results to Identify At-Risk Rural Students **28**
7.1 Variables That Could Be Used to Identify Students Who Would Benefit from a Ninth Grade Intervention 28
7.1.1 Math Test Scores 29
7.1.2 Absences 29
7.1.3 Economic Disadvantage 29
7.1.4 Parent Education 29
7.1.5 Retention 30
7.1.6 Gender 30
7.2 Variables That Should Not Be Used To Identify Students 30
7.2.1 Race 30
7.2.2 Learning Disability 31
7.2.3 LEP Status 31
7.2.4 Reading Test Scores 31

8 Policy Implications for a Ninth Grade Intervention Program **32**
8.1 Exemplary Dropout Prevention Programs 32
8.2 Exemplary Dropout Programs that Address Specific Risk Factors 34
8.2.1 Math Achievement 34
8.2.2 Absences 35
8.2.3 Economic Disadvantage 37
8.2.4 Parent Education 38
8.2.5 Retention 38
8.2.6 Gender 39

9 Conclusions **41**
9.1 Identifying At-Risk Students Who Could Benefit from a Ninth Grade Intervention Program 41
9.2 Methods for a Ninth Grade Intervention Program to Address Significant Risk Factors 41

A List of Variables **43**

B Results from Analysis **45**

Bibliography **53**

Chapter 1

Introduction

1.1 Motivation for Research

North Carolina has the second largest rural student population in the United States and a rural dropout rate on par with the national dropout rate [1]. The transition to ninth grade is crucial for students who are at risk of dropping out since most students drop out in the first two years of high school and ninth grade intervention programs have been successful in lowering dropout rates [2][3]. To help identify students that could benefit from these programs, a large body of research has explored dropout risk factors in urban areas and across the nation [4]. Yet little research has considered dropout risk factors specific to rural areas [4]. The purpose of this study is to identify the individually significant predictors of high school dropout for rising ninth graders in rural North Carolina to close this research gap. This study also identifies the programs and methods that have successfully addressed these risk factors and lowered dropout rates. Policymakers can consider some of these methods as possible opportunities for a ninth grade dropout intervention program in rural North Carolina.

1.2 Organization of Paper

Chapter 2 introduces the research question and Chapter 3 describes relevant background information. Chapter 4 explores the data and Chapter 5 outlines the methodology. Chapter 6 describes the results of the analysis. Chapter 7 discusses the implications for identifying students who would benefit from a ninth grade intervention program while Chapter 8 suggests evidence-based methods to address these risk factors. Chapter 9 concludes the paper with key points. Appendix A lists the variables and Appendix B includes all regression results.

Chapter 2

Research Questions

1. What are the most significant predictors of high school dropout for rising ninth graders in rural North Carolina schools?
2. What methods could policymakers use to address these significant predictors in a ninth grade dropout prevention program?

Chapter 3

Background

This chapter explores the literature on the implications of and factors contributing to high school dropout as well as the literature on the unique strengths and challenges faced by rural communities.

3.1 Motivation for Preventing High School Dropout

Graduating from high school is important for an individual's stability and opportunity. Individuals who drop out are less likely to be healthy and more likely to commit a crime [5] [6]. They are less likely to find a job since 75% of jobs require more than a high school degree and they will make an average of $550,000 less than individuals with a high school degree over the course of their lifetimes [7] [6]. A high school dropout also contributes an average of $139,100 less than a high school graduate to federal and state income taxes. For each new high school graduate, the public health system saves an average of $40,500 from reduced dependence on public health services while the criminal justice system saves an average of $26,600 from diminished criminal activity [8].

In 2009, North Carolina invested $11.7 million dollars in dropout prevention. The North Carolina Committee on Dropout Prevention estimated that this investment would be recovered from taxes and decreased reliance on social services if only 670 fewer students dropped out [9]. To accomplish this benchmark, only one third of the rural students who dropped out from the cohort considered in this study would have needed to graduate.

3.2 Who is Considered a Dropout?

This study uses the North Carolina definition of dropout: any student who leaves school before completion of a program of studies without transferring to another elementary or secondary school [10].

In practice, this is a student who was enrolled at some time during the previous school year but who was not enrolled and did not meet reporting exclusions on day 20 of the current school year. Schools that cannot document a former student's enrollment in a US school must report that student as a dropout unless they meet one of the following exclusions:

- Students who are known to have left the country
- Students who are serving suspensions
- Students who are expelled (expelled students are counted as dropouts for federal but not state reporting)
- Students who transfer to a private school, home school, or state-approved educational program
- Students who are not enrolled on day 20 because they have serious illnesses

Students who are reported as dropouts but are not included in the dropout rate are students who leave within the first 20 days of enrollment, students incarcerated in an adult facility, and students who fail to return to school after a long-term suspension. Students who are reported as dropouts and included in the dropout rate are students who leave the public schools to attend community college or who leave school to obtain a GED since it is not known whether or not they obtain a high school diploma [11].

3.3 Factors in High School Dropout

Students who drop out from high school often cite challenges internal and external to the school setting as reasons for dropping out. Internal challenges include strict school discipline policies and academic obstacles while external challenges include factors like out-of-school employment or having to take care of family members [12]. Since this study does not have access to data on students' reasons for dropping out, it will focus on the predictors that make a student more likely to drop out. The following factors are explored in this study:

3.3.1 Demographic Factors: Gender and Race

Past research has indicated that male students are less likely to graduate than female students [13] [14]. One theory for this difference is that girls display a more self-determined motivational profile than boys in high school [15]. Vallerand et al (1997) theorizes that teacher-student relationships account for some portion of this difference as female students are more likely than male students to perceive their teachers as supporting their own autonomy [15]. This lack of support combined with teachers increased propensity for critical, controlling, and punitive interactions with male students may lead to the development of a non-self-determined motivational profile that triggers undesirable consequences like dropping out of high school [15]. One limitation of the Vallerand et al (1997) study it that is was conducted in Canada, but the author suggests that it is likely generalizable to the US since similar teacher-student relationships have been observed in the US. Gender is included in this analysis to see whether status as a male student is also a significant predictor of dropout in rural areas of North Carolina.

Students from minority groups, especially Native American students, are more likely to drop out than White students [13][16]. The notable exception are Asian students who have the same or lower likelihood of dropping out when compared to White students [17][18]. Griffin (2002) suggests that this outcome arises since Black and Hispanic students tend to show increased detachment from academics than Asian and White students [17]. Rumberger (1983) suggests that most differences in racial dropout outcomes can be explained through family characteristics like socioeconomic status [19]. Lofstrom (2007) used a Texas student sample to corroborate Rumberger's national findings and also noted the influence of community effects like the neighborhood the student lives in [18]. Race disaggregated into Asian, Black, Hispanic, Multiracial, Native American, and White is included in this analysis to see whether or not race is a significant predictor independent of the other variables included in this study.

3.3.2 Home Factors: Economic Disadvantage and Parent Education Level

Students who receive free or reduced price lunches are labeled as economically disadvantaged and are less likely to graduate [14][20]. The effects of a low-socioeconomic background are deeply entrenched before a student arrives for the first day of kindergarten. For example, by age four, students from high-income families have heard approximately 30 million more words than students from low-income families [21]. Students from households with low parent education levels are also less likely to graduate [19][14]. Parent education is closely connected to economic disadvantage as lower education levels often correspond to lower income levels. Yet Rumberger (1983) found evidence that education level may be more than simply an indicator of income status as it has effects that are not captured by an economic disadvantage variable alone [19]. For example, parent education levels can correspond to educational expectations for a student. Both economic disadvantage and parent education level are considered in this analysis to understand their effects as predictors of dropout.

3.3.3 Behavioral Factor: Days Absent

Students with a significant number of days absent in elementary or middle school [22]. The benchmark for having a significant number of days absent is usually set at 10 percent of instructional time. Silver et al (2008) found that seventh and eighth grade attendance were significant predictors of high school dropout and recognized that more days absent likely reflects lower school engagement [13]. Janosz et al (2008) finds that students with lower school engagement are more likely to drop out and suggests early identification based on disengagement factors to prevent high school dropout [23]. A count of the number of days absent in each grade from fourth grade through eighth grade is included in this analysis.

3.3.4 Achievement Factors: Test Scores

Student academic achievement on both math and reading exams has been positively correlated with the odds of graduation, indicating that higher achieving students are more likely to graduate [24] [13][19]. Achievement in grades below high school is important as those who fail a class in sixth or seventh grade are significantly less likely to graduate [13]. Battin-Pearson (2000) suggests that the influence of academic achievement on the likelihood of dropout is stronger than some peer and family effects and that academic achievement should be a main focus of dropout intervention programs [2]. Reading and math end-of-grade test scores are included in this analysis.

3.3.5 Learning Disabilities and Limited English Proficiency

Studies appear to suggest that students with learning disabilities are more likely to drop out [25][26]. Ingrum (2006) further suggests that the effect of having a learning disability on the likelihood of dropping out is amplified by the effect of socioeconomic status [26]. Both status as learning disabled in reading and learning disabled in math are considered in this study.

While there is not a substantial amount of research on Limited English Proficiency (LEP) status in the United States and its relationship with high school dropout, LEP students in California are significantly more likely to drop out than non-LEP students [27]. Driscoll (1999) also finds that Hispanic students with lower English proficiency are more likely to drop out, even when controlling for other demographic and family background characteristics [28]. LEP status is included in this study

3.3.6 Retention

Students who were retained before eighth grade are significantly more likely to drop out and multiple studies have found that retention is the most significant predictor of high school dropout [29][30][31]. Students are sometimes retained in an effort to improve their academic achievement or socioemotional and behavioral adjustment, though the success of retention in affecting these outcomes is often questioned. In a metanalysis, Jimerson and Kaufman (2003) found that while some studies suggest small gains in test scores for retained students relative to similarly performing peers who were not retained, these gains typically disappear over time[30]. Roderick (1994) suggested that students who are retained are more likely to drop out because they are overage for their grade and show increasing disengagement with their school as they progress [31]. Retention before Grade 9 is considered in this analysis.

3.4 The Rural Landscape in North Carolina

In addition to identifying predictors of dropout, this study seeks to draw attention to the state of rural education in North Carolina and the state of rural education research. Currently, 49.2% of North Carolina schools are considered rural. North Carolina ranks second in the nation for both the number of rural students and the number of rural minority students. It ranks in the top ten states nationwide for percent of rural students designated as English Language Learners in 2014 (6.1%) and percent of rural students who are Title 1 eligible (23.7%). The Rural School and Community Trust releases a biannual Rural Education Priority Gauge. In the 2013-2014 report, North Carolina ranked fourth, indicating that rural North Carolina students are facing significant challenges like low achievement and high rural unemployment rates that are not being adequately addressed by state policy [1]. The body of rural education research is considerably smaller than the general body of education research. Arnold et al (2005) found only 490 research papers published between 1993 and 2001 that specifically addressed rural education [4]. Of these, only three explored high school dropout in rural areas [4].

3.4.1 Defining Rural

Rural is defined using the 2010-2011 district urban-centric locale codes. Schools in areas coded as Fringe Town, Distant Town, Remote Town, Rural Fringe, Rural Distant, and Rural Remote are considered rural. Schools in areas coded as Large City, Midsize City, Small City, Large Suburb, Midsize Suburb, and Small Suburb are considered non-rural. This definition of rural is preferred to the alternative option of the 2002-2003 National Center for Education Statistics location codes that were also included in the dataset as the 2010-2011 urban-centric codes were more detailed since the had more levels and more complete since there were fewer cases missing.

3.5 Unique Assets of Rural Communities

Rural America is often recognized for its small, tight-knit communities. Focus groups in rural Tennessee, Kentucky, and Alabama highlighted the importance of informal social networks in rural areas where adults in a neighborhood know and watch out for other children, ask for advice, do favors for each other, and often participate in faith-based activities [32]. Small communities reduce local government bureaucracy as every resident feels the impact of local policy and can increase inter-generational relationships [33].

Rural schools also have unique benefits. Rural schools tend to have lower student-teacher ratios compared to urban schools and parents often value the increased teacher attention [33][32]. Harde (2003) applied self-determination theory to education and suggested that the effect of a teacher's support on a student's motivation was noticeably stronger for rural students, potentially indicating that rural students' academic motivation is closely related to the quality of their teachers' motivating styles [34].

3.6 Unique Challenges Faced by Rural Communities and Schools

A major challenge in many rural areas is unemployment. The North Carolina rural unemployment rate is around 8.6%, the seventh highest in the nation for rural areas [1]. Mining and farming jobs have been replaced by low-wage jobs in retail and service industries, contributing to the economic disadvantage of these areas. In focus groups in rural Tennessee, Kentucky, and Alabama, the majority of participants said that if they had a magic wand, they would create more jobs with higher wages as participants believe that a higher income would improve the lives of their children and the community [32].

These low-wage environments contribute to increased poverty. The rural child poverty rate in North Carolina is higher than the urban child poverty rate in 19 states [35]. North Carolina ranks eighth out of all states for the number of students who are Title 1 eligible with 23.7% of students meeting the qualifying standards [1]. Families in rural areas often have high property taxes, sometimes up to half of a family's income in very poor areas [33]. Yet rural areas still have a smaller tax base when compared to urban areas [33].

Rural schools face a number of challenges due to limited funding and small school size. First, rural schools tend to have lower teacher salaries and subsequently struggle to recruit high qualified teachers and teachers with advanced degrees [36] [33]. Rural schools also struggle to offer advanced courses like AP classes and are forced to limit counseling and psychological services [36]. Students travel significant distances to get to school and many may not be able to participate in after school activities since they must take the only available bus home right after school ends. There also are not as many extracurricular opportunities in rural areas due to the small school size, contributing to the discovery that the highest level of drug use for youth ages 12 to 17 is in rural communities [37].

Finally, rural communities tend to have less exposure to diversity and higher education. Though North Carolina has the ninth largest rural minority population and ranks eighth for number of ELL students, many rural areas are not exposed to racial or cultural diversity as they have been a majority white for generations. Adults in rural areas are more likely to have a high school degree than adults in non-rural areas, but rural adults are considerably less likely to have a college degree [35]. Parents in the Tennessee, Kentucky, and Alabama focus groups recognized that the lack of higher education in rural areas may be contributing to poor early childhood home environments relative to peers in more educated households [32].

Chapter 4

Data

4.1 Data Overview

The data for this analysis was provided by the North Carolina Department of Public Instruction. The North Carolina Education Research Data Center (NCERDC) created this dataset by concatenating the data across school years to create one file per student. NCERDC also provided files on each school attended by a student in the cohort. School indicators were used for grouping in the hierarchical model.

4.1.1 Student Data

The data consists of the 53,996 students initially enrolled in third grade at any North Carolina school in 2002-2003, who remained enrolled in a North Carolina school through the expected eighth grade year of 2007-2008, and who were enrolled in a school classified as rural in their expected eighth grade year. Not all students in the cohort were actually in eighth grade in the 2007-2008 school year as some students were retained.

Data was collected on this cohort from third grade through expected graduation in 2012. Students who moved to North Carolina after the 2002-2003 school year were not included as they were not added to the cohort. Students who left the cohort before their expected eighth grade year were removed from the dataset since the goal of this analysis is to predict the likelihood of dropout for rising ninth graders. Approximately 3.9% of the cohort (2068 students) dropped out. This is slightly above the national 2011-2012 dropout percentage of 3.3%.

The student-level dataset initially contained 148 variables. Predictors collected in high school, such as End of Course test scores and days absent in ninth, tenth, eleventh, or

Table 4.1: Percentage and Count of Students By Race and Gender

Gender	White	Black	Asian	Hispanic	Native American	Multiracial	Total
Male	31.1	14.0	0.6	3.1	1.3	0.9	50.9
	(16786)	(7563)	(311)	(1673)	(675)	(512)	(27520)
Female	30.1	13.3	0.5	3.0	1.1	0.9	49.1
	(16254)	(7197)	(281)	(1634)	(600)	(510)	(26476)
Total	61.2	27.3	1.1	6.1	2.4	1.9	100
	(33040)	(14760)	(592)	(3307)	(1275)	(1022)	(53996)

twelfth grade, were removed as the goal of this analysis is to predict dropout based on data available through eighth grade. The final student-level dataset was reduced to 34 variables. A full list of variables is in Appendix A.

4.1.2 Demographic Variables: Gender and Race

Table 4.1 illustrates the racial and gender distribution of the students in the cohort. The largest racial group in the cohort is white students followed by African American students, Hispanic students, and Native American students. Multiracial and Asian students compose an especially small proportion of the cohort. There are slightly more male students than female students.

4.1.3 Dropout, Economic Disadvantage, LEP, and Retention

Table 4.2 further disaggregates the gender and race data by other risk factors. This table illustrates that male and female students in the cohort have similar rates of being identified as economically disadvantaged (ED) or Limited English Proficiency (LEP). Male students are more likely to drop out, more likely to be retained, and more likely to have a learning disability in reading (LDR) or a learning disability in math (LDM).

Table 4.2 also reveals racial disparities among these risk factors. Native American males and females have remarkably high rates of dropout. African American and Hispanic students have high rates of economic disadvantage. More than 90% of both male and female Hispanic students qualify for free or reduced price lunch. Asian and Hispanic males and females have high rates of LEP status which is expected since these are two of the largest immigrant groups in America. One particularly notable statistic is that more than 20% of the Hispanic, Native American, and Multiracial male students were

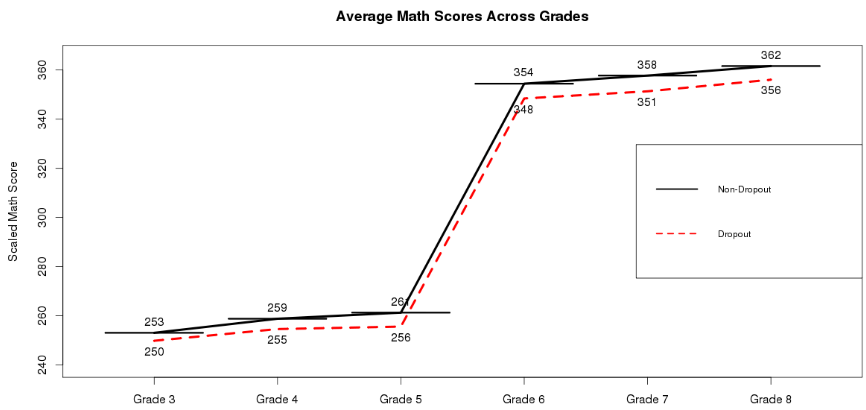

FIGURE 4.1: This figure illustrates the difference between the average math test scores across grades for students who do not drop out and the average math test scores across grades for students who do drop out.

retained before eighth grade. A different retention trend was seen for female students as Black female students were most likely to be retained.

4.1.4 Days Absent

Table 4.3 shows that in fourth grade, students who go on to drop out already have more absences on average than students who do not go on to drop out.[1] This trend widens over time as students who go on to drop out see a 6.3 day increase in the average number of days absent while the students who do not go on to drop out see a 2.3 day increase in the average number of days absent.

4.1.5 Test Score Data

Figure 4.1 and Figure 4.2 show that, starting in third grade, students who do not go on to drop out have higher average test scores in both reading and math. The gap in reading scores remains around five points while the gap in math scores increases from around three points to between six and seven points. Since the test scores are scaled, they are normally distributed around the mean. Due to this normalcy, Welch's Two Sample T-Test was used to confirm that the average test scores in both math and reading for students who went on to dropout were lower than the average test scores for students who did not go on to drop out at a statistically significant level.

[1] Third grade absence data was not available in this dataset

Table 4.2: Percentage and Count of Students by Gender, Race, and Risk Factors

Males	Drop Out	ED	LEP	Retained	LDR	LDM
White	4.1	44.9	0.4	6.1	14.0	9.8
	(700)	(7548)	(64)	(1031)	(2354)	(1646)
Black	5.4	87.9	0.3	14.0	18.1	14.5
	(412)	(6646)	(19)	(1061)	(1371)	(1100)
Asian	1.9	67.9	58.5	3.9	6.8	4.8
	(6)	(211)	(182)	(12)	(21)	(15)
Hispanic	4.7	94.5	88.2	28.9	16.2	12.2
	(79)	(1581)	(1476)	(484)	(271)	(204)
Native American	8.0	88.6	0.0	26.1	20.7	15.7
	(54)	(598)	(0)	(176)	(140)	(106)
Multiracial	4.5	75.2	3.5	24.2	16.4	12.7
	(23)	(385)	(18)	(124)	(84)	(65)
All Males	4.6	61.7	6.4	8.8	15.4	11.4
	(1274)	(16969)	(1759)	(2400)	(4241)	(3136)
Females	Drop Out	ED	LEP	Retained	LDR	LDM
White	2.8	45.3	0.3	3.8	7.4	5.4
	(455)	(7355)	(54)	(611)	(1210)	(873)
Black	3.4	88.4	0.2	8.0	9.3	7.9
	(248)	(6362)	(16)	(572)	(670)	(569)
Asian	.4	61.6	55.5	1.4	5.7	4.3
	(1)	(173)	(156)	(4)	(16)	(12)
Hispanic	2.9	93.3	87.6	6.1	9.4	7.1
	(48)	(1523)	(1431)	(100)	(154)	(116)
Native American	5.2	87.0	0.8	6.0	10.2	7.8
	(31)	(522)	(5)	(36)	(61)	(47)
Multiracial	2.2	72.7	3.7	3.5	8.8	5.7
	(11)	(371)	(19)	(18)	(45)	(29)
All Females	2.9	61.6	6.3	5.1	8.1	6.2
	(794)	(16306)	(1681)	(1341)	(2156)	(1646)

Percentage and count (in parentheses) of each gender disaggregated by race and risk factors. Example interpretation is that 4.1% of White, Male students drop out. ED = Economically Disadvantaged; LEP = Limited English Proficiency; LDR = Learning Disability in Reading; LDM = Learning Disability in Math

TABLE 4.3: Mean and SD of Days Absent by Dropout Status and Grade

Days Absent	Dropout		Non-Dropout	
Grade	Mean	SD	Mean	SD
Fourth	8.6	7.6	6.2	5.8
Fifth	8.9	7.4	6.4	6.2
Sixth	11.1	9.8	7.2	7.1
Seventh	13.1	11.1	8.1	8.2
Eighth	14.9	12.7	8.5	9.1

SD = Standard Deviation

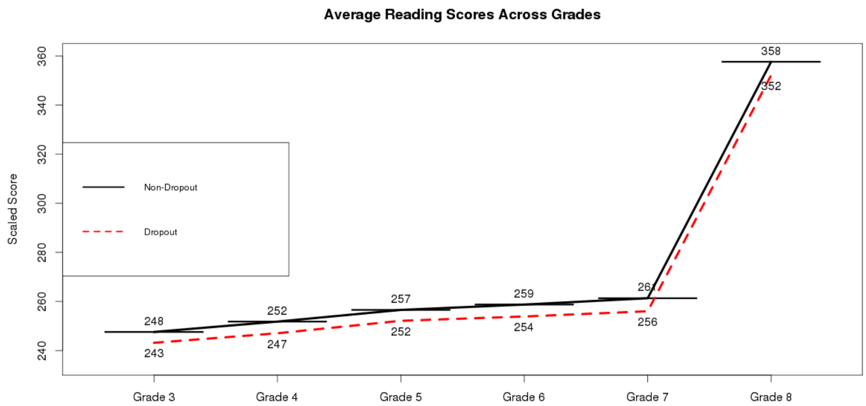

FIGURE 4.2: This figure illustrates the difference between the average reading test scores across grades for students who do not drop out and the average reading test scores across grades for students who do drop out.

4.1.6 School Data

Data was also collected for the 1226 rural schools attended by the students in this cohort. The strongest correlation between dropout and a school-level variable was a small 0.05 correlation between dropout and the average percent of students proficient in math in a particular school. This suggests that student-level variables are better predictors of dropout than school-level variables. Since school factors were incorporated into the final model by using multilevel analysis which intrinsically controls for school-level effects, individual school variables were not included in the model as they would be redundant.

4.2 Missing Data

While the outcome variable of dropout status is recorded for all students, some predictor values were missing. These included items like test score data for students who were retained, learning disability indicators, and LEP status. Rather than eliminate cases with missing data, multivariate imputation through chained equations in the **MICE** package in **R** was used to impute all missing values.[2]

4.3 The Limitation of "Ever Identified"

One limitation of this dataset is that the following variables are based on whether students were *ever identified* as having these characteristics. Since the initial dataset went from third grade through high school, it is possible that the students in the cohort did not have the following characteristics until high school:

- Economic Disadvantage
- Learning Disability in Reading
- Learning Disability in Math
- Limited English Proficiency

This discrepancy threatens the analysis because the models in the subsequent chapters are attempting to predict dropout based on data from third through eighth grade and it is impossible to know when students were assigned these labels.

Yet, there is reason to believe that late labeling in high school for these four variables is unlikely. Late labeling as economically disadvantaged is unlikely as students in high school are less likely to sign up for free/reduced price lunch. If students were ever going to be labeled as economically disadvantaged, it is more likely that they would first receive the label in elementary school [39]. It is also unlikely that a student would not have been identified as Learning Disabled in Reading, Learning Disabled in Math, or Limited English Proficiency until high school since all students in this study were part of the 2002-2003 third-grade cohort and it is unlikely that these characteristics would not have been noticed until ninth grade or after. This study therefore assumes that these labels were assigned before high school while acknowledging that this assumption is still a limitation of the analysis.

[2] The **MICE** package uses multivariate imputation through chained equations which is an iterative process that imputes missing variables based on the other observations in the dataset. A detailed description is available in Azur et al (2012) [38].

Chapter 5

Methodology

This study seeks to maximize both accuracy and interpretation to identify the significant predictors of high school dropout in rural North Carolina. A multivariate, multilevel logistic regression with clustering is used to achieve these goals.

5.1 Logistic Regression Model

Logistic regression is often used to model situations with binary outcomes. In this study, the binary outcome is 'dropout' or 'non-dropout'.

The odds of a student being a dropout are defined as the probability that a particular student is a dropout π_i divided by the probability that a particular student is not a dropout $1 - \pi_i$. Equation 5.1 shows how logistic regression predicts the natural log of these odds in terms of an intercept B_0, a series of predictors X_n, and their respective coefficients B_n.

$$ln(\frac{\pi_i}{1 - \pi_i}) = B_0 + B_n X_n \tag{5.1}$$

The coefficients B_n for the predictors X_n represent the change in the logit outcome for a unit increase in the predictor.

5.2 Clustering

K-means clustering is an effective way of summarizing data across time to minimize the effects of multicollinearity between years. In k-means clustering, an iterative algorithm

TABLE 5.1: SD of Test Scores in Reading and Math Clusters Across Grades

Reading Cluster — Grade	3rd	4th	5th	6th	7th	8th	N
Considerably Above Average	1.36	1.41	1.42	1.41	1.39	1.50	4616
Above Average	0.74	0.77	0.76	0.77	0.77	0.78	8116
Slightly Above Average	0.21	0.20	0.21	0.22	0.22	0.18	11269
Slightly Below Average	-0.31	-0.34	-0.32	-0.33	-0.35	-0.42	12544
Below Average	-0.95	-0.96	-0.96	-0.97	-0.97	-1.01	11338
Considerably Below Average	-1.76	-1.78	-1.83	-1.85	-1.78	-1.57	6113
Math Cluster — Grade	3rd	4th	5th	6th	7th	8th	N
Considerably Above Average	1.56	1.59	1.72	1.65	1.70	1.75	5490
Above Average	0.85	0.90	0.94	0.97	0.97	0.94	10445
Slightly Above Average	0.33	0.35	0.33	0.37	0.34	0.25	12374
Slightly Below Average	-0.17	-0.20	-0.26	-0.27	-0.27	-0.33	11798
Below Average	-0.72	-0.78	-0.85	-0.95	-0.94	-0.80	9131
Considerably Below Average	-1.71	-1.68	-1.55	-1.43	-1.40	-1.33	4758

N=Number of students in each cluster

is used to assign each data point to the cluster with the nearest mean. The number of clusters is fixed at the number where adding another cluster doesn't lead to better modeling of the data.[1] In this study, k-means clustering was utilized to summarize data for math scores from grades 3-8, reading scores from grades 3-8, and absence data from grades 4-8. Each student was placed in one of six reading and math score clusters. Each student was also placed in one of five absence clusters.

Table 5.1 shows the six score clusters that each student is assigned to. For example, a student who consistently performed 1.6 standard deviations above the average test score would, through the k-means algorithm, be placed in the "Considerably Above Average" cluster group. Table 5.1 also shows the mean for each cluster across all six years of data collection. Distinct differences are clear between each cluster, indicating that individual students often follow a pattern of high, average, or low performance. The clusters with the most students are those that are closest to the average while the clusters with the smallest numbers of students are the furthest from the average. This distribution of students is expected since the test score data for all grades is distributed normally.

Table 5.2 shows the mean number of days absent for each cluster across five years of data collection. While differences between clusters are not as defined as in the test score clusters, each cluster still exhibits a pattern that is described by the cluster name. The

[1]Hartigan and Wong (1979) give a detailed description of a K-Means clustering algorithm [40].

Table 5.2: Number of Days Absent in Each Grade by Absence Cluster

Attendance — Grade	4th	5th	6th	7th	8th	N
Low Absence	3.14	2.93	3.13	3.45	3.70	24837
Mild Absence	7.87	7.92	8.51	8.67	8.09	17368
Increasing Absence	7.15	7.54	9.99	14.62	20.0	5384
Moderate Absence	15.66	17.13	19.97	19.25	16.09	5152
Severe Absence	15.15	17.65	24.53	37.37	46.30	1255

N=Number of students in each cluster

Increasing Absence, Moderate Absence, and Severe Absence clusters are all notable as they each have at least one grade where the average number of days absent is more than 10% of the school year (generally 18 days). The cluster with the most students is the cluster with the lowest number of average absences and the cluster with the fewest students is the cluster with the highest number of average absences. This distribution of students is expected since absence data exhibits a right skew where students are far more likely to have fewer absences.

5.3 Multilevel Model with Random Effects

Multilevel modeling is used to control for school level effects since many students in this study attend the same schools. Guo and Zhao highlighted three benefits of using multilevel modeling:

- Multilevel modeling permits analysis of how variables collected at both the student and school level may affect the outcome of dropping out.
- Multilevel modeling can correct biases in parameter estimates as there can be specific correlations between variables collected within a school and these correlations, if strong, may not be adequately adjusted for in a simple logistic regression.
- Multilevel modeling also corrects standard errors as it adjusts for cluster effects. Independence among observations is a crucial assumption of a binary regression, yet this assumption may not be valid because students are clustered at a school level. Multilevel modeling takes this into account and provides correct standard errors which then contribute to correct confidence intervals and significant levels. [41]

The multilevel equation is:

$$y_{ij} = B_0 + B_n x_n + u_j + e_{ij} \tag{5.2}$$

In Equation 5.2, y_{ij} is the log odds of dropping out for the ith student at the jth school. Like in the logistic model, B_0 is the intercept, X_n is the explanatory variable for the ith student at the jth school, and B_n is the accompanying coefficient. u_j is a random effect accounting for variation at the school level and e_{ij} is the error term [41].

5.4 Change-in-Deviance Test

Logistic regression uses deviance as a statistic to describe the overall fit of a model. Deviance measures the difference in the observed values from the expected values. The larger the deviance, the worse the model fit. A chi-square change-in-deviance test is used in the following chapter to select the significant variables that ought to be used in the multilevel multivariate logistic regression model to ensure the best fit.

Chapter 6

Results

In the following regressions, the coefficients for continuous variables can be interpreted as the effect of a one-unit increase in that variable on the log odds of the outcome. For categorical variables, the coefficient represents the effect of having that variable on the log odds of the outcome. In both cases, exponentiating the coefficient allows for interpretation of that particular variable on the odds of dropout.

Race, education level, test score clusters, and absence clusters are treated as factors in this analysis. This means that each of the groups within these variable are compared to a baseline group. Race is compared to the baseline of White since White students are the most common in the dataset. A negative coefficient indicates that a different race has lower odds of dropping out while a positive coefficient indicates that a different race has higher odds of dropping out. Parent Education is compared to a base level of not having a high school degree as this is the lowest level of education. Test Score clusters are compared to the lowest performance level clusters. Absence clusters are compared to the lowest average Days Absent cluster.

6.1 Selecting the Model with the Best Fit

6.1.1 Testing Marginal Associations through Univariate Models

Marginal associations for each individual variable are calculated by regressing the log odds of dropout on each variable and an intercept while also accounting for the random school-level effects. Equation 6.1 illustrates this calculation where B_0 is the intercept, u_j is the school random effect, x is the variable being analyzed, and B_1 is the resulting coefficient.

$$logit(dropout) = B_0 + u_j + B_1 x \tag{6.1}$$

Table 6.1 shows the B_1 coefficient for each variable and its standard deviation. Nearly all variables have significant marginal associations with the log odds of dropping out when no other variables are included in the model.

Notable exceptions include Hispanic, Multiracial, and LEP status. The odds of a Hispanic or Multiracial student dropping out are not significantly different from the odds of a White student. Status as an LEP student does not significantly predict high school dropout.

While it appears from Table 6.1 that many variables could be individually used to predict high school dropout, marginal association models are poor predictive models as they only include one variable. By controlling for the effects of other variables in the regression model, one can better isolate the true effect of specific variables on the likelihood of dropping out. The following regression models are better predictive models since they control for the effects of multiple variables.

6.1.2 Initial Multivariate Logistic Regression Model

The initial model is intended to be a baseline with which to add or subtract variables as necessary to build the model with the best fit. The initial model was created with all variables that did not lead to dramatic differences in coefficients from the marginal associations. To be specific, Economic Disadvantage, Retained, Math Cluster, and Reading Cluster were not included in the initial model as they have dramatic effects on Race coefficients that will be investigated in separate regressions. Learning Disabled in Reading (LDR) was not included as though its marginal association was significant, Learning Disabled in Math (LDM) appears to account for the effects of LDR and a change-in-deviance test indicated that LDR did not significantly improve the model when LDM was already included. LEP status was also not included as a change-in-deviance test suggested that it was not a significant addition to the model. Therefore, the initial model is:

$$\begin{aligned} logit(dropout) = B_0 + u_j + B_1(Gender) + B_2(Race : Factored) + B_3(LDM) \\ + B_4(AbsenceCluster : Factored) + B_5(ParentEducation : Factored) \end{aligned} \tag{6.2}$$

In Equation 6.1, B_0 is the intercept, u_j are the school level random effects, and B_n with n from 1 to 5 are the coefficients. The full results of the regression are in Appendix B.

Table 6.1: Univariate Regression Results

Variable	Coefficient	SD
Male	0.46***	0.05
Asian	-1.03***	0.38
American Indian	0.65***	0.14
Black	0.21***	0.05
Hispanic	0.05	0.10
Multiracial	-0.08	0.18
LD: Reading	0.68***	0.06
LD: Math	0.76***	0.06
Economic Disadvantage	1.47***	0.07
LEP	-0.06	0.10
Retained	1.16***	0.06
Mild Absences	0.61***	0.06
Increasing Absences	1.29***	0.07
Moderate Absences	1.37***	0.07
Severe Absences	1.71***	0.10
ParEd: HS Degree	-0.48***	0.05
ParEd: Some Ed After HS	-0.73***	0.09
ParEd: Trade School	-1.15***	0.09
ParEd: Junior College	-1.19***	0.17
ParEd: Four-Year College	-2.04***	0.12
ParEd: Grad Degree	-2.18***	0.27
Math: Below	-0.36***	0.06
Math: Slightly Below	-0.83***	0.07
Math: Slightly Above	-1.33***	0.08
Math: Above	-2.02***	0.11
Math: Considerably Above	-2.94***	0.21
Reading: Below	-0.35***	0.07
Reading: Slightly Below	-0.70***	0.07
Reading: Slightly Above	-1.09***	0.07
Reading: Above	-1.66***	0.09
Reading: Considerably Above	-2.35***	0.15

*p<0.1; **p<0.05; ***p<0.01

All coefficients are significant indicating that they have a significant impact on the odds of dropping out, but will not be discussed in detail as this is not the final model.

6.1.3 Adding Individual Coefficients to the Initial Model

The four variables left out of the initial model due to correlation concerns that were evident from their dramatic effects on the coefficients of Race were Economic Disadvantage, Retained, Math Cluster, and Reading Cluster. Each of these variables was significant when added individually to the initial model. The inclusion of these variables impacted the coefficients of the other variables in the dataset in the following ways:[1]

- When Economic Disadvantage is added to the model, American Indian students no longer have significantly higher odds than White students of dropping out and Black, Hispanic, and Multiracial students all have significantly lower odds of dropping out. The coefficients for both Days Absent and Parent Education are not as large, though they all maintain the same sign.
- When Retained is added to the model, American Indian students return to having significantly higher odds than White students of dropping out but Hispanic and Multiracial students have significantly lower odds. The odds of dropout for Black students are not significantly different from the odds of dropout for White students. A difference between adding Retained and adding Economic Disadvantage is that when adding Retained, LDM is no longer significant. A similarity between adding Economic Disadvantage and adding Retained is that the coefficients for both Days Absent and Parent Education are again not as large as in the initial model.
- When Math Cluster is added to the model, American Indian students no longer have significantly higher odds than White students of dropping out and Black, Hispanic, and Multiracial students all have significantly lower odds of dropping out with even more negative coefficients than when Retained was added. The coefficients for both Days Absent and Parent Education are again not as large as in the initial model. LDM is also no longer significant.
- When Reading Cluster is added to the model, the effects are similar to adding Math Cluster to the model. American Indian students again no longer have significantly higher odds than White students of dropping out and Black, Hispanic, and Multiracial students all have significantly lower odds of dropping out. The coefficients for both Days Absent and Parent Education are again not as large as in the initial model. LDM is also no longer significant.

[1]Regression results for these models are in Appendix B.

The results from adding individual coefficients to the model suggest that there may be a substantial amount of multicollinearity, particularly between these individual variables and Race since the inclusion of any of these four variables had dramatic effects on the sign and strength of the coefficients for Race. The influence of LDM also appears to be better described by the other variables in the model as LDM was not significant when three of the four variables considered above were included.

6.1.4 Developing the Final Model

Development of the final model began with all variables from the initial baseline model as well as Economic Disadvantage, Retained, Math Cluster, and Reading Cluster. Reading Cluster was dropped from the model after a change-in-deviance tests showed that the Reading Clusters did not significantly contribute to the model.

The next step was to address the multicollinearity discovered when adding individual coefficients to the initial model. Options to address multicollinearity include combining, dropping, or interacting variables. Combining variables did not make sense for this study as there were few variables that could be combined in an interpretable manner.

For dropping variables, each variable in the model was dropped while the other variables remained in the model to see if that improved model fit. Each of these regressions had small p-values for their change-in-deviance test when compared to the first full model, indicating that each variable (including LDM) contributed significantly to the model.

For the third option - interacting variables - the focus was on interactions with Race since including Economic Disadvantage, Retained, or Math Cluster in the model had a substantial impact on the coefficients for Race. The three interactions considered were:

1. *Economic Disadvantage * Race (factored)*
2. *Retained * Race (factored)*
3. *Math Cluster * Race (factored)*

Change-in-deviance tests showed that none of the interactions were significant so they were not included in the final model.

6.1.5 The Final Model

The final model is:

$$\begin{aligned} logit(dropout) &= B_0 + u_j + B_1(Gender) + B_2(Race : Factored) + B_3(LDM) \\ &+ B_4(AbsenceCluster : Factored) + B_5(ParentEducation : Factored) \\ &+ B_6(EconomicDisadvantage) + B_7(Retained) + B_8(MathCluster : Factored) \end{aligned} \tag{6.3}$$

In Equation 6.2, B_0 is the intercept, u_j are the school level random effects, and B_n with n from 1 to 8 are the coefficients for the variables. Table 6.2 shows the regression output.

TABLE 6.2: Regression Results from Final Model

School Random Effects: Variance = 0.12, SD = .34

Variable	Coefficient
Constant	-3.579^{***}
	(0.119)
Male	0.395^{***}
	(0.048)
Asian	-0.895^{**}
	(0.387)
American Indian	-0.012
	(0.136)
Black	-0.409^{***}
	(0.059)
Hispanic	-0.572^{***}
	(0.104)
Multiracial	-0.445^{**}
	(0.181)
LDM	-0.147^{**}
	(0.070)
Mild Absence	0.387^{***}
	(0.062)
Increasing Absence	0.791^{***}
	(0.072)
Moderate Absence	0.835^{***}
	(0.073)

Continued on next page

Table 6.2 – *Continued from previous page*

Severe Absence	0.910***
	(0.107)
ParEd: HS Degree	−0.226***
	(0.061)
ParEd: Some Ed After HS	−0.241**
	(0.099)
ParEd: Trade School	−0.490***
	(0.098)
ParEd: Junior College	−0.447**
	(0.175)
ParEd: Four-Year Degree	−0.774***
	(0.130)
ParEd: Grad Degree	−0.536*
	(0.285)
ED	0.867***
	(0.076)
Retained	0.391***
	(0.069)
Math: Below Average	−0.199***
	(0.067)
Math: Slightly Below Average	−0.512***
	(0.075)
Math: Slightly Above Average	−0.849***
	(0.088)
Math: Above Average	−1.345***
	(0.118)
Math: Considerably Above Average	−1.974***
	(0.221)
Observations	53,957
Log Likelihood	−7,883.281
Akaike Inf. Crit.	15,820.560
Bayesian Inf. Crit.	16,060.750

*p<0.1; **p<0.05; ***p<0.01
Odds of dropout obtained by exponentiation of coefficient

Nearly all of the predictors significantly impact the odds of dropping out. Status as a male student increased the odds of dropping out in all models. In the final model, the odds of dropping out for a male student are 1.48 times greater than the odds for a female student when holding all other variables constant.

Coefficients for Race are not nearly as consistent as the coefficient for male across these regressions. This suggests that the any effect of Race on dropout is better captured by other variables like retention or economic disadvantage. In the final model, Asian students are 0.41 times as likely, Black students are 0.66 times as likely, Hispanic students, are 0.56 times as likely, and Multiracial students are 0.62 times as likely to drop out as White students when holding all other variables constant. Since the Native American coefficient is not significant, Native American students do not appear to have significantly different odds of dropping out than White students when controlling for the other variables included in this model.

This model suggests that students who are learning disabled in math are 0.86 times as likely to drop out as students who are not learning disabled in math. This is an unexpected result and it may arise from the earlier exploration that showed that Retained and Math Cluster may be better at capturing the influence of a math learning disability. Math is kept in the model due to its significant change-in-deviance tests.

Students with more absences have higher odds of dropping out. Students in the Moderate and Severe absences category which both have years where students miss 10% or more of the school year increase the odds of dropping out by 2.30 and 2.48 respectively when compared to a baseline of low number of days absent.

Students whose parents have obtained a high school degree or higher level of education have lower odds of dropping out than students whose parents did not graduate from high school. Students with parents who had a degree from a four-year college were 0.48 times as likely to drop out as a student with parents who did not have a high school degree when all other variables were held constant.

Students who were economically disadvantaged were 2.38 times as likely to drop out as students who were not economically disadvantaged when all other variables were held constant. Students who were retained were 1.48 times as likely to drop out than students who were not retained when all other variables were held constant.

Students with higher math scores were significantly less likely to drop out. Students who scored considerably above average were 0.14 times as likely to drop out as students who scored considerably below average.

Chapter 7

Using Results to Identify At-Risk Rural Students

Chapter 7 demonstrates how the results of this analysis illustrate which variables policymakers should and should not use for identifying rural North Carolina students who are at-risk of dropping out and who could benefit from a ninth grade intervention program. Before reading this chapter, policymakers should note two limitations of this analysis:

- First, this study establishes predictive factors not causality. The significant predictors cannot be said to cause drop out but are simply indicators of what makes a student more likely to drop out. For example, a policymaker can recognize that students who are retained would likely benefit from a ninth grade intervention program as this study shows that they are at significantly at-risk of dropping out, but the policymaker should not conclude that retention causes dropout.
- Second, due to the effects of multicollinearity, policymakers should consider a combination of the following predictors when identifying students for a program as one absolute eligibility characteristic may not adequately capture all students at-risk of dropping out.

7.1 Variables That Could Be Used to Identify Students Who Would Benefit from a Ninth Grade Intervention

The following is a rank order of the most significant predictors of high school dropout that could be used to identify students who would likely benefit from a ninth grade intervention program.

7.1.1 Math Test Scores

Math Cluster is the strongest predictor of dropout. A student who scores above average in math is significantly less likely to drop out than a student who scores below average in math. This finding is consistent across regression models. Policymakers should consider intervention programs directed at students who are low-performing in math.

7.1.2 Absences

Students with larger numbers of absences are more likely to drop out and this effect is consistent in all regression models. Three groups are especially concerning: those who see a dramatic increase in absences in seventh and eighth grade (an average of 7 days absent in fourth grade to an average of 20 days absent in eighth grade), those who are moderately absent across all grades (miss an average of 15 to 19 days), and those with severe, increasing absences (miss an average of 15 days in fourth grade and 46 days in eighth grade). Missing 10% of the school year (18 days in a 180 degree school year) is the typical benchmark for problematic absences. Given the results of this study, a policymaker should consider lowering that benchmark and targeting students who miss 15 or more days of school for a ninth grade intervention program.

7.1.3 Economic Disadvantage

Rising rural ninth graders who have ever been identified as economically disadvantaged have higher odds of dropping out than their non-disadvantaged classmates. Economic disadvantage has a significant positive effect on high school dropout in all regression models considered. Policymakers should consider ninth grade intervention programs specifically designed for students from low-income families.

7.1.4 Parent Education

Students whose parents have a high school degree or higher level of education have lower odds of dropping out when compared to students whose parents did not obtain a high school degree. This effect is consistent through all regression models, though the strength of the coefficients is reduced when other variables are included. Policymakers should consider identifying students from families with low education levels for an intervention program.

7.1.5 Retention

Retention before eighth grade is a significant but moderate predictor of high school dropout for rising ninth graders in rural North Carolina. This finding is consistent across all models. Policymakers should consider targeting students who were retained for a ninth grade intervention program.

7.1.6 Gender

Rising male ninth graders in rural areas are significantly more likely to drop out than rising female ninth graders in rural areas. A policymaker may consider targeting male students in a ninth grade intervention program since their odds of dropping out are significantly higher.

7.2 Variables That Should Not Be Used To Identify Students

The following variables should not be used for identifying students for a ninth grade intervention as they are either not accurate or not good predictors. They are described in no particular order.

7.2.1 Race

In the final model, Asian, Black, Hispanic, and Multiracial students were less likely to drop out than White students while Native American Students were not significantly different from White Students. Yet, before economic disadvantage, retention, or achievement in math were included in the model, Black and American Indian students were significantly more likely to drop out that White students while Hispanic and Multiracial students were not significantly different than White students. These conflicting results suggest that race is not a consistent predictor of dropout and that economic disadvantage, retention, or achievement in math may better capture the influence that race has on the odds of dropping out. A policymaker should not include race as factor in identifying students for a ninth grade intervention program as it is difficult to isolate the true effects of race on dropping out and there are alternative measures that are more accurate and better predictors.

7.2.2 Learning Disability

In the initial model, a learning disability in math had a significant, positive effect on the odds of dropout but in the final model, a learning disability in math had a significant, negative impact on the odds of dropout. This change indicates that the effects of a learning disability on the likelihood of dropout are not robust and are affected by the other variables in the model. Policymakers should not use learning disabilities as identifiers for at-risk students who could benefit from an intervention program because the influence of a learning disability on the likelihood of dropout is difficult to isolate and other variables are better predictors.

7.2.3 LEP Status

LEP status was not significant in any model tested. This suggests that the odds of a rural LEP student dropping out are not significantly different from the odds of a rural non-LEP student dropping out. A policymaker should not target LEP students for dropout intervention programs as resources could be better focused on students who are more likely to drop out.[1]

7.2.4 Reading Test Scores

Reading test scores are significant predictors of high school dropout until math scores are included in the model. This suggests that math scores are better predictors of high school dropout than reading scores. Policymakers should use low performance on math tests to identify students at risk at dropping out instead of low performance on reading tests when identifying students for a ninth grade intervention program.

[1]It is important to note that North Carolina has seen a dramatic increase in the Latino student population in the last 10 years and many Latino students qualify for LEP status. Research on a more recent cohort may change this recommendation if LEP status has become a significant predictor of high school dropout in recent years [42].

Chapter 8

Policy Implications for a Ninth Grade Intervention Program

Chapter 8 discusses existing prevention programs that have successfully reduced dropout rates by addressing the risk factors identified as important in Chapter 7. Again, this study does not imply a causal relationship between the significant predictors and dropout, but policymakers can still use the predictors to select and build ninth grade intervention programs.

8.1 Exemplary Dropout Prevention Programs

Hammond et al (2007) completed one of the most rigorous analyses of high school dropout prevention programs across the nation through a collaboration between the National Dropout Prevention Center/Network, the National Dropout Center for Students with Disabilities, and leading dropout prevention organization Communities in Schools, Inc [43]. Hammond et al (2007) used existing research on evidence-based dropout programs to identify fifty exemplary programs. To be exemplary, a program had to meet a series of standards:

- Were ranked in a top tier or level by at least two sources
- Were currently in operation
- Had no major revisions since the ranking of the program
- Had consistent, positive evaluation outcomes
- Targeted K-12 school populations

TABLE 8.1: Specific Methods Used by Exemplary Programs to Address Each Risk Factor

Program	Ret	ED	Att	Ach	ParEd
Across Ages			X		
AVID				X	
Big Brothers, Big Sisters			X	X	
Career Academy			X	X	
CASASTART	X				
Check and Connect			X	X	
Coca-Cola Valued Youth				X	
Families and Schools Together				X	X
Helping the Non-Compliant Child				X	
LA's BEST			X	X	
Nurse-Family Partnership		X			
Preventative Treatment Program	X				
Project GRAD				X	
Quantum Opportunities				X	
SAFE Children				X	
SOAR				X	
STEP			X	X	
Success for All				X	
Teen Outreach Program				X	

Ret = Retained; ED = Economic Disadvantage; Att = Attendance; Ach = Achievement; ParEd = Parent Education

For each exemplary program, Hammond et al (2007) identified specific risk factors that the program addressed and specific methods that the program used. Table 8.1 shows the 19 exemplary programs that address at least one of the risk factors identified in Chapter 6. Many existing programs have successfully lowered dropout rates by focusing on Attendance and Achievement. Very few programs have successfully lowered dropout rates by specifically addressing the needs of students from low-SES families and students from families with low levels of education [43]. The following analysis illuminates the approaches policymakers could take to lower the dropout rate in rural North Carolina for at-risk students.

8.2 Exemplary Dropout Programs that Address Specific Risk Factors

8.2.1 Math Achievement

Students who achieve lower in math than their peers are less likely to graduate. Of the 50 exemplary programs, Hammond et al (2007) found that 14 successfully addressed achievement to reduce high school dropout. Table 8.2 shows the methods these programs used to address the effects of low achievement. One of the most common methods was Academic Support which directly focused on student learning in the classroom. The most popular form of Academic Support was high-quality tutoring by trained tutors. AVID, Check-and-Connect, Coca-Cola Valued Life, LA's Best, SAFE, and Success for All all incorporated tutoring into their programming. SAFE and Success for All used early-age tutoring in reading, so they are likely not the best reference point for policymakers looking to design a ninth grade intervention program. But even these programs reinforce the impact high-quality tutoring can have on student achievement in the short term and graduation in the long term.

Nine of the 14 schools utilized family involvement as they recognized that family interest and dedication could motivate a student to improve their achievement level. The intensity of this involvement varied from parent training in SOAR and Success for All to intensive interventions that included home visits in Check and Connect or Families and Schools Together.

Mentoring and Structured Extracurricular Activities were both used by five programs to increase student interest in school and thereby increase student achievement and likelihood of graduation. Check and Connect, Project GRAD, and Quantum Opportunities used in-school mentors that built relationships with students to ensure that they attended school and completed their work. Big Brothers, Big Sisters used adult community mentors to give participants role models. LA's Best and Coca-Cola Valued Life used field trips and motivational speakers to increase student interest.

Finally, After-School Activities, Life Skills Development, School Environment, Career Development, and Truancy Prevention were all used to lesser degrees in programs that sought to address low achievement by providing students with support systems and increasing their interest in learning. The Teen Outreach program used volunteer opportunities for Life Skills Development, the Career Academy used specific career programming to get participants interested in careers and to understand the relevance of their coursework, and STEP created smaller student cohorts within schools to improve the School Environment and encourage students to want to attend school.

None of these programs are specifically ninth grade intervention programs, but all of them successfully reduced high school dropout rates by focusing on improving student achievement. With the exceptions of SAFE Children and Success for All which are intended for elementary school students, these achievement programs have been designed to apply to diverse populations. A policymaker looking to address the low achievement risk factor through a ninth grade dropout prevention program in rural North Carolina should make the following considerations:

- High quality tutoring with trained tutors can have a significant impact on achievement and graduation rates
- Family participation can improve student motivation, thereby increasing achievement and odds of graduation
- Mentoring and structured extracurricular opportunities appear to increase student interest in school which also increases achievement and graduation rates
- Policymakers should note that programs with after-school activities may not be successful in rural areas due to the challenge of transportation since many students must ride the bus home at the end of the school day

8.2.2 Absences

Students with poor attendance are more likely to drop out as a significant number of absences is often a sign of disengagement from school and the community [23]. Six programs in Table 8.3 specifically address disengagement in students with poor attendance. Across Age and Big Brothers, Big Sisters connect participants with adult community mentors who act as role models, help the students feel valued, and encourage students to attend school. Check and Connect uses a mentor in the school who actively engages in truancy prevention by keeping track of student attendance and developing strong relationships with students who are frequently absent. STEP increases student connection to school by utilizing smaller student cohorts to improve student-student and student-teacher relationships. Career Academy utilizes career skills to increase student interest in school. LA's best provides students with incentives like field trips so long as they maintain a minimum attendance level.

A policymaker looking to address the low attendance risk factor through a ninth grade dropout prevention program in rural North Carolina should make the following considerations:

Table 8.2: Methods Used by Programs that Serve Students with Low Achievement

Program	AfS	AS	CD	F	LSD	M	SE	SEA	TP
AVID		X		X				X	
Big Brothers, Big Sisters	X					X			
Career Academy			X			X			
Check and Connect		X		X		X			X
Coca-Cola Valued Life		X		X				X	
Families and Schools Together				X				X	
LA's Best	X	X		X	X			X	
Project GRAD		X		X		X			
Quantum Opportunities	X	X			X	X		X	
SAFE Children		X		X					
SOAR		X		X	X		X		
STEP							X		
Success For All		X		X					
Teen Outreach	X				X				

AfS = After School; AS = Academic Support; BI =Behavioral Intervention; CM = Case Management; CD = Career Development/Job Training; F = Family Strengthening; LSD = Life Skills Development; M = Mentoring; SE = School Environment; SEA = Structured Extracurricular Activities; TP = Truancy Prevention

- Community mentors have helped at-risk students feel valued and supported, which has contributed to increased interest in school, attendance, and likelihood of graduation
- In-school mentors can also build relationships with students as they have direct knowledge of their students' attendance and personal situations
- Activities like career skill-building and incentives like field trips can increase a students interest in school, thereby increasing attendance
- A program like STEP may not be necessary in rural areas as the student cohort is already small
- Since transportation is an issue, the after-school aspects of Across Ages, Big Brothers, Big Sisters, and LA's Best may not be feasible in rural areas

Table 8.3: Methods Used by Programs that Serve Students with High Absence Rates

Program	AfS	AS	CD	F	LSD	M	SE	SEA	TP
Across Ages	X			X		X		X	
Big Brothers, Big Sisters	X					X			
Career Academy			X						
Check and Connect		X		X		X			X
LA's Best	X	X		X	X			X	
STEP							X		

AfS = After School; AS = Academic Support; BI =Behavioral Intervention; CM = Case Management; CD = Career Development/Job Training; FE = Family Engagement; FS = Family Strengthening; LSD = Life Skills Development; M = Mentoring; SE = School Environment; SEA = Structured Extracurricular Activities; TP = Truancy Prevention

8.2.3 Economic Disadvantage

Economic disadvantage is a strong predictor of high school dropout, yet many of the effects of a low-SES background are already entrenched before a student reaches kindergarten. To that end, Hammond et al (2007) identified only one exemplary program that specifically addressed economic disadvantage. The Nurse-Family Partnership provides first-time, low-income mothers of any age with home visitations during their pregnancies and for two years following birth. The primary participant is the mother, but relatives, friends, parents, and partners are all included in programming. The Nurse-Family Partnership seeks to overcome the typical effects of economic disadvantage by providing parents with the education necessary to support and raise their child in the crucial early years. Positive behavioral and health effects have been observed in the children of these mothers when they reached 15 years of age and are compared to peers in control groups, but the evaluations did not go long enough to confirm the effect on dropout rates. Nevertheless, Hammond et al (2007) included the Nurse-Family Partnership as an exemplary program as it has substantially improved outcomes for students from at-risk, low-income families [43].

The lack of existing programming that specifically addresses economic disadvantage reveals how much of a challenge it is to actually address the effects of low socioeconomic status in adolescence. The Nurse-Family Partnership is likely able to counteract the effects of low-income status because it focuses on early childhood when some of the strongest income effects like lack of parent communication with children first appear. A policymaker looking to address the economic disadvantage risk factor through a ninth grade dropout prevention program in rural North Carolina should make the following considerations:

- A ninth grade intervention program is unlikely to directly combat the effects of economic disadvantage as these effects often emerge in early childhood
- However, policymakers can still serve low-income students with a ninth grade intervention program as these students are still significantly at-risk of dropping out (other programs in this chapter like the Preventative Treatment Program specifically recruited low-income students)

8.2.4 Parent Education

Similar to Economic Disadvantage, only one exemplary program attempted to specifically address the obstacles that arise from having parents with low levels of education. Families and Schools Together recruits families with low-levels of education and 4-12 year old children to participate in family and group therapy and training programs. It seeks to overcome low levels of parent education by providing parents with the skills and knowledge necessary to develop strong family partnerships. Families and Schools Together has been shown to increase parent involvement in their child's education and improve student behavior in classrooms.

The effects of low parental education levels are similar to the effects of Economic Disadvantage because parents with low levels of education are more likely to be economically disadvantaged and because these effects have the strongest impact on a child during the earliest stages of their development. A policymaker looking to address the low parental education risk factor through a ninth grade dropout prevention program in rural North Carolina should make the following considerations:

- A ninth grade intervention program is unlikely to directly combat the effects of low parental education as these effects often emerge in early childhood
- However, policymakers can still target students with parents who have low levels of education for a ninth grade intervention program as these students are still significantly at-risk of dropping out

8.2.5 Retention

Students who are retained are overage for their grade and often show increasing disengagement with their school as they progress [31]. Hammond et al (2007) identified

Table 8.4: Methods Used by Programs that Serve Students who were Retained

Program	AS	CM	CA	CR	FS	LSD	M	SEA
CASASTART	X	X	X		X	X	X	X
Preventative Treatment Program				X	X	X		

AS = Academic Support; CM = Case Management; CA = Court Advocacy; CR = Conflict Resolution; FS = Family Strengthening; LSD = Life Skills Development; M = Mentoring; SEA = Structured Extracurricular Activities; TP = Teen Parent Support

two programs that have successfully targeted and impacted students who are retained: CASASTART and the Preventative Treatment Program [43]. Table 8.4 summarizes the different strategies each program uses. CASASTART is a community-based program that seeks to keep eight to thirteen-year-old students away from drugs and criminal involvement while promoting achievement and attendance in school. CASASTART employs a combination of family services, after-school and summer activities, mentoring, and intensive individual student case management. The Preventative Treatment Program in Montreal targets seven to nine-year-old male students and uses family consultation and small-group peer modeling to reduce disruptive classroom behavior. Both programs have seen reduced drug and crime rates and a greater likelihood of promotion through grade levels.

A policymaker looking to address the retention risk factor through a ninth grade dropout prevention program in rural North Carolina should make the following considerations:

- The current programs that address retention serve remarkably different populations, so they are not ideal for expansion in rural North Carolina
- However, they do illustrate that students who are retained are well-served by comprehensive programs that include both family and community partners

8.2.6 Gender

Male students often find school to be a less supportive, more punishing environment [15]. Hammond et al (2007) did not identify gender as a risk factor as it is not a characteristic that can be controlled or manipulated. However, the exemplary Preventative Treatment Program which focused on behavior and improving student-teacher relationships showed that a program can successfully improve graduation rates by recruiting and focusing on male students.

A policymaker looking to address the gender risk factor through a ninth grade dropout prevention program in rural North Carolina should make the following considerations:

- A program focused on gender has limitations since gender is a demographic characteristic and cannot be manipulated like achievement or attendance
- However, a program may successfully target male students and improve graduation rates if it focuses on the issues faced by male students that contribute to dropout like behavioral issues and poor student-teacher relationships

Chapter 9

Conclusions

9.1 Identifying At-Risk Students Who Could Benefit from a Ninth Grade Intervention Program

The most significant predictors of dropout in rural North Carolina that could be used to identify at-risk students who could benefit from a ninth grade intervention program are, in rank order:

1. Low Math Achievement
2. Poor Attendance
3. Economic Disadvantage
4. Low Parent Education Levels
5. Retention Before Eighth Grade
6. Male Gender

Race, learning disability, LEP status, and reading test scores should not be used to identify at-risk students who could benefit from a ninth grade intervention program as they are either not significant predictors or they are not as good at predicting dropout.

9.2 Methods for a Ninth Grade Intervention Program to Address Significant Risk Factors

In selecting methods for a ninth grade intervention program to address the significant predictors above, policymakers should make the following considerations:

- Low math achievement should be addressed through a program that combines high quality tutoring with family participation, mentoring, or structured opportunities to increase student motivation and achievement levels
- Low attendance should be addressed through community, mentor, or activity-based programs that increase student interest in school and learning
- Retention should be addressed through comprehensive community and family programs that account for retained students' disengagement with school since they are overage for their grade levels
- Male students can be specifically targeted through programs that address issues like behavior

While policymakers can still target and recruit low-income students and students with parents who have low levels of education, the challenges of Economic Disadvantage and Low Parent Education levels emerge in early childhood and are unlikely to be successfully addressed by a ninth grade intervention program. Policymakers should focus on the evidence-based methods above for addressing dropout risk factors.

Appendix A

List of Variables

Table A.1: Student-Level Variables

Variable	Description	Mean	SD
Male	Indicator for male gender	0.51	0.50
Female	Indicator for female gender	0.49	0.50
White	Indicator for identifying as White	0.61	0.49
Asian	Indicator for identifying as Asian	0.01	0.10
American Indian	Indicator for identifying as American Indian	0.02	0.15
Black	Indicator for identifying as Black	0.27	0.45
Hispanic	Indicator for identifying as Hispanic	0.06	0.24
Multiracial	Indicator for identifying as Multiracial	0.19	0.14
Ever AIGM	Ever identified as academically/intellectually gifted in math	0.32	0.47
Ever AIGR	Ever identified as academically/intellectually gifted in reading	0.32	0.47
Ever LDM	Ever identified as learning-disabled in math	0.09	0.28
Ever LDR	Ever identified as learning-disabled in reading	0.12	0.32
Ever ED	Ever identified as economically disadvantaged (eligible for free/reduced price lunch)	0.62	0.49
Ever LEP	Ever identified as Limited English Proficiency	0.06	0.24

Continued on next page

Table A.1 – *Continued from previous page*

Variable	**Description**	**Mean**	**SD**
Retained	Student was retained	0.18	0.38
Days Absent 04	Number of days absent in 2003-2004 school year	6.33	5.85
Days Absent 05	Number of days absent in 2004-2005 school year	6.47	6.23
Days Absent 06	Number of days absent in 2005-2006 school year	7.38	7.26
Days Absent 07	Number of days absent in 2006-2007 school year	8.31	8.41
Days Absent 08	Number of days absent in 2007-2008	8.75	9.32
Parent Education	Seven Levels of Parent Education:		
	Less than HS degree	count 6635	NA
	HS degree	count 23504	NA
	some education after HS	count 4424	NA
	Trade School	count 6970	NA
	Junior College	count 1685	NA
	Four-Year College	count 9030	NA
	Graduate Degree	count 1601	NA
Reading 03	Grade 3 scaled reading score	247.39	9.05
Math 03	Grade 3 scaled math score	252.96	6.19
Reading 04	Grade 4 scaled reading score	251.60	8.86
Math 04	Grade 4 scaled math score	258.65	7.54
Reading 05	Grade 5 scaled reading score	256.32	8.16
Math 05	Grade 5 scaled math score	261.08	9.66
Reading 06	Grade 6 scaled reading score	258.54	8.30
Math 06	Grade 6 scaled math score	354.20	9.57
Reading 07	Grade 7 scaled reading score	261.09	8.73
Math 07	Grade 7 scaled math score	357.49	9.59
Reading 08	Grade 8 scaled reading score	357.39	8.92
Math 08	Grade 8 scaled math score	361.39	7.90

Appendix B

Results from Analysis

Table B.1: Regression Results from Initial Model

School Random Effects: Variance = 0.13, SD = .36

Variable	Coefficient
Intercept	-3.448^{***}
	(0.083)
Male	0.385^{***}
	(0.047)
Asian	-0.841^{**}
	(0.386)
American Indian	0.301^{**}
	(0.135)
Black	0.038
	(0.056)
Hispanic	-0.360^{***}
	(0.104)
Multiracial	-0.225
	(0.180)
LDM	0.312^{***}
	(0.064)
Mild Absence	0.500^{***}
	(0.061)
Increasing Absence	1.032^{***}
	(0.071)

Continued on next page

Table B.1 – *Continued from previous page*

Moderate Absence	1.080***
	(0.072)
Severe Absence	1.292***
	(0.106)
ParEd: HS Degree	−0.426***
	(0.060)
ParEd: Some Ed after HS	−0.620***
	(0.097)
ParEd: Trade School	−0.995***
	(0.096)
ParEd: Junior College	−1.012***
	(0.173)
ParEd: Four-Year College	−1.761***
	(0.123)
ParEd: Graduate Degree	−1.857***
	(0.277)
Observations	53,996
Log Likelihood	−8,154.142
Akaike Inf. Crit.	16,348.280
Bayesian Inf. Crit.	16,526.220

*p<0.1; **p<0.05; ***p<0.01

Table B.2: Regression Results from Initial Model + Retained

School Random Effects: Variance = 0.13, SD = .36

Variable	Coefficient
Intercept	−3.483***
	(0.084)
Male	0.366***
	(0.047)
Asian	−0.840**
	(0.386)
American Indian	0.297**
	(0.136)

Continued on next page

Table B.2 – *Continued from previous page*

Black	0.003
	(0.056)
Hispanic	−0.368***
	(0.104)
Multiracial	−0.232
	(0.181)
LDM	0.109
	(0.069)
Mild Absence	0.485***
	(0.061)
Increasing Absence	0.989***
	(0.072)
Moderate Absence	1.045***
	(0.072)
Severe Absence	1.164***
	(0.107)
ParEd: HS Degree	−0.403***
	(0.061)
ParEd: Some Ed After HS	−0.581***
	(0.097)
ParEd: Trade School	−0.956***
	(0.096)
ParEd: Junior College	−0.967***
	(0.173)
ParEd: Four-Year College	−1.709***
	(0.123)
ParEd: Grad Degree	−1.798***
	(0.277)
Retained	0.652***
	(0.068)
Observations	53,957
Log Likelihood	−8,104.819
Akaike Inf. Crit.	16,251.640
Bayesian Inf. Crit.	16,438.450

*p<0.1; **p<0.05; ***p<0.01

Table B.3: Regression Results from Initial Model + ED

School Random Effects: Variance = 0.13, SD = .36

Variable	Coefficient
Intercept	-4.233^{***}
	(0.104)
Male	0.398^{***}
	(0.047)
Asian	-0.989^{**}
	(0.386)
American Indian	0.133
	(0.135)
Black	-0.171^{***}
	(0.057)
Hispanic	-0.541^{***}
	(0.104)
Multiracial	-0.381^{**}
	(0.180)
LDM	0.275^{***}
	(0.064)
Mild Absence	0.445^{***}
	(0.061)
Increasing Absence	0.916^{***}
	(0.072)
Moderate Absence	0.959^{***}
	(0.073)
Severe Absence	1.146^{***}
	(0.106)
ParEd: HS Degree	-0.342^{***}
	(0.060)
ParEd: Some Ed After HS	-0.447^{***}
	(0.098)
ParEd: Trade School	-0.728^{***}
	(0.097)
ParEd: Junior College	-0.702^{***}
	(0.174)

Continued on next page

Table B.3 – *Continued from previous page*

ParEd: Four-Year College	-1.219^{***}
	(0.128)
ParEd: Grad Degree	-1.170^{***}
	(0.281)
ED	1.008^{***}
	(0.076)
Observations	53,996
Log Likelihood	−8,053.393
Akaike Inf. Crit.	16,148.780
Bayesian Inf. Crit.	16,335.610

*p<0.1; **p<0.05; ***p<0.01

TABLE B.4: Regression Results from Initial Model + Math Cluster

School Random Effects: Variance = 0.12, SD = .35

Variable	Coefficient
Intercept	-2.773^{***}
	(0.098)
Male	0.397^{***}
	(0.047)
Asian	-0.757^{*}
	(0.387)
American Indian	0.110
	(0.136)
Black	-0.255^{***}
	(0.058)
Hispanic	-0.425^{***}
	(0.104)
Multiracial	-0.317^{*}
	(0.181)
LDM	−0.052
	(0.068)
Mild Absence	0.434^{***}
	(0.061)

Continued on next page

Table B.4 – *Continued from previous page*

Increasing Absence	0.898***
	(0.072)
Moderate Absence	0.937***
	(0.073)
Severe Absence	1.076***
	(0.106)
ParEd: HS Degree	−0.293***
	(0.061)
ParEd: Some ED After HS	−0.383***
	(0.098)
ParEd: Trade School	−0.703***
	(0.097)
ParEd: Junior College	−0.700***
	(0.174)
ParEd: Four-Year College	−1.201***
	(0.125)
ParEd: Graduate School	−1.072***
	(0.281)
Math: Below Average	−0.250***
	(0.067)
Math: Slightly Below Average	−0.626***
	(0.074)
Math: Slightly Above Average	−1.011***
	(0.086)
Math: Above Average	−1.543***
	(0.116)
Math: Considerably Above Average	−2.238***
	(0.220)
Observations	53,996
Log Likelihood	−7,979.910
Akaike Inf. Crit.	16,009.820
Bayesian Inf. Crit.	16,232.240

*p<0.1; **p<0.05; ***p<0.01

Table B.5: Regression Results from Initial Model + Reading Cluster

School Random Effects: Variance = 0.12, SD = .35

Variable	Coefficient
Intercept	-2.869^{***}
	(0.103)
Male	0.325^{***}
	(0.047)
Asian	-0.932^{**}
	(0.387)
American Indian	0.152
	(0.136)
Black	-0.184^{***}
	(0.057)
Hispanic	-0.464^{***}
	(0.104)
Multiracial	-0.273
	(0.181)
LDM	-0.004
	(0.070)
Mild Absence	0.476^{***}
	(0.061)
Increasing Absence	0.959^{***}
	(0.071)
Moderate Absence	1.018^{***}
	(0.072)
Severe Absence	1.170^{***}
	(0.106)
ParEd: HS Degree	-0.312^{***}
	(0.061)
ParEd: Some Ed After HS	-0.417^{***}
	(0.098)
ParEd: Trade School	-0.754^{***}
	(0.097)
ParEd: Junior College	-0.768^{***}
	(0.174)

Continued on next page

Table B.5 – *Continued from previous page*

ParEd: Four-Year College	-1.313^{***}
	(0.126)
ParEd: Grad Degree	-1.257^{***}
	(0.280)
Below Average	-0.174^{**}
	(0.074)
Slightly Below Average	-0.427^{***}
	(0.077)
Slightly Above Average	-0.699^{***}
	(0.084)
Above Average	-1.120^{***}
	(0.101)
Considerably Above Average	-1.569^{***}
	(0.159)
Observations	53,996
Log Likelihood	−8,044.054
Akaike Inf. Crit.	16,138.110
Bayesian Inf. Crit.	16,360.520

*p<0.1; **p<0.05; ***p<0.01

Bibliography

[1] J. Johnson, D. Showalter, R. Klein, and C. Lester. Why rural matters 2013-2014. *Rural School and Community Trust*, May .

[2] Sara Battin-Pearson. Predictors of early high school dropout: A test of five theories. *Journal of educational psychology*, 92(3):568–582, 2000.

[3] Ruth Curran Neild, Scott Stoner-Eby, and Frank Furstenberg. Connecting Entrance and Departure: The Transition to Ninth Grade and High School Dropout. *Education and Urban Soceity*, 40(5):543–569, July 2008.

[4] Michael Arnold, John Newman, Barbara Gaddy, and Ceri Dean. A Look at the Condition of Rural Education Research: Setting a Direction for Future Research. *Journal of Research in Rural Education*, 20(6), 2005.

[5] O. Yeboah, P.E. Faulkner, and G. Appiah-Danquah. North Carolina High School Dropout Rates: An Econometric Analysis. 2010.

[6] Solving the Graduation Crisis: Identifying and Using School Feeder Patterns in your Community, 2013.

[7] E. Pascarella and P. Terenzini. Predicting freshman persistence and voluntary dropout decisions from a theoretical model. *The Journal of Higher Education*, 51 (1):60–75, 1980.

[8] Henry Levin, Clive Belfield, Peter Muennig, and Cecilia Rouse. The Costs and Benefits of an Excellent Education for All of America's Children, 2006.

[9] Model Dropout Prevention Programs, 2011.

[10] Policy Regarding dropout prevention and students at-risk, 2004.

[11] J. Owen, J. Rosch, C Muschkin, J. Alexander, and C. Wyant. Dropout prevention: Strategies for improving high school graduation rates. *Center for Child and Family Policy*, 2008.

[12] E. Stearns and E. Glennie. When and Why Dropouts Leave High School. 2006.

[13] D. Silver, M. Saunders, and E. Zarate. What Factors Predict High School Graduation in the Los Angeles Unified School District, 2008.

[14] J.L. Jordan, G. Kostandini, and E. Mykerezi. Rural and urban high school dropout rates: Are they different? *Journal of Research in Rural Education*, 21(12):1–21, 2012.

[15] Robert Vallerand, Michelle Fortier, and Frederic Guay. Self-determination and persistence in a real-life setting: Toward a motivational model of high school droput. *Journal of Personality and Social Pscyhology*, 72(5):1161–1176, May 1997.

[16] Kelsey Sheehy. Graduation Rates Dropping Among Native American Students. *US News*, June 2013.

[17] Bryan Griffin. Academic Disidentification, Race, and High School Dropouts. *The High School Journal*, 85(4):71–81, May 2002.

[18] Magnus Lofstrom. Why are Hispanic and African-American Dropout Rates So High? In *IZA Discussion Paper*, December 2007.

[19] Russell Rumberger. Dropping out of High School: The Influence of Race, Sex, and Family Background. *American Educational Research Journal*, 20(2):199–220, 1983.

[20] B.R. Subdei and B. Johnson. Predicting High School Graduation and Dropout Using a Hierarchical Generalized Linear Model Approach. 2007.

[21] Daphna Bassok and Loeb Susanna. Early Childhood and the Achievement Gap. In Ladd Helen and Goertz Margaret, editors, *Handbook of Research in Education Finance and Policy*, pages 510–527. Routledge, New York, 2 edition, 2015.

[22] T.C. West. Just the Right Mix: Identifying Potential Dropouts in Montgomery County Public Schools Using an Early Warning Indicators Approach, 2013.

[23] Michel Janosz, Isabelle Archambault, Julien Morizot, and Linda Pagani. School Engagement Trajectories and Their Differential Predictive Relations to Dropout. *Journal of Social Issues*, 64(1):21–40, 2008.

[24] R.S. Subedi and M. Howard. Predicting high school graduation and dropout for at-risk students: A multilevel approach to measure school effectiveness. *Advances in Education*, 2(1):11–17, 2013.

[25] Larry Kortering and Patricia Braziel. School Dropout among Youth with and without Learning Disabilities. *Career Development and Transition for Exceptional Individuals*, 21(1):61–74, April 1998.

[26] Adrienne Ingrum. High School Dropout Determinants: The Effect of Poverty and Learning Disabilities. *The Park Place Economist*, 14(1):73–79, 2006.

[27] Patricia Gandara. Review on Research on the Instruction of Limited English Proficient Students: A Report to the California Legislature. Technical report, 1997.

[28] Anne Driscoll. Risk of High School Dropout among Immigrant and Native Hispanic Youth. *International Migration Review*, 33(4):Winter 1999, January 857.

[29] Shane Jimerson, Gabrielle Anderson, and Angela Whipple. Winning the Battle and Losing the War: Examining the Relation Between Grade Retention and Dropping Out of High School. *Psychology in the Schools*, 39(4), 2002.

[30] Shane Jimerson and Amber Kaufman. Reading, Writing, and Retention: A Primer on Grade Retention Research. *The Reading Teacher*, 56(7):622–635, April 2003.

[31] Melissa Roderick. Grade Retention and School Dropout: Investigating the Association. *American Educational Research Journal*, 31(4):729–759, 1994.

[32] Linda Tilly, Apreill Curtis Hartsfield, Lisa Parrish, Debra Miller, Valerie Salley, Linda O'Neal, Pam Brown, and Edwina Chappell. The Rural South: Listening to Families in Alabama, Kentucky, and Tennessee, 2004.

[33] Allan Porowski and Caitlin Howley. Dropout Prevention: Challenges and Opportunities in Rural Settings, 2013.

[34] Patricia Harde and John Marshall Reeve. A motivational model of rural students' intention to persist in, versus drop out of, high school. *Journal of Educational Psychology*, 95(2):347–356, June 2003.

[35] Jay Smink and Mary Reimer. Rural School Dropout Issues: Implications for Dropout Prevention Strategies and Programs.

[36] Rebecca Droessler Mersch. *Student Academic Achievement in Rural vs. Non-Rural High Schools in Wisconsin*. Dissertation, Edgewood College, 2012.

[37] David Lambert, John Gale, and David Hartley. Substance Abuse by Youth and Young Adults in Rural America. *The Journal of Rural Health*, 24(3):221–228, 2008.

[38] Melissa Azur, Elizabeth Stuart, Constantine Frangakis, and Philip Leaf. Multiple imputations by Chained Equations: What is it and how does it work? *International Journal of Methods in Psychiatric Research*, 20(1):40–49, March 2012.

[39] Katherine Ralston, Constance Newman, Annette Clauson, Joanne Guthrie, and Jean Buzby. The National School Lunch Program: Background, Trends, and Issues. Technical Report Economic Research Report 61, USDA, July 2008.

[40] JA Hartigan and MA Wong. A K-Means Clustering Algorithm. *Journal of the Royal Statistical Society Series C (Applied Statistics)*, 28(1):100–108, 1979.

[41] Guang Guo and Hongxin Zhao. Multilevel Modeling for Binary Data. 26:441–462, 2000.

[42] Chris Burritt and Timothy Homan. Rural North Carolina Absorbing Huge Surge of Hispanics. *The Seattle Times*, March 2011.

[43] Cathy Hammond, Dan Linton, Jay Smink, and Sam Drew. Dropout Risk Factors and Exemplary Programs: A Technical Report, May 2007.