2017

Testing the Ability of Two Series of Models to Predict High School Graduation Status

David T. Marshall
Virginia Commonwealth University

TESTING THE ABILITY OF TWO SERIES OF MODELS TO PREDICT HIGH SCHOOL GRADUATION STATUS

A dissertation submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy at Virginia Commonwealth University

by

David Thomas Marshall
Bachelor of Arts, Virginia Polytechnic Institute and State University, 2002
Master of Arts, Virginia Polytechnic Institute and State University, 2009
Master of Arts, La Salle University, 2012

Director: Dr. James H. McMillan
Interim Associate Dean for Academic Affairs and Professor
VCU School of Education

Virginia Commonwealth University
Richmond, Virginia
April 2017

Acknowledgement

Writing a dissertation is not a simple task, and it would be a huge understatement to claim that I could have made it this far without the love, support, collegiality, and friendship of many individuals. It would be difficult, if not impossible, to properly thank everyone in an acknowledgement section, however there are a few individuals who deserve some special words of recognition. First and foremost, I would like to thank my amazing wife, who has been supportive of my pursuits – whatever they may be – since the day I met her. This doctoral program was no exception, and I am beyond fortunate to have found such a loving and wonderful human being with which to share my life.

My dissertation committee has been amazing. I would like to thank my advisor and committee chair, Dr. Jim McMillan, for his constant guidance, encouragement, positivity, and unyielding support throughout the dissertation process and throughout my entire doctoral program. I would like to thank Dr. Lisa Abrams for sharing her expertise and for the numerous encouraging conversations we have had, especially over the course of the past year. I would like to thank Dr. Michael Broda for spending countless hours with me talking through my ideas and for working with me throughout my data analysis. And finally, I would like to thank Dr. Karren Bailey. It was her willingness to allow me to work with her and her department for my externship that planted the seed for this study in the first place. Her expertise, insights, and support have been tremendous as well.

There are other individuals who did not work directly with me on my dissertation that deserve recognition for their role in my doctoral program. Without the support of these individuals, I would not be where I am today: Ann Cherry, Dr. Henry Clark, Dr. Terry Dozier, soon-to-be-Dr. Thy Nguyen, Dr. LaRon Scott, Dr. Jesse Senechal, Dr. Kurt Stemhagen, Dr. Chriss Walther-Thomas, and Dr. Guofang Wan.

The experiences that I had teaching in traditional public and public charter schools in Philadelphia, Pennsylvania were foundational for this work. I would especially like to thank a handful of former colleagues, without which I would not have become the educator that I did: Aileen Burr, Rick Garlock, Cara Harvey (formerly Cammalleri), the late Napoleon Horton, David Mandell, and Tyrone Taylor. I thank each of you for the countless days and hours spent helping me to reflect on my teaching experiences and for helping me to become a better educator.

Finally, I would like to thank my family and friends who have been supportive of me throughout this process. I would especially like to acknowledge my mother, Melinda Marshall, who has devoted more than 30 years of her life to educating first and second grade students in York County, Virginia. Her students were blessed to call her their teacher, and I have been beyond blessed to call her my mom.

Dedication

On April 16, 2007, the world changed for me and for many of the people I know. My alma mater suffered a tremendous tragedy and 32 lives ended far too soon. Brian Bluhm, a graduate student in engineering, was among the fallen. Brian was my brother's roommate and best friend, and he was one of the most decent humans I have had the honor to count among my friends. For all of the sorrow that came from this event, several positive lessons emerged, and the most salient among these was a reminder that life is short. It was around this time that I re-enrolled at Virginia Tech to begin my career in education. In the decade that has passed since that awful day, I have tried to live each day to the fullest and live my life with a sense of urgency in the service of our nation's students. I dedicate this work to the memory of Brian and the other 31 individuals who lost their lives on that brisk April morning. We will prevail.

Table of Contents

List of Tables

List of Figures

Abstract

TESTING THE ABILITY OF TWO SERIES OF MODELS TO PREDICT HIGH SCHOOL GRADUATION STATUS

By David T. Marshall, Ph.D.

A dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy at Virginia Commonwealth University

Virginia Commonwealth University, 2017

Director: Dr. James H. McMillan
Interim Associate Dean for Academic Affairs, Professor, Foundations of Education
School of Education

The purpose of this study was to create and test two series of predictive models aimed at projecting high school graduation status. Secondary data were obtained in partnership with an urban school district. All of the predictor variables included in the models tested in this study were academic and nonacademic variables that were found to be significant predictors of high school graduation in previous empirical work. In the first series of models tested, individual academic and nonacademic variables were tested together along with school-level variables. Eighth and ninth grade variables were tested separately to avoid multicollinearity issues. The second series of models tested included similar individual-level academic and nonacademic variables, along with community-level predictors to analyze their ability to predict high school

graduation status. Logistic regression and multilevel logistic regression analyses were conducted to analyze the data. The model including community-level predictors yielded a pseudo R-squared value of .40, approximating that 40% of the variance was explained by the predictors in the model. Most of the individual predictors included in the models yielded findings similar to those found in previous literature on high school graduation status projection; however, this was not true for all of the predictor variables included. These differences highlight the tension that can exist between generalizability and local specificity. Significant findings from studies utilizing large nationally-representative longitudinal datasets and other large data sources do not always generalize to settings with samples that differ demographically. This study represents a first step in a line of research aimed at developing a better understanding of high school graduation status, particularly in challenging school contexts

I. INTRODUCTION

Background for the Study

The pinnacle moment in a K-12 education is high school graduation. Those who graduate from high school have many doors opened for them. They are better positioned to continue their education and have a wider array of employment opportunities than someone who does not achieve this milestone (Balfanz, Fox, Bridgeland, & McNaught, 2009; Kim, 2013; Rumberger, 2011). An individual who fails to earn a high school diploma will only earn two-thirds of what a high school graduate will earn in his or her first job (Balfanz et al., 2009; Breslow, 2012). A survey of high school graduates in 2015 found that individuals who earned a high school diploma who were not enrolled in college were more than twice as likely to be a part of the labor force as individuals who failed to complete high school (Bureau of Labor Statistics, 2016). Individuals who do not complete high school are likely to earn substantially less income, are less likely to vote, have shorter life spans and live in poorer health, and are more likely to engage in crime and become incarcerated than those who finish high school (Orfield, 2006; Rumberger, 2011). There are also larger societal costs associated with students failing to complete high school. Individuals who drop out of school are less likely to participate in the economy, pay taxes, or vote; at the same time, they are more likely to consume government resources by accessing services, commit crime, or experience adverse health outcomes (Levin,

1972; Rumberger, 2011). Every time a student fails to complete high school, the government brings in less money in revenue, and spends more money on services related to poverty, public health, and the criminal justice system, and has fewer resources left over for other public services – public education among them.

In 2012 the national cohort graduation rate in the United States eclipsed 80 percent for the first time, representing more than a 10-point increase since the early 1990s (National Center for Education Statistics, 2016). At the conclusion of the 2015-2016 school year, the national graduation rate reached 83.6%. However, this graduation rate does not universally exist across all school districts in the United States. Some school districts have graduation rates that far surpass this figure. However, graduation rates are often much lower in urban and low income schools (Rumberger, 2011; Swanson, 2006). Even if the national rate was uniform, one in five students who attends high school would still fail to graduate. When this is considered in conjunction with the diminished possibilities that those who fail to complete high school face, the phenomenon of high school dropout is what Durkheim (1982) would deem a *social fact*. He defined a social fact as "any way of acting, whether fixed or not, capable of exerting over the individual an external constraint" (Durkheim, 1982, p. 59). Being a high school dropout is certainly not a fixed condition; individuals who drop out often return to school at some point, and approximately half complete their education by their mid-twenties (Entwisle, Alexander, & Olsen, 2004; Rumberger, 2011; Rumberger & Rotermind, 2004). However, it is equally true that those who do not complete high school are clearly disadvantaged in life. Having the tools to predict high school graduation status becomes a necessary part of intervening and improving the academic outcomes of students who are at-risk of not earning a high school diploma.

Rationale for the Study

The purpose of this study was to create and test two series of predictive models that project student high school graduation status. These models contribute to the literature on high school graduation status and models designed to predict high school graduation status in two distinct ways. Both series of models include a set of predictors that previous empirical studies have found to predict high school graduation status. The findings from this study build on previous literature by examining the predictive ability of variables found to be significant in previous studies work when they are included together in the same model within a new context. Previous literature has explored relationships between community-related variables and whether or not students complete high school, and previous studies testing models predicting high school graduation status have employed multilevel modeling. However, all of the studies included in this review of the literature that employed multilevel modeling nested students (Level 1) in schools (Level 2). In the second series of models, students (Level 1) are nested in home address zip codes (Level 2) to approximate predictive ability of community-level variables in the context of high school graduation. This study represents the first study in a line of research focusing on high school graduation status. The findings from this study form the foundation on which the next steps in this line of work can build.

Overview of the Literature

This review of the literature explores predictors of high school graduation status that previous empirical work has found to be useful in predicting high school graduation status in the United States. Studies using samples outside of the United States were excluded from the search, as were studies on interventions aimed at increasing high school graduation. There are many studies in the literature that examine the relationship between attitudinal variables, such as

student goal-setting behaviors, self-efficacy, and self-perceptions and graduation status (Lee, Cornell, Gregory, & Fan, 2011; Parr & Bonitz, 2015), as well as studies that examine the relationship between behavioral deviance variables, such as teenage alcohol and drug use, criminal behavior, and early sexual activity, and graduation status (Bayliss et al., 2011; Ellickson et al., 1996; Henry, Knight, & Thornberry, 2012; Rumberger, 2011). This review excludes such work from the review of the literature for practical purposes; these are variables that would not be found in an administrative dataset, and that was the source of the data used in this study.

Operationalizing graduation status. The most important concept in need of operationalizing for a study examining a model predicting high school graduation status is *high school graduation status*. Most studies in the literature dichotomize the outcome variable of graduation status as either (1) graduates and non-graduates, or (2) completers and non-completers. This is far from a simple designation; rather it can be quite complex. For one, there are more than two possible student outcomes; these include (1) earning a high school diploma; (2) dropping out; (3) still enrolled in school but not having earned enough credits to graduate after four years; (4) transferring to another school; (5) earning an alternative credential such as a passing score on the Graduate Educational Development (GED) test; and (6) being deceased (Rumberger, 2011). There may be several types of diplomas that students can earn depending on where a study is conducted geographically, but these are typically collapsed into a generic "graduate" classification for predictive studies. Students enrolled in special education programs who earn modified diplomas are almost always counted as graduates. Special education students certainly appear among those classified as dropouts; it would be inappropriate to not count them among the graduates. Most studies exclude the deceased and those who transfer out of the population from being included in analyses. Studies tend to differ in how they classify students

who are still enrolled in school, as well as individuals who earn alternative credentials. Studies also differ in how they treat students who transfer in to the population who began their high school careers elsewhere. Failing to adequately operationalize and describe the outcome variable can limit the extent to which proper inferences can be made from findings.

One of the challenges in conducting research on high school completion is in operationalizing what this concept means. High school graduation status is typically measured in one of two ways (Cratty, 2012; Orfield, 2006; Rumberger, 2011). Most commonly, graduation status is calculated as a graduation cohort rate, which is the total number of students who enter the ninth grade together. This number becomes the denominator in a proportion formula. The number of students within this cohort who earn diplomas becomes the numerator; the rate is calculated with simple division. Students who transfer out of the population being studied are typically removed from the cohort. Consider an example. If 115 students entered the ninth grade together, 15 students transferred out of the school district prior to graduating, and 90 students earned diplomas, the cohort graduation rate would be 90% (90/100). Graduation status is sometimes calculated as an event dropout rate (Rumberger, 2011). This rate represents the proportion of a high school’s population that dropped out in a given year. This is calculated by dividing the total number of students who drop out by the total population of the high school. For example, if a high school has 100 students and five students drop out in a given school year, that high school would have an event dropout rate of 5% (5/100).

In many facets of life, understanding the positive result rate is enough to understand the negative result rate. For example, if a basketball player makes 80% of her free throws, one can infer from that statistic that she also missed 20% of her attempts. The same is not true with high school graduation status. High school graduation rates and high school dropout rates are not

necessarily the inverse of each other. This is the due to the fact that there are more than two possible outcomes for students after four years, including the possibility of still being enrolled due to retention or earning an alternative credential such as a GED.

Predictors of high school graduation status. The factors that lead to a student's graduation status are complex. There are three general categories of predictor variables that are discussed in this review and included in the models tested in this study: (1) academic predictors; (2) nonacademic predictors related to student engagement; and (3) nonacademic demographic predictors. Five types of academic predictor variables were found to be significantly related to high school graduation status. No matter how they were operationalized, student grades (Doren, Murray, & Gau, 2014; Mac Iver & Messel, 2013; Rumberger, 2011), failed courses (Allensworth & Easton, 2005; Mac Iver & Messel, 2013; Mac Iver & Messel, 2012; Rumberger, 2011); and retention (Alexander, Entwisle, & Horsey; 1997; Jimerson, Anderson, & Whipple, 2002; Neild, Stoner-Eby, & Furstenberg, 2008; Stroup, 1972; Swanson, 2006) were all found to be significant predictors of high school graduation status. Course enrollment patterns have also been found to be a significant predictor of high school graduation status (Cratty, 2012; Goldschmidt & Wang, 1999; Soland, 2013). Interestingly, standardized test scores were among the most often used academic predictors, and the least reliable. Some studies found standardized test scores to be significant predictors of graduation status, whereas others did not (Barrington & Hendricks, 1989; Cratty, 2012; Hernandez, 2011; Mac Iver & Messel, 2013).

Three types of nonacademic predictors related to student engagement have also been found to be significant predictors of high school graduation status. Regardless of how they were operationalized, student attendance (Bayliss et al., 2011; Cratty, 2012; Mac Iver & Messel, 2013; Mac Iver & Messel, 2012; Rumberger, 2011) and student behavior (Bayliss et al., 2011; Cratty,

2012; Doren, Murray, & Gau, 2014; Jimerson et al., 2000; Mac Iver & Messel, 2013) were found to be significant predictors of high school graduation status. Participation in extracurricular activities and sports has also been found to be related to increased graduation rates (Mahoney & Cairns, 1997; Rumberger, 2011).

Finally, nonacademic demographic variables have been found to be significant predictors of high school graduation status. Particular emphasis has been placed on exploring disparities that exist between students by race (Cratty, 2012; Jordan, Lara, & McPartland, 1996; Lee et al., 2012; Losen, 2006; Orfield, 2006; Parr & Bonitz, 2015; Swanson, 2006), gender (Cratty, 2012; Jordan et al., 1996; Rumberger, 2011), and socioeconomic status (Cratty, 2012; Rumberger, 2006; Swanson, 2006) and their propensity to complete high school. No matter how socioeconomic status was operationalized, it was consistently found to be a significant predictor of high school graduation status. Race and gender were less reliable predictors of graduation status. In some studies, a relationship between these variables and graduation status was found; in others no relationship was found between these demographic predictors and a student's odds of completing high school. Family structure (Cratty, 2012; Doren et al., 2014; Rumberger, 2006; Song, Benin, & Glick, 2012) and student mobility (Rumberger, 2011; Rumberger & Larson, 1998) were also consistently found to predict high school graduation status; students from families with two parents in the home and students who did not change schools, except for promotion to middle or high school, were consistently found to graduate from high school at higher rates than peers who did not fit these criteria.

Previous predictive models. Over the last four decades, studies have been conducted to understand the relationships between various predictor variables and high school graduation status. The literature on models designed to predict student graduation status has examined

nationally representative longitudinal datasets (e.g. Parr & Bonitz, 2015), statewide graduation cohorts (e.g. Cratty, 2012), and administrative data derived from a single school district (e.g. Mac Iver & Messel, 2013). A search for peer-reviewed literature based in the United States that explored the utility of models predicting high school graduation status found 24 studies that fit the criteria. These studies reported their findings in three general ways. First, almost all of the studies reported the predictive ability of the individual independent variables included in the models. Second, many of the studies reported the amount of the variance explained by the model as a pseudo R-squared value. Third, many of the models also reported the degree to which they correctly classified cases in three ways: (1) the percent of graduates or completers correctly classified; (2) the percent of dropouts or non-completers correctly classified; and (3) the overall rate at which cases were correctly classified.

Over 80% of the studies included in this review of the literature employed either discriminant function analysis (e.g. Lloyd, 1978) or logistic regression analysis (e.g. Ou & Reynolds, 2008) as their method of analysis. Discriminant function analysis is appropriate for use to predict group membership using multiple predictor variables (Dattalo, 2010). However, this method of analysis assumes multivariate normality and only continuous predictor variables can be used (Dattalo, 2010; Tabachnick & Fidell, 2013). Logistic regression is another statistical procedure that has been used to analyze the utility of models predicting high school graduation status (Cratty, 2012; Hernandez, 2011; Mac Iver & Messel, 2013; Neild et al., 2008; Rumberger & Larson, 1998). Logistic regression is also appropriate for use when predicting group membership with multiple predictor variables; however, it can be used with categorical and binary predictor variables, as well as with continuous predictors (Hosmer & Lemeshow, 2000; Tabachnick & Fidell, 2013). Logistic regression is necessary to use in lieu of linear regression

for categorical outcomes because it corrects for violations of the assumptions of normal distribution and linearity (Tabachnick & Fidell, 2013). Three of the studies included in this review employed multilevel logistic regression analysis to analyze similar predictive models (e.g. Mac Iver & Messel, 2013). Multilevel modeling is appropriate for use when nested data is present (Raudenbush & Byrk, 2002). For models predicting high school graduation status, this usually features individual students (Level 1) nested in schools (Level 2). This is the preferred type of analysis to use if there are at least ten units for Level 2.

In summary, several factors have been found to contribute to whether or not a student completes high school. Academic predictor variables have been found to be associated with high school graduation status, as have nonacademic variables related to student engagement, and nonacademic demographic variables. Models designed to predict high school graduation status that include these predictor variables should employ logistic regression analysis, ideally using multilevel modeling if there are enough units at the second level of analysis.

Research Questions

This study was guided by the following research questions.

1. To what extent do individual academic variables predict high school graduation status?
    a. To what extent do individual-level academic variables predict student high school graduation status?
    b. To what extent do school building-level variables predict high school graduation status?
2. To what extent do nonacademic variables predict high school graduation status?
    a. To what extent do individual levels of student engagement predict high school graduation status?

b. To what extent do demographic variables predict high school graduation status?

c. To what extent do community-level variables predict a student's high school graduation status?

Design and Methods

Using a mid-sized urban school district's administrative longitudinal data, two series of models were tested to assess their ability to predict student high school graduation status. The first series of models included eighth and ninth grade academic, nonacademic, and demographic predictors, as well as school-level predictors. Eighth and ninth grade variables were tested separately, and logistic regression analyses were performed on each set of models. The second series of models included demographic and ninth grade variables, along with community-level variables. Multilevel logistic regression and logistic regression analyses were conducted to test these models.

For both series of models, models were built and tested, progressively adding additional clusters of variables. This demonstrated the predictive ability of variables tested together with different sets of variables, creating the possibility of a more nuanced understanding of the predictive ability of the variables explored in this study. The models were compared with each other, as well as with other predictive models in the literature in two ways. First, the models were evaluated on the basis of their pseudo R-squared values. Models were also evaluated in terms of their ability to correctly classify cases. The overall classification rates and non-graduate classification rates for each series of models were compared with each other, as well as with the findings from other predictive models in the literature.

Definition of Terms

Graduate - In this study, a graduate is an individual who has earned a high school diploma

within four years of enrolling in high school.

<u>High school completion</u> – In this study, this refers to having either earned a high school diploma or having earned an alternate credential such as the GED.

<u>High school completion rate</u> - In this study, this refers to a four-year cohort completion rate. Students who earn any diploma or alternate credential are considered to be completers. Students who are still enrolled after four years are considered to be non-completers.

<u>High school graduation</u> – In this study, this refers to having earned a high school diploma, including advanced, International Baccalaureate, standard, and modified diplomas.

<u>High school graduation rate</u> - In this study, this refers to a four-year cohort graduation rate. Students who earn high school diplomas in four years are considered to be graduates. Students earning alternate credentials and students who are still enrolled in school after four years are counted as non-graduates in this calculation. By law of the state in which this study was conducted, students with an Individual Education Plan (IEP) are afforded six years to complete their program of study and still be counted as on time graduates.

<u>High school graduation status</u> – In this study, this describes a student's status four years after enrolling in high school. In this time, students will: (1) have earned a diploma; (2) have earned an alternate credential such as a GED; (3) still be enrolled in high school; (4) have dropped out of high school; (5) have transferred out of the school district; or (6) died.

<u>Non-graduate</u> – In this study, a non-graduate is an individual who has not earned a high school diploma within four years of enrolling in high school.

<u>On-time graduation</u> – In this study, this refers to having graduated high school in four years. Students who have an IEP are afforded six years to complete their program of study, and

those who do so are considered to have completed school on time.

II. REVIEW OF LITERATURE

Overview of the Review of Literature

Prior research on developing and testing statistical models to predict student high school graduation status is organized in four sections. This chapter begins with a brief discussion of how high school graduation status has been operationalized in previous work, and how it is currently being operationalized in federal education policy. Next, previous empirical work investigating variables that have been demonstrated to be predictors of high school graduation status are presented, including those related to both academic and nonacademic student experiences and traits. Third, previously tested models that predict high school graduate status are explored. The findings, limitations, and utility of these models are discussed. Finally, the methodologies employed in previous predictive models are explored. The merits and limitations of discriminant function analysis, logistic regression analysis, and multilevel modeling are examined, providing the foundation for the methodology employed in this study.

Literature Review Methodology

Several national databases were searched to retrieve previous literature, including ERIC via ProQuest, Sociological Abstracts, Education Research Complete, and Google Scholar. Each of six journals published by the American Educational Research Association was also searched. Search terms used included “high school,” “completion,” “graduation,” “dropout,”

“predictor(s),” “factors,” and “cause(s).” Peer-reviewed literature published between 1996 and 2016 that included the above-listed terms in abstracts was included in the review. Other inclusion and exclusion rules were employed as well. Studies that explored the effectiveness of specific interventions aimed at influencing high school graduation outcomes were excluded since interventions are not the focus of this study. Also, studies that were based outside of the United States were also excluded. A number of studies in the literature focus on how latent constructions like motivation, student engagement, and academic self-efficacy impact student graduation outcomes. These studies were also excluded from the literature search since variables corresponding to these constructs were unlikely to be found in an administrative dataset. Finally, there are a number of studies that examine the relationship between high school graduation status and certain behavioral variables, such as teenage alcohol and drug use, early sexual activity, and teenage pregnancy. These are also not variables that would be found in administrative data and these studies were excluded as well.

Additional searches of ERIC via ProQuest and Google Scholar were conducted using the terms “high school,” “graduation,” “dropout,” “prediction,” and “model,” with no date range. Of interest was to locate and explore an inclusive collection of the literature exploring models predicting high school graduation status. See Table 1 for a list of the studies on high school graduation status prediction included in this review, listed chronologically with an indication of the source of the data used and sample size for each study.

Table 1

Studies Predicting High School Graduation Status

Author(s), Year	Data source	*N* of Study
Stroup & Robins, 1972	St. Louis long. data, AA males*	223
Lloyd, 1978	California (no further description)	1562
Barrington & Hendricks, 1989	Two high school cohorts	651
Astone & McLanahan, 1991	HSB 80, 82, 84, 86	1968
Ensminger & Slusarcick, 1992	Long. data from Chicago Woodlawn	1242
McNeal, 1995	HSB 80	14249
Alexander et al., 1997	Cohort from 20 elem. schools, Baltimore	790
Rumberger & Larson, 1998	NELS 88, 92	11671
Goldschmidt & Wang, 1999	NELS 88, 90, 92	25000**
Tobin & Sugai, 1999	Single high school, urban, northwest US	526
Batin-Pearson et al., 2000	Long. data from Seattle, high-crime areas	808
Jimerson et al., 2000	Long. data from Minn., low-SES	177
McWilliams, Everest, & Bass, 2000	Senior cohort, rural low-SES district SE US	80
Owens et al., 2001	Rural Florida county cohort	208
Rumberger & Palardy, 2005	NELS 88	14199
Jimerson & Ferguson, 2007	Single rural school district, Western US	72
Neild et al., 2008	PELS (Philadelphia)	1457
Ou & Reynolds, 2008	Chicago Long. Study	1286
Hernandez, 2011	NLSY-79	3975
Song, Benin, & Glick, 2012	NELS 88, 90, 92	21420
Cratty, 2012	North Carolina statewide cohort	68401
Mac Iver & Messel, 2013	Two cohorts, Baltimore City	14541
Soland, 2013	NELS 88, 90, 92, 94, 00	9482

Parr & Bonitz, 2015	NELS 02, 04	15753

*Notes: "Long." = Longitudinal; * "AA" =African American; **Study reports an approximate sample, no exact sample reported*

Operationalizing High School Graduation Status

High school graduation status is always operationalized as a categorical variable in the literature; however, there are many nuances to the treatment of this outcome variable across studies. Table 2 displays a range of ways that this outcome variable has been operationalized across studies

Table 2

Methods of Operationalizing High School Graduation Status

Author(s), Year	Description of the Outcome Variable
Astone & McLanahan, 1991	*Two ways*: a) Enrolled continuous through graduation/ All other outcomes; b) Earned diploma or GED by age 22/All other outcomes
Ensminger & Slusarcick, 1992	Graduates/Non-graduates (Still enrolled considered non-graduate)
Rumberger & Larson, 1998	High school diploma/GED/Still enrolled/Dropout
Battin-Pearson et a., 2000	Dropout/All other outcomes (measured at age 16)
Oh & Reynolds, 2008	Completers/Non-completers
Cratty, 2012	Graduates/Non-graduates (Still enrolled and GEDs considered non-graduates)
Parr & Bonitz, 2015	Still attending in 12th grade/Not attending, not graduated

Most frequently, high school graduation status is treated as a binary outcome. All but one of the studies included in Table 2 operationalize the outcome variable this way. Astone and McLanahan (1991) investigated high school graduation status using two different outcome variables and comparing the findings obtained using each. Some studies categorized graduation status as those who earned a high school diploma versus those who did not (i.e. Ensminger & Slusarcick, 1992). Oh and Reynolds (2008) compared completers and non-completers, including students who earned an alternative credential such as a GED in the completers category. Contrary to that, Cratty (2012) classified students who earned GEDs as not having completed high school. Parr and Bonitz (2015), and Battin-Pearson et al. (2000) examined the outcome more as a measure of persistence rather than completion; both studies explored high school graduation status in terms of whether a student had dropped out of school by a certain age. In January 2017, the U.S. Department of Education (2017) provided guidance on measuring high school graduation status for the Every Student Succeeds Act. Their non-regulatory guidance suggests measuring this as an Adjusted Cohort Graduation Rate, similar to Ensminger and Slusarcick's (1992) and Cratty's (2012) operationalization.

Predictors of High School Graduation Status

Using the established inclusion and exclusion criteria for this review of the literature, two categories of high school graduation status predictors emerged – academic predictors and nonacademic predictors. Academic predictors are observed variables that pertain to activities student achievement outcomes. Nonacademic predictors include measures of student engagement, as well as student demographic predictors.

Academic predictors. A high school diploma is the summative product of having achieved a certain level of academic success. It should come as no surprise that academic

achievement is the most reliable predictor of whether or not a student will finish high school. One of the primary reasons cited for students failing to complete high school is unsatisfactory academic progress (Astone & McLanahan, 1991; Ensminger & Slusarcick, 1992; Jordan et al., 1996; Mac Iver & Messel, 2012; Rumberger, 2011). Previous research on the relationship between academic achievement and graduation has examined how academic variables at different points in a student's K-12 career are associated with graduation status (e.g. Lloyd, 1978). Lan and Lanthier (2003) conducted survey research with a sample of over 1100 high school dropouts and found that academic performance significantly declined from eighth grade to tenth grade, and again from tenth grade to twelfth grade. Deteriorating academic progress was a hallmark trait for this population of students. A qualitative study examining the perspectives of students who dropped out of high school found that almost one in two dropouts cited poor preparation in elementary and middle school years as a reason that they did not complete school (Bridgeland, Dilulio, & Morison, 2006). A similar qualitative study conducted in Philadelphia echoes this finding; study participants who dropped out of high school described academic struggles as being a central reason for failing to complete school (Bayliss et al., 2011). Academic achievement has been operationalized in the literature in five primary ways: (1) grades; (2) courses failed; (3) course enrollment; (4) test scores; and (5) retention.

Grades. Graduating from high school is the result of passing a requisite sequence of courses, and it logically follows that a student's grades would be an indication of whether or not he or she will reach this apogee. However, several studies have been able to predict a student's graduation status using grades earned early in their academic career. Among the studies included in this review, three studies used elementary school grades as predictors of high school graduation and in every study grades were found to be significant predictors (Alexander,

Entwisle, & Horsey, 1997; Ensminger & Slusarcick, 1992; Lloyd, 1978). What differed between these studies was how "student grades" was operationalized. Two studies examined specific grades earned in courses (i.e. third grade reading), one examined GPA in the third grade, and the third study dichotomized grades earned in the first grade as As and Bs versus Cs and Ds. Ensminger and Slusarcick found that male students who earned As and Bs in the first grade were more than two times more likely to graduate from high school than were students who earned lower grades. Female students were more than 1.5 times more likely to graduate. Overall, the literature suggests that elementary school grades can be useful predictors of a student's graduation status.

Four studies explored the relationship between grades earned in middle school and a student's eventual graduation status, all of which found grades to be statistically significant predictors (Barrington & Hendricks, 1989; Rumberger & Palardy, 2005; Rumberger & Larson, 1998; Tobin & Sugai, 1999). Two of the studies used grade point average as a predictor, one used record data, and a fourth study relied on an average of self-report grades in middle school core subjects. Although Rumberger and Larson's reliance on self-report data limits the inferences that can be drawn from that data, it does provide another piece of evidence that can be considered with the rest of the studies included in this review. Similar to the findings on elementary school grades, middle school grades have utility as predictors of high school graduation status. However, administrative record data is the preferred method of obtaining these. Self-report survey data introduces several possible sources of error that are much less of a concern with record data (Mitchell & Jolley, 2013).

Another four studies explored the relationship between grades earned in high school and a student's graduation status. Barrington and Hendricks (1989) found that failing grades in any

year between seventh grade and ninth grade were a predictor of dropout. Astone and McLanahan's (1991) study found that students maintaining at least a B-/C+ grade point average were likely to graduate from high school. Soland (2013) found a student's grade point average in the ninth grade to be a significant predictor of high school dropout, however when tenth grade English and mathematics grade point average were added as predictors, no relationship was found. This finding runs counter to Jimerson and Ferguson's (2007) longitudinal study which found grade point average in the tenth grade to be a significant predictor.

Overall, grades earned during each of level of K-12 schooling - elementary, middle, and high school - were found to have an association with high school graduation status. Regardless of how student grades were operationalized, they were found to be significant predictors of whether or not a student would complete high school. Better grades were associated with an increased likelihood of completing high school and poor grades at every level were associated with an increased likelihood of dropping out of school.

Courses failed. Several studies have operationalized student achievement in terms of failing grades earned in middle and high school courses (Allensworth & Easton, 2005; Balfanz & Herzog, 2005; Neild & Balfanz, 2006; Rumberger, 2011; Rumberger, 1987). Six studies in this review of the literature explored the relationship between failing courses during the secondary years and completing high school. A study based in Los Angeles explored the relationship between ninth grade student course failure and graduation status, and found that each course a student failed was associated with a 10% reduction in their likelihood to graduate from high school (Silver, Saunders, & Zarate, 2008). A similar study in Los Angeles found that failing a core subject course in middle school had an even stronger association with dropping out of school (Saunders, Silver, & Zarate, 2008). Each course failed by middle school students was

associated with an additional 20% increase in the likelihood of dropping out. Balfanz and Herzog (2005) found that fewer than one in ten students who failed sixth grade English or math went on to graduate on time. Similarly, Neild and Balfanz (2006) found that fewer than one in four students who failed eighth grade English or math graduated on time. Allensworth and Easton (2005) measured course failure in terms of the number of credits a student earns during his or her freshman year. Students who earned enough credits to be promoted to the tenth grade and failed no more than one semester of a core subject course were considered on-track for graduation; those who failed to meet these criteria were considered to be off-track. Eighty-one percent of students who were on-track by the end of the ninth grade year went on to graduate from high school, a rate four times greater than that of students who were off-track after their freshman year (22% graduation rate).

Early warning indicators have been described as signs that a student is at risk of not graduating from high school (Allensworth & Easton, 2005; Mac Iver & Messel, 2012; National Research Council, 2011). In Mac Iver and Messel's (2013) study exploring the predictive power of early warning indicators, failing an eighth grade English or mathematics course was found to be a significant predictor of dropout. Students who failed one of these two courses in the eighth grade were almost twice as likely to drop out of school as a student who passed both courses. In the same study, Mac Iver and Messel (2013) found that high school students who failed a core subject course in the ninth grade were 2.5 to five times more likely to drop out as students who did not. Neild et al. (2008) also examined the impact that eighth grade course failure had on a student's eventual graduation outcome. They operationalized course failure as a percent of courses failed, and found that the percent of courses failed in the eighth and ninth grade were predictors of high school dropout. Neild et al., and Owens, Morris, and Lieberman (2001)

operationalized failure in middle school courses as earning the grade of D or F. They justified how they operationalized this by citing anecdotal evidence that middle school teachers were less inclined to give students failing marks even when they were earned than were high school teachers.

Overall, course failure was consistently found to be associated with an increased likelihood of high school dropout regardless of whether it was operationalized as a percentage of courses failed, as the number of courses failed, or as a dichotomous variable of having failed a course or not. The authors of two of the studies included earning a D in a middle school course as a failure. Additional studies should consider operationalizing middle school course failure this way to develop a stronger literature around this practice.

Course enrollment. Specific course enrollment has been found to predict whether or not a student will graduate from high school. Two studies found enrollment in Algebra I to be a significant predictor of graduation status. Soland (2013) found enrollment in Algebra I, as well as enrollment in Geometry, by the tenth grade to be predictors of graduation status. Cratty (2012) found enrollment in Algebra I by the eighth grade to be a strong predictor of graduating from high school. The statewide graduation cohort examined in Cratty's (2012) study had an overall dropout rate of 19.3%; less than 5% of those enrolled in Algebra I by the eighth grade dropped out of school. Silver, Saunders, and Zarate (2008) found that students who passed Algebra I by the ninth grade were twice as likely to graduate as students who failed to do so.

Interestingly, studies found that both enrollment in any Advanced Placement (AP) course and enrollment in a remedial English course to both be associated with a decreased likelihood of dropout (Goldschmidt & Wang, 1999; Soland, 2013). The first finding is supported by the

literature on tracking. Students placed on more academically challenging tracks, which tend to include more advanced curricula including AP courses, drop out at substantially lower rates than students placed on less challenging academic tracks (Oakes & Guiton, 1995; Rumberger, 2011; Werblow, Urick, & Duesbury, 2013). The second finding is certainly counterintuitive. One explanation for this could be that the remediation did what it was designed to do; students who needed additional supports to succeed in high school received those supports and went on to graduate from high school (Bridgeland et al., 2006).

Overall, enrollment in specific courses and specific types of courses was found to have an association with high school graduation status. Enrollment in specific courses in math, Algebra I and Geometry, were found to be statistically significant predictors of high school graduation status. Enrollment in both AP courses and remedial English were found to be associated with an increased likelihood of graduation. Only one study included in this review explored the predictive ability of enrollment in remedial courses. This variable should be considered for inclusion in future models predicting high school graduation.

Test scores. The second most common indicator of academic achievement used to predict high school graduation status is standardized test scores. In many states, including Virginia and New York, students must pass a series of state-mandated standardized tests in addition to meeting course requirements in order to be eligible to graduate (Rumberger, 2011; Virginia Department of Education, 2016). Studies aimed at using test scores to predict high school graduation status have yielded mixed findings. In the studies included in this review, three used state-mandated standardized tests as predictors, four used nationally-normed achievement tests (Iowa Test of Basic Skills [ITBS] and Peabody Individual Achievement Test [PIAT]), and two studies used composite scores for reading and mathematics standardized tests.

Four of the studies examined the predictive ability of standardized test scores from elementary school grades. Barrington and Hendricks (1989) and Hernandez (2011) examined the relationship between third grade test scores and graduation, and both studies found a significant relationship to exist between the two variables. Hernandez (2011) used PIAT reading test scores as a predictor and found that one in six children who are not reading at a proficient level in the third grade do not graduate from high school in four years, however significant limitations exist with this study. High school completion was noted by whether or not the student reported that he or she had completed; whether or not a student had actually dropped out was not confirmed. Also, third grade scores were only used for approximately half of the sample. The other half of the sample's "third grade reading score" was an average of their second and fourth grade reading scores. Jimerson et al. (2000), and Alexander, Entwisle, and Henry (1997) explored the relationship between first grade test scores and eventual graduation status and both studies found no significant relationship to exist.

Five studies included in this review examined the relationship between middle school test scores and graduation status, and their findings were mixed. Rumberger and Larson (1998) and Barrington and Hendricks (1989) found eighth grade reading and mathematics scores to be significant predictors of whether or not a student dropped out of high school. Jimerson et al. (2000) found sixth grade PIAT scores to predict graduation status, and found the variable to correctly classify cases 77% of the time. The findings from Neild et al.'s (2008) study examining high school graduation in Philadelphia were more nuanced. The model that only included eighth grade predictors found a significant relationship to exist between mathematics scores on statewide standardized tests and graduation status, but not for reading and graduation status. When ninth grade academic variables were introduced, neither of the eighth grade

variables was a significant predictor. This was likely due to multicollinearity that existed between the eighth and ninth grade variables. To avoid the violation of this assumption, the variables should have been tested in separate models.

Similarly, Mac Iver and Messel (2013) included a dichotomous predictor variable with their ninth grade variables indicating whether a student had a non-proficient score on their eighth grade state standardized reading test. The sample in this study was comprised of two graduation cohorts, tested separately. Non-significant eighth grade test scores yielded significant *p*-values in both models, however only one of the two cohort models found this predictor to have practical significance. In the first cohort, non-significant test scores produced an odds ratio of 1.00. This indicates that students who did not pass an eighth grade reading test were just as likely to graduate as they were not to graduate. In the second cohort's model, non-significant test scores produced an odds ratio of approximately 1.8. This finding does have practical significance; students who did not pass an eighth grade reading test were almost twice as likely to fail to graduate from high school.

Three studies examined the relationship between high school test scores and graduation status, with mixed findings. Ou and Reynolds (2008) and Jimerson et al. (2000) found no relationship to exist between test scores and graduation status, whereas Parr and Bonitz (2015) did find a relationship to exist. One possible explanation for this discrepancy is that the Parr and Bonitz study focused primarily on two latent constructs and included far fewer covariates in their models than other authors did in their studies, and in these studies it is possible that some of the additional covariates were able to account for some of the variance that was explained by test scores in Parr and Bonitz's study.

Nine studies included in this review that explored the relationship between standardized test scores and high school graduation status. Of these, three studies included the raw test scores in the models (Neild et al., 2000; Ou & Reynolds, 2008; Parr & Bonitz, 2015); four studies included an overall composite score that included math and reading (Alexander et al., 1997; Barrington & Hendricks, 1989; Jimerson et al., 2000; Rumberger & Larson, 1998); and two studies treated test scores as a categorical variable placing the emphasis on whether or not a student earned a proficient score (Hernandez, 2011; Mac Iver & Messel, 2013). See Table 3 for a breakdown of the type of test score variable used and the associated findings.

Table 3.

Type of Test Score Variable Used and Associated Findings

Type of variable	Type of Findings		
	Significant	Non-Significant	Mixed
Raw score	0	1	2
Composite score	2	1	1
Categorical	2	0	0

How this variable is treated has implications on the inferences that can be made as a result of the findings. The cleanest interpretations can be made with dichotomous categorical predictor variables. Each of the studies included in this review of the literature featured dichotomous variables for standardized reading tests. For studies that include only composite test scores, inferences can only be made about how a student's overall performance on standardized tests is related to his or her graduation status. Raw scores for specific tests allow

for the most specific inferences to be made based on findings. None of the three studies included in this review that utilized raw test scores produced a significant finding. However, if one of these studies found a standardized test score to be a significant predictor, it might be possible to state that for every additional point earned, a student's odds of graduating complete by *x* amount.

Overall, the literature reviewed for this study contains mixed findings on whether an association exists between standardized test scores and high school graduation status. At every level – elementary, middle, and high school – some studies found statistically significant relationships to exist between test scores and graduation, and others did not. Standardized test scores should be considered for inclusion in future studies investigating the ability for variables to predict graduation status. Enough evidence exists which indicates that test scores could be a factor in whether or not a student completes high school. Even if test scores are found to be insignificant predictors in future work, their inclusion in the model could alter the findings regarding the predictive ability of other academic variables.

Retention. The premise underlying student retention is straight forward. Students should master the material at their present grade level before being promoted to the next level (Jimerson, Anderson, & Whipple, 2002). Students who enter the next grade lacking the prerequisite knowledge to succeed will be more likely to struggle academically (Feden & Vogel, 2003; Slavin, 2006; Willingham, 2009). Therefore, the rationale for retention is that it allows students who did not initially master the requisite material the opportunity to do so before facing more academically challenging content. On its face, this rationale seems reasonable. However, an abundance of literature suggests that students who are retained are less likely to complete high school.

A study conducted by Stroup and Robins (1972) found retention to be the strongest predictor of a student dropping out of high school. Neild, Stoner-Eby, and Furstenberg (2008) found that being retained at any grade level was the second strongest predictor of student graduation status behind being enrolled in remedial coursework. Other studies that have examined retention as a dichotomous variable (ever been retained versus never been retained) have also found it to be a significant predictor of dropout (Alexander, Entwisle, & Horsey, 1997; Cratty, 2012; Mac Iver & Messel, 2013). Studies included in Jimerson, Anderson, and Whipple's (2002) systematic review of the literature consistently found that students who were retained a grade were more likely to drop out of high school than students who were promoted annually. Rumberger and Larson's (1998) longitudinal study found that students who were retained at least once prior to high school were four times more likely to drop out of school than those who were not. Andrews (2014) discusses grade retention as a "triggering event" (p. 658) that leads students to develop a "low-value status" (p. 658) as a learner. Viewed as a triggering event, retention potentially plays a major role in the process of students dropping out of high school.

Although retention generally is related to a student's high school graduation status, *when* a student is retained seems to make a difference. First grade is the most common point in K-12 education when students are retained (Warren, 2014) and retention during elementary school grades has been consistently found to be a significant predictor of student dropout status (Jimerson, Anderson, & Whipple, 2002; Rumberger, 2011; Rumberger, 1987). Lloyd (1978) found retention from grades 1-3 to be a significant predictor of students failing to complete school. Jimerson (1999) also examined the impact of early grade retention and found that students being retained between kindergarten and third grade were 25% more likely to drop out

out than a comparison group of low-achieving students who were continuously promoted. Taken together, these findings would suggest that being retained during the traditional elementary school grades (K-5) has a more detrimental effect on long-term academic outcomes than being promoted without all of the requisite knowledge and skills for the next grade, which runs counter to the theoretical underpinnings that promote the use of retention.

Ambiguity exists in the literature examining the relationship between retention during middle school years (6-8) and high school graduation status. Cratty (2012) conducted a longitudinal study of a third-grade statewide cohort in North Carolina and one of the findings was that retention between the sixth and eighth grade was associated with higher rates of dropout. An astounding 69% of students in the study who were retained during the middle school grades dropped out of high school. Ou and Reynolds (2008) conducted a study that yielded similar findings; students who had been retained between the ages of 10 and 14 were significantly more likely to fail to complete high school. However, Jacob and Lefgren (2004) found the opposite to be true; they found no relationship to exist between retention in the middle school years and high school graduation status.

Retention at the beginning of a student's high school career is also related to graduation status. Behind first grade, ninth grade is the year that students are second-most likely to be held back (Warren, 2014). At the turn of the 21st century, students were increasingly more likely to be retained during the ninth grade year than they were in previous decades (Abrams & Haney, 2006). In Cratty's (2012) longitudinal study, more than 70% of students who were retained during the ninth grade year went on to drop out of high school. In general, students who are not promoted to the tenth grade are substantially less likely to complete school than those who are.

Although the vast majority of the literature examining the relationship between retention and high school graduation status has found this to be a significant predictor, two studies included in this review did not. McWilliams (2000) tested a pair of models that were designed to predict high school dropout, using separate models for male and female students and retention was not found to be a predictor of high school graduation status. A possible confound lies in the participants included in the study. Because McWilliams used a twelfth grade cohort in a single rural school district, one possible explanation is that students who made it to their senior year despite being retained are outliers. The students who were retained that failed to graduate from high school in this school district never made it to the twelfth grade.

Another study that failed to find a significant relationship between retention and graduation status investigated early warning indicators and teacher prediction of student graduation status as predictor variables (Soland, 2013). Mac Iver and Messel (2013) used early warning indicator variables in their predictive model and students being overage for grade as a significant predictor of graduation status. Students are typically designated overage for grade if they are older than 14 years old entering the ninth grade, and this is most often the result of being retained at some point in an earlier academic year (Rumberger, 2011). The discrepancy between Soland's and Mac Iver and Messel's findings could be explained two ways. First, Mac Iver and Messel relied on data obtained from a single school district; Soland used data from a nationally-representative longitudinal data set. It is possible that differences between the samples could partially explain the findings. However, this would still run counter to the overall trend of retention as a predictor in the literature. A better explanation could be the interaction between the variables included in the study. In Soland's study, teacher intuition was included as a variable. Teachers predicted whether or not a student would graduate from high school, and this

was found to be one of the strongest predictor variables included in the study when controlling for early warning indicator and test score variables. It is possible that teachers consider whether or not a student had previously been retained as a factor in their prediction process and this is already accounted for in that variable.

Overall, the literature reviewed on the relationship between retention and high school graduation status indicates that a strong association exists between retention and graduation. Students who repeat a grade are less likely to graduate from high school and more likely to drop out. In the case of retention, an intervention designed to improve long-term student outcomes is not effective in doing so.

Nonacademic predictors. Factors unrelated to student achievement have also been found to be significant predictors of high school graduation status. In this review of the literature, two types of nonacademic predictors emerged – those related to student engagement and student demographic predictors.

Predictors related to student engagement. Nonacademic predictors related to student engagement included in this review of the literature included: (1) student attendance; (2) student behavior; and (3) extracurricular participation.

Student attendance. Poor student attendance has been consistently found to have a positive relationship with high school graduation in the literature (Balfanz, Herzog, & Mac Iver, 2007; Rumberger, 2011; Rumberger, 1987). In order to participate in classroom activities and grow academically, students have to be in attendance. Rumberger (2011) discusses attendance as “one of the most direct and visible indicators of engagement” (p. 169). While this might be true, the inverse is not so simple. The factors that lead to students missing school are complex,

and are the result of academic progress as well as outside-of-school factors (Rumberger, 2006). A study in Philadelphia found that more than half of those responding to a survey reported that outside-of-school issues contributed to not attending school (Bayliss et al., 2011). Bridgeland, Dililio, and Morison (2006) conducted a series of focus groups and interviews with students who dropped out of school. Students described academic struggles as a reason for their decision to drop out, indicating that they had crossed a point of no return and that there was no way for them to catch up and graduate. However, students also indicated that life circumstances led to their failure to complete school, for reasons similar to those provided by the students in Bayliss et al.'s Philadelphia study. Patterns of poor attendance often manifest well in advance of a student dropping out of school; poor attendance is often identified as an early warning indicator of high school dropout (Balfanz, Herzog, & Mac Iver, 2007; Mac Iver & Messel, 2013; Mac Iver & Messel, 2012).

Studies that have examined the relationship between student attendance and high school completion have operationalized attendance in three primary ways: (1) as a percent (Neild, Stoner-Eby, & Furstenburg, 2008); (2) as the number of days attended over a given time frame, typically an academic year (Alexander, Entwisle, & Horsey, 1997; Ou and Reynolds, 2008); and (3) as a categorical variable (; Cratty, 2012; Parr & Bonitz, 2015; Rumberger & Larson, 1998; Soland, 2013). When attendance is expressed as a percent or a decimal between zero and one, the numerator is the number of days attended and the denominator is the number of days enrolled. Attendance is operationalized categorically in three ways in the studies included in this review. First, attendance can be operationalized dichotomously, with a pre-determined cut point determining that a student has excessive absences or not. Mac Iver and Messel (2013) set their cut point at 20 days, representing approximately one month of instruction missed. Students who

missed more than 20 days were categorized as "chronically absent" (Mac Iver & Messel, 2013, p. 53). Those who missed 20 days or fewer were categorized as not being chronically absent. Cratty (2012) operationalized attendance cumulatively in four categories: (1) 0-7 days missed; (2) 8-14 days missed; (3) 15-21 days missed; (4) more than 21 days missed. Rumberger and Larson (1998) and Soland (2013) both used the National Educational Longitudinal Survey 1988 dataset, which operationalizes attendance by soliciting self-report data on how many days participants missed school over the past four weeks: (1) zero days; (2) one or two days; (3) three or four days; (4) five to ten days; (5) more than ten days (U.S. Department of Education, 1988). This is an imperfect method of measuring student attendance for two reasons. First, the reliance on self-report data could yield responses that fall prey to social desirability bias or faulty recollection, either of which introduces error into the data (McMillan, 2012; Mitchell & Jolley, 2013; Thorndike & Thorndike-Christ, 2010). Also, by measuring attendance with only four weeks (approximately one-tenth of a typical school year), it is possible that the data obtained from this item would not be representative of a full academic year, which also lends the possibility of making faulty inferences from the findings.

Using data from the National Education Longitudinal Study (NELS) of 2002, Parr and Bonitz (2015) operationalized student attendance by asking teachers how often students were absent from school, with possible values ranging from 1 to 4 (1 indicated that students "never" missed school and a 4 indicated that students missed school "all the time"; Parr & Bonitz, 2015, p.510). This treatment of student attendance is unwieldy for the same two reasons as those listed above regarding the 1988 survey. Reliance on self-report data introduces error in the data and measuring attendance by only examining a small fraction of an academic year is problematic. A third cause for concern with this treatment is teacher subjectivity. Whereas the 1988 survey

asked students for an actual number of the days he or she missed, this measure asks teachers for a subjective measure of attendance which is flawed for two reasons. First, it is far more likely that a student can accurately account for his or her experiences than it is that a teacher can accurately account for the experiences of 100 or more students. Second, subjective categories like "all the time" could mean very different things from one teacher to the next without further operationalization. As such, the inferences drawn from the use of these inexact measures of attendance should be heeded with caution.

Overall, attendance was consistently found to be a significant predictor of high school graduation status. Parr & Bonitz (2015) found that high absenteeism, as defined by a student's math or English teacher scoring them as a 3 or 4 on the NELS 2002's nebulous scale, was significantly related to an increased likelihood of dropping out. Mac Iver and Messel (2013), and Neild, Stoner-Eby, and Furstenberg (2008) tested models including eighth and ninth grade attendance, and attendance was found to be a significant predictor in each. In the Mac Iver and Messel models, students missing more than 20 days in the eighth grade were almost three times as likely to drop out as those who missed fewer days, and students who surpassed this threshold in the ninth grade were between three and five times more likely to drop out than those who did not. Rumberger and Larson (1998) also discussed attendance in terms of the relative odds of dropping out. Students who reported missing five or more days over a four-week period in the eighth grade were significantly more likely to drop out. For their model testing including twelfth grade attendance as well, only twelfth grade attendance was found to be a significant predictor with students who report missing five or more days being over 2.6 times more likely to drop out than those who reported missing fewer days. Eighth grade attendance was not a significant predictor in this model, and that is likely because most students who experience significant

attendance issues in the eighth grade are not on track to complete high school never make it to the twelfth grade.

Three studies discuss the relationship between attendance and graduation status in terms of the effect that missing additional days has on the probability of completing high school. Cratty (2012) found that one in three students who miss 15 days or more fail to graduate from high school. Half of students who miss 21 days or more fail to graduate. Ou and Reynolds (2008) operationalized absenteeism by averaging the number of days missed annually in the fifth and sixth grades, and found that each day absent per year represented a 7% decrease in the likelihood to graduate high school. Similarly, Alexander, Entwisle, and Horsey (1997) found that on average, dropouts missed 16 days of school annually, whereas students who graduated from high school missed 10 days of school annually. Each additional day of school missed annually represented a 5% increase in the likelihood to eventually drop out of school.

Soland's (2013) study examining early warning indicators (EWI) and teacher prediction is the lone study that yields mixed findings regarding student attendance. In his study, three models are tested: (1) a baseline model that includes demographics and variables representing three EWI (GPA, course failures, attendance); (2) the full model without teacher prediction; and (3) the full model including teacher prediction. Attendance is found to be a significant predictor for the baseline model, however it is not significant for either of the other two models. Soland's model produced similar findings to the EWI models tested by Mac Iver and Messel (2013). When additional variables were introduced in the second model, attendance no longer served as a useful predictor of dropout. Attendance yielded the smallest beta coefficient for the full model including teacher prediction. Similar to this study's finding with retention, it is possible that

teachers already factored in variables like attendance related to student tenacity in their predictions.

Overall, the literature reviewed indicates that a relationship exists between student attendance and high school graduation. Regardless of how attendance is operationalized, missing fewer days of schools is associated with an increased likelihood of completing school. Of the researchers who used categorical operationalizations of attendance, Mac Iver and Messel's and Cratty's treatments were the most useful. Mac Iver and Messel's dichotomous treatment of the variable is concrete, clean, and simple to understand. Cratty's more complex treatment yields the most information of all of the categorical attendance variables. In the findings, the log odds of dropping out of high school for each of these four variables was presented so that, for example, the log odds of missing between 8-14 days represented the impact that missing those additional seven days had on a student's odds of dropping out above and beyond the impact that missing the first seven days had on his or her outcome.

Student behavior. An association between misbehavior in school and high school graduation status has been found in the literature. In Rumberger's (2011) review of the literature, almost half of the studies that examined misbehavior at the high school level, and more than three-fourths of the studies that examined misbehavior at the middle school level found it to be significantly related to higher dropout and lower graduation rates. Lee, Cornell, Gregory, and Fan (2011) found that African American students were more likely to be suspended from school and found suspension as a predictor of high school dropout. Studies included in this review primarily operationalized misbehavior in one of two ways: (1) in terms of the number of suspensions, or number of days suspended from school (Cratty, 2012; Goldschmidt & Wang, 1999; Neild, Stonery-Eby, 2008; Mac Iver & Messel, 2013); or (2) in terms of in-school

misbehavior (Goldschmidt & Wang, 1999; Jimerson, Egeland, Sroufe, & Carlson, 2000; Rumberger & Larson, 1998; Rumberger & Palardy, 2005).

When misbehavior was operationalized as being suspended from school, it was consistently found to be a predictor of high school dropout (Cratty, 2012; Goldschmidt & Wang, 1999; Mac Iver & Messel, 2013; Neild, Stoner-Eby, and Furstenberg, 2000). Cratty found that 36.6% of students who were suspended once during “one year of middle school” (p. 648) and 59.7% of students who were suspended more than once during the same year went on to drop out of school. Overall, Cratty’s description of how predictor variables were operationalized is very detailed, with a high level of specificity. This was not the case with the description of this variable; it would have been useful to know which year was selected to represent misbehavior in middle school. Mac Iver and Messel found that being suspended three or more days in the eighth grade was a significant predictor of dropping out of school. Students who met this criterion were 1.75 times more likely to drop out of school than those who did not. They also found being suspended three or more days in the ninth grade to be a significant predictor of dropping out of school; students meeting this threshold were 1.3 to 1.9 times as likely to drop out as those who were never suspended or suspended for fewer than two days. Neild et al.’s study examining longitudinal data in Philadelphia found the number of suspensions in the tenth grade to be the largest predictor of dropout. This echoed Goldschmidt and Wang’s finding that students suspended in the tenth grade were almost three times as likely to drop out of school as those who were not.

Three of the four studies included in this review that operationalized student misbehavior in terms of in-school performance examined eighth grade student data. Rumberger and Larson (1998), Goldschmidt and Wang (1999), and Rumberger and Palardy (2005) all relied on data

from the NELS 1988 dataset. Four items asked about student misbehavior and a principal components factor analysis was conducted, yielding a single factor that Rumberger and Larson, as well as Goldschmidt and Wang, described as *misbehaving*. Rumberger and Palardy described this as *classroom disruptions*. In each study, the single factor describing student misbehavior was found to be a significant predictor of dropout status. Jimerson, Egeland, Sroufe, and Carlson (2000) used the Child Behavior Checklist-Teacher (CBCL-T) instrument (Edelbrock & Achenbach, 1985) to measure "problem behaviors" (p. 532). Data on problem behaviors from first grade, sixth grade, and age sixteen were included in the models they tested. In a model that also included third grade IQ and variables related to early childhood, misbehavior in the first grade and at age 16 were strong, statistically significant predictors of dropout, with each classifying between 76% and 77% of cases correctly. Although sixth grade behavior was not significant in that model, it was the strongest predictor in a model that also included first grade behavior. One possible explanation for the discrepancy is that a large amount of overlap exists between misbehavior in the sixth grade (typically age 12) and at age 16. Students who were apt to misbehave at one age, likely misbehaved at the other age as well.

Overall, the literature reviewed found a relationship between students' behavior in school and their propensity to graduate from high school. Students who misbehave in school are less likely to graduate from school than those who do not. None of the studies included in this review yielded non-significant findings for student misbehavior, however operationalized, as a predictor of high school graduation status.

Extracurricular participation. The relationship between participating in extracurricular activities and eventual high school graduation status is another facet of student engagement that has been explored in the literature (Rumberger, 2011). In Rumberger's (2011) review of the

literature, he found that "participation in extracurricular activities reduced the odds of dropping out" (p. 170) in twelve of nineteen studies that examined this at the high school level. Most of the studies included in this review yielded a similar finding. Song, Benin, and Glick (2012) found a significant relationship to exist between participating in extracurricular school activities and a decreased likelihood of dropping out of school. McNeal (1995) found that participation in sports clubs and clubs related to the arts to be significant predictors of high school completion, however he found there to be no relationship between participating in academic or vocational-related clubs. The model also included variables for tracking; students on an academic track were less likely to drop out of school. No relationship was found between being on a vocational track and dropping out of school. The inclusion of these two variables likely decreased the impact participating in extracurricular clubs related to either track. Rumberger and Larson's (1998) study examining the relationship between student mobility and high school graduation status included participation in extracurricular activities as a variable in their models and found no relationship to exist.

McWilliams, Everett, and Bass (2000) grouped predictor variables in clusters for their study. The cluster that included extracurricular activity participation was labeled *Social Snythesis* [sic]*/Integration* (McWilliams et al., 2000, p. 45), and also included variables for student popularity, being in trouble with law enforcement, and school attended. All four variables were retained for the analysis done for male students, and all but trouble with law enforcement were retained for females. The cluster was found significant in both analyses, classifying more than 80% of cases correctly. There is no indication of a factor analysis being conducted to provide justification for the grouping of these four variables, and this unusual,

otherwise unexplained combination makes it difficult to draw concrete inferences from the finding.

Overall, the literature on the association between participation in extracurricular activities and high school graduation yields mixed findings. Most of the studies included in this review find that participating in extracurricular activities, especially those involving sports and the arts, is related to having an increased likelihood of completing high school. The non-significant finding for participation in academic extracurricular activities makes sense given that academic track was also included in the model. Extracurricular activity participation is an example of a measure of student engagement that can be captured as an observed variable, and this should be considered for inclusion in future predictive models.

Demographic predictors. Six types of demographic predictor variables were explored in this review of the literature. These include: (1) race; (2) gender; (3) socioeconomic status; (4) family structure; (5) disability status; and (6) student mobility.

Race. One of the quintessential challenges in public education has been to work to decrease academic achievement gaps that exist between demographic groups, especially pertaining to race and gender (Hauser, Simmons, & Pager, 2006; Losen, 2006; Orfield, 2006; Rumberger, 2011). Vast differences often exist along racial and gender lines in terms of student graduation outcomes (Rumberger, 2011; Rumberger 1983). Race and gender differences are often variables included in models designed to predict graduation rates. Among the predictive models included in this review of the literature, seven studies explored racial differences in graduation rates. The most common racial difference explored was that which existed between White and African American students. Four of the seven studies found that African American

students were more likely to drop out of school than were White students. Two studies produced more nuanced findings around race. In their study exploring the relationship between high school graduation and family structure and transitions, Song, Benin, and Glick (2012) found that when family-related variables including socioeconomic status were removed from the model, no significant difference existed between White and African American students in terms of high school graduation. However, when the family variables were introduced in the model, a significant difference existed between the two racial groups. Goldschmidt and Wang (1999) explored the difference between early dropouts (between eighth and tenth grade) and late dropouts (between tenth and twelfth grade) along a set of predictor variables including race. African Americans were more likely to drop out of school early than were White students. However, no significant difference existed between the groups for late dropouts.

For studies that explored differences that existed between more than two racial groups, findings were almost always presented in terms of differences with White students. For models tested using logistic regression analysis, White students were selected as the reference group in every study except for one. Only one out of five of the studies that explored graduation outcome disparities between White and Latino students found a significant difference to exist. Five studies that explored the graduation outcomes between White and Asian students, of which two found Asian students to outperform all other races, and another two found no significant differences to exist. Interestingly, the study with the largest sample size found when a number of in-school variables were accounted for, White students dropped out at higher rates than African American, Latino, and Asian students (Cratty, 2012). The only group White students outperformed in this study was American Indians, who dropped out at substantially higher rates than students identified as members of the other racial groups.

Gender. Ten studies included in this review examined the relationship between gender and high school graduation status. Of the ten studies, five found that males were either less likely to graduate from school or more likely to drop out, depending on how the outcome variable was operationalized. Two additional studies found no significant difference between the two genders. Rumberger and Larson's (1998) study examining the impact of student mobility on high school graduation produced a more nuanced finding. When examining students' eighth grade variables, female students were actually found to be between 1.2 and 1.6 times more likely to not complete school than were males, findings that held true two years later when school characteristics and student engagement were included in the model. When the same cohort of students was examined as twelfth grade students, there were no statistically significant differences between the genders in terms of six-year graduation rates.

Socioeconomic status. Ever since the Coleman Report (Coleman, 1966), much has been written about the relationship that exists between poverty and student academic outcomes. Twelve of the studies in this review included some variable(s) reflecting the socioeconomic status of a student's family and their graduation status (e.g. Song, Benin, & Glick, 2012) and this was operationalized five ways. Five studies included a composite variable representing socioeconomic status, which typically included measures of family income and parent education, and sometimes included parent occupation as well. In each of these studies, low socioeconomic status was associated with increased likelihood of dropout and decreased graduation (Battin-Pearson et al., 2000; Jimerson et al., 2000; Rumberger & Palardy, 2005; Rumberger & Larson, 1998), and low socioeconomic status was associated with students having upwards to twice the likelihood of dropping out of school as a student of a higher socioeconomic status (Goldschmidt & Wang, 1999). Parent income and parent education were also found to be significantly related

to higher levels of dropout in each study examined. Parent income was operationalized as eligibility for free or reduced lunch, family receipt of welfare payments, income earned, and having lived below the federal poverty line (Alexander, Entwisle, & Horsey, 1997; Hernandez, 2011; Song, Benin, & Glick, 2012). Using a nationally-representative sample, Hernandez (2011) found that 22% of students who had lived in poverty went on to drop out of school, whereas only six percent of students who had never experienced poverty dropped out. Ou and Reynolds (2008) explored family income as a Level 2 predictor and found no relationship between the percent of low-income families living near a student and their graduation outcome.

Family structure. Another variable that has been explored is the difference that exists in the graduation status of students who come from different types of families. Most literature examines this in terms of the association that a family's parent/guardian makeup has on high school graduation, and in terms of the association that divorce has on a student's graduation outcomes. Eleven studies included in this review explored the relationship that exists between family structure and student graduation outcomes. Nine out of ten studies found that students who come from single parent families had a greater propensity to drop out of school than students who came from two parent families. Lagana's (2004) study exploring family cohesion found that students who had two parents who stayed married were less likely to be held back a grade, which has been associated with higher levels of dropout (e.g. Andrews, 2014). Cavanagh, Schiller, and Riegle-Crumb (2006) found that students from two parent families were more likely to have passed Algebra I and Algebra II, and more apt to graduate from high school – a finding consistent with previous empirical work demonstrating the relationship between Algebra and graduation (Cratty, 2012; Silver, Saunders, & Zarate, 2008). Song, Benin, and Glick (2012) found that while students from two parent households had the lowest odds of dropping out.

Among all other family structures, students living with just a single mother had the lowest odds of dropping out.

Song et al. (2012) also explored the impact that divorce has on student graduation outcomes and found that students whose parents divorced within the past year were twice as likely to dropout as students for whom that was not their situation. Painter and Levine's (2000) study examining the impact of family structure on student outcomes also found divorce to be a significant predictor of student dropout after controlling for eighth grade and demographic variables. Parents remarrying had no impact on student graduation outcomes (Painter & Levine, 2000; Song et al., 2012).

Disability status. Some scholars would describe a student's eligibility for special education services as an academic variable, and not an individual demographic predictor. Similar to mobility, whether or not a student has a learning disability is almost certainly as much of a choice as is his or her race or gender. Surprisingly few of the studies included in this literature review explored special education status as a predictor of high school graduation status. One theory for this is that other academic outcomes, such as test scores and grades, explain most of the differences in graduation outcomes that exist between students with disabilities and those without (Powell & Steelman, 1993). Ou and Reynolds (2008) found no relationship between receiving special education services and graduation status; test scores and grades were also included in their model and found to be significant predictors. Two other studies found that students with a specific learning disability were more likely to drop out of school than students without disabilities, while students diagnosed with emotional and behavioral disorder had dropout rates of 50% or more (Cratty; 2012; Reschly & Christenson, 2006). Although it is possible that the differences in graduation and dropout between students with and without

disabilities disappear when other academic predictors are considered, this is an area in need of further exploration.

Student mobility. In Rumberger's (2011) comprehensive book on the many facets of high school dropout, student mobility – or students changing schools unrelated to promotion from elementary to middle school, or middle to high school – is considered an educational factor. It is not in this review of the literature; whether or not a student moves likely is not his or her decision and is more similar to a demographic variable in that regard. Five studies included in this review included a variable related to student mobility as a predictor in their model(s), all with findings indicating a significant relationship between a student changing schools and an increased likelihood of dropout and decreased likelihood of graduation. The most striking was found in Rumberger and Larson's (1998) study; students who changed schools between eighth and twelfth grade were between four and seven times more likely to drop out than were students who only moved for promotion to high school.

Summary of nonacademic predictors. The studies included in this review of the literature explored two primary categories of nonacademic predictor variables and their relationship with graduation. For predictors related to student engagement, student attendance and student behavior were found to be significant predictors of high school graduation status across studies, no matter how they were operationalized. Findings surrounding participation in extracurricular activities were more nuanced. Participation in sports or art-based activities were more likely to be significant predictors of graduation status, whereas participation in academic-related extracurricular activities was not a consistent predictor of graduation status.

For demographic predictor variables, studies that included race and gender as predictors tended to find that students who were White and female were more likely to graduate from high school than were their counterparts. However, findings from some studies including demographic variables yielded findings than ran counter to the dominant findings. More than any other out-of-school predictor variables, the relative odds of graduating high school associated with race and gender depended on the nature of the other variables included in the models. Many of the studies that explored the predictive ability of race examined the differences between White and African American students. Although useful and interesting findings can emerge when race is treated in this manner, studies exploring the predictive ability of race should consider Latino students as well at a minimum. According to the 2015 Census statistics, Latinos comprise almost 14% of the population of the United States and are now the second largest racial group in the country (U.S. Census Bureau, 2015). Conducting research that excludes this group from consideration ignores the fastest growing group in the United States.

Overall, the studies included in this review found that family structure, socioeconomic status, and mobility were all statistically significant predictors of high school dropout. Students from two-parent households were consistently found to perform well academically and were increasingly likely to graduation compared to students from other family structures. Students who came from low socioeconomic status backgrounds were found to be more likely to drop out of high school, regardless of how socioeconomic status was operationalized. Also, students who switched schools for reasons other than grade promotion were found to be more likely to drop out than peers who did not. Each of these variables should be considered for inclusion in future predictive studies.

School-level predictors. Seven of the studies included in this review of the literature explored school-level variables using multilevel modeling (e.g. Goldschmidt & Wang, 1999), and six primary findings emerged. Low socioeconomic status, frequently operationalized by student eligibility for free and reduced lunch, was found to be a significant predictor of an increased likelihood to drop out of high school (Christle, Jolivette, & Nelson, 2007; Goldschmidt & Wang, 1999; Rumberger & Palardy, 2005; Werblow & Duesbergy, 2009; Zvoch, 2006). School-wide student retention, attendance, suspension, and dropout rates were also associated with an increased likelihood of student dropout (Christle, Jolivett, & Nelson, 2007; Goldschmidt & Wang, 1999; Zvoch, 2006). The one consistent nonsignificant finding occurred when a student's English language learner status was tested as a predictor of dropout; no relationship was found (Werblow & Duesburg, 2009; Zvoch, 2006). Future predictive models employing multilevel modeling with schools at the second level should consider including retention, attendance, suspension, dropout rates, and free and reduced lunch eligibility as predictors.

Neighborhood effects. Schools are not the only setting in which students exist; in fact, students spend more of their waking hours outside of school than they do in school. Foster and McLanahan (1996) define *neighborhood effects* as "the influence of neighbors and neighborhood institutions on individual choice" (p. 251). In other words, where a student resides can have an impact on the decisions that he or she makes, including those related to schooling. Their study examining the impact that neighborhood effects have on high school graduation status found that the types of occupations that existed, poverty rate, and community dropout rate to be significant predictors of high school graduation status (Foster & McLanahan, 1996). Students who grew up in neighborhoods where poverty was more common and more individuals dropped out of school were more likely to fail to complete school themselves. Crowder and Smith (2003) estimated

neighborhood socioeconomic status using a composite measure approximating neighborhood disadvantage. Neighborhood poverty rate, the percentage of families receiving public assistance, male unemployment rate, the percentage of workers in non-professional occupations, and the percentage of residents without a college education were all factors comprising this measure. As neighborhood disadvantage increased for students, their odds of graduating from high school decreased. Crowder and Smith (2003) also included a race by neighborhood disadvantage interaction term in their model and found that the impact of neighborhood disadvantage was twice as large for African American students; African American students from disadvantaged neighborhoods were twice as likely to drop out of high school as were their White peers from similar neighborhoods. Another study found neighborhood effects, including neighborhood poverty rates, to be indirectly related to early math achievement, teenage drug use, and the decision to drop out of high school (Ensminger, Lamkin, & Jacobson, 1996).

Taken together, the setting in which students reside can impact their academic performance, and models that strive to predict graduation status should attempt to account for this. None of the predictive studies reviewed for this study used multilevel modeling, nesting students in their neighborhoods; an opportunity to add to the literature on neighborhood effects and high school graduation status exists here.

Previous Predictive Models

Studies designed to test models predicting high school graduation status with a set of predictor variables have been found in the literature for over four decades. Models have employed a variety of variables aimed at predicting graduation outcomes including observable academic variables (e.g. Soland, 2013), latent constructs (e.g. Battin-Pearson et al., 2000), family

and parent-related variables (e.g. Song, Benin, & Glick, 2012), and other nonacademic variables (e.g. McWilliams, Everett, & Bass, 1997). Studies exploring the predictive nature of in-school variables like academic performance, student behavior, and student attendance have done so using elementary school data (e.g. Alexander, Entwisle, & Horsey, 1997), middle school data (e.g. Rumberger & Larson, 1998), and high school data (Parr & Bonitz, 2015).

The literature exploring the utility of these predictive models has reported findings in three ways predominantly. Some studies have reported on the number (usually expressed as a percent) of cases that are correctly classified. Data analyzed using discriminant function analysis or logistic regression analyses are able to produce this output. Studies that present their findings in this manner have tended to correctly classify cases between 72% and 78% of the time (Jimerson et al., 2000; Lloyd, 1978; Ou & Reynolds, 2008; Owens, Morris, & Lieberman, 2001). Of these studies, they were split in their ability to better predict students completing school (Lloyd, 1978; McWilliams, Everett, & Bass, 2000) and in their ability to better predict students dropping out of school (Jimereson et al., 2000; Owens, Morris, & Lieberman, 2001). McWilliams, Everett, and Bass's (2000) model produced the strongest classification figures, correctly classifying up to 90% of cases included in the study. However, the way in which they tested their variables makes it difficult to draw substantive inferences from the findings. Instead of testing variables for their individual predictive power, they grouped variables into one of six clusters. For example, one of their findings was that *Social Snythesis* [sic]/*Integration Variables* correctly classified cases 88% of the time for females and 83% of the time for males. However, this cluster of variables was comprised of four variables: (1) extracurricular activities; (2) school membership; (3) popularity; and (4) whether or not the student had been arrested. Classification rates over 80% are potentially strong; however, it is difficult to draw inferences about what that

might mean if the analysis ends at the cluster level. It is also difficult to adequately assess the strength of a predictive model based solely on an overall classification rate. For example, a model that yielded a high measure of overall classification success that also poorly classified non-graduates would have little utility if the overarching goal was to improve the academic outcomes for students who are at risk of not earning a high school diploma. If a factor analysis had been able to distill a single factor out of these four variables, that would have been more useful, however that was not reported to be the case in this study.

Another common way findings are reported for studies testing predictive models is in terms of the percent of variance that is explained by the predictor variables (e.g. Mac Iver & Messel, 2013). With linear regression this is typically expressed as an R-squared statistic (Tabachnick & Fidell, 2013). For logistic regression analyses, there is no R-squared statistic that provides a similar goodness-of-fit indication. However, pseudo R-squared values ranging from zero to one are often reported that can be interpreted the same way as a standard R-squared statistic (Hosmer & Lemeshow, 2000). Most of the studies included in this review of the literature produced a pseudo R^2 value between .25 and .35 for their full model. Two studies performed better than the rest. In terms of Level 1 predictors, Mac Iver and Messel's ABC early warning indicator models including ninth grade variables produced pseudo R-squared values between .43 and .46. Their relatively simple model with very few predictor variables substantially outperformed several of the other models that featured several times more predictor variables, and since most of their variables were in-school variables, it has the potential of being very useful to practitioners in school buildings. In terms of Level 2 predictors, Rumberger and Palardy's (2005) model investigating predictors using multilevel modeling explained over 57%

of the variance at the school level (Level 2). These two studies in particular present marks that future studies testing predictive models should strive to meet.

Finally, studies testing predictive models are often reported in terms of log likelihood (e.g. Tobin & Sugai, 1999). These studies typically build models from a null model up, adding additional sets of variables and testing the difference that exists between models (Raudenbush & Byrk, 2002). Four models included in this review of the literature reported log likelihood improvement for their models. However, this statistic has no comparative usefulness outside of a given study. Whereas the pseudo R-squared statistic can be used to compare the utility of models across studies, log likelihood values are relatively useless, except when used to compare models within a study (Hosmer & Lemeshow, 2000).

Empirical work analyzing the ability of variables to predict student graduation status has employed a number of types of statistical analyses. The type of statistical procedure that is appropriate is always informed by the nature of the research question(s). Although logistic regression analysis was the most often used type of analysis in the studies included in this review, several approaches were used. See Table 4 for the types of analyses employed by each study.

Table 4

Data Analysis Used in Studies Predicting High School Graduation Status

Authors, Year	Type of Analysis Employed
Stroup & Robins, 1972	OLS regression analysis
Lloyd, 1978	Discriminant function analysis

Barrington & Hendricks, 1989	ANOVA
Astone & McLanahan, 1991	OLS & Logistic regression analysis
Ensminger & Slusarcick, 1992	Logistic regression analysis
McNeal, 1995	Logistic regression analysis
Alexander et al., 1997	Logistic regression analysis
Rumberger & Larson, 1998	Logistic regression analysis
Goldschmidt & Wang, 1999	Multilevel logistic regression analysis
Tobin & Sugai, 1999	OLS & Logistic regression analysis
Batin-Pearson et al., 2000	Structural equation modeling
Jimerson et al., 2000	Discriminant function analysis
McWilliams, Everest, & Bass, 2000	Logistic regression analysis
Owens et al., 2001	Discriminant function analysis
Rumberger & Palardy, 2005	Multilevel logistic regression analysis
Jimerson & Ferguson, 2007	ANCOVA
Neild et al., 2008	Logistic regression analysis
Ou & Reynolds, 2008	Logistic regression analysis
Hernandez, 2011	Not reported*
Song, Benin, & Glick, 2012	Logistic regression analysis
Cratty, 2012	Logistic regression analysis
Mac Iver & Messel, 2013	Multilevel logistic regression analysis
Soland, 2013	OLS & Logistic regression analysis
Parr & Bonitz, 2015	Structural equation modeling

*Notes: * The report does not identify the type of analysis employed, however the findings are reported in terms of the number of cases correctly classified. This type of classification table is a common output for either logistic regression analysis or discriminant function analysis.*

Predictive Model Methodology. When the aim of a study is to test the impact that latent theoretical constructs have on the graduation outcomes of students, structural equation modeling

is the most appropriate type of analysis to employ (Bollen, 1989; Kline, 2016). That is not the most useful procedure for this study since only observed variables are being investigated. Some past models have used discriminant function analysis for their predictive models (Jimerson et al., 2000; Lloyd, 1978; Owens, Morris, & Lieberman, 2001). Discriminant function analysis is appropriate to use to predict group membership using multiple predictor variables. Group membership must be mutually exclusive for this to be a viable analytical tool (Dattalo, 2010). Since a student cannot be simultaneously classified as both a graduate and a dropout, discriminant function analysis is appropriate for predicting group membership for these two discrete variables. However, limitations exist with this type of analysis. Depending on the size of group membership, sample size can be an issue. The membership of the smallest group should be greater than the number of independent variables included in the analysis (Tabachnick & Fidell, 2013), which can be problematic for models investigating numerous variables with few students who drop out of school.

Logistic regression analysis is another statistical procedure that has been used to analyze the utility of models predicting high school graduation status (e.g. Cratty, 2012). Logistic regression analysis is appropriate for use when more than one predictor variable, any of which can be continuous, categorical, or binary in nature, are used to predict a binary outcome (Hosmer & Lemeshow, 2000; Tabachnick & Fidell, 2013). Logistic regression is necessary to use in lieu of linear regression for categorical outcomes because it corrects for two assumptions that are violated when a variable is categorical. Categorical variables are neither normally distributed, nor linear in nature (Tabachnick & Fidell, 2013). Logistic regression is preferred to discriminant function analysis for models using categorical variables because of the violation of the multivariate normality assumption (Tabachnick & Fidell, 2013). Although data transformations

can be used to correct for this with skewed continuous data, there is no such fix for categorical data (Raudenbush & Byrk, 2002; Tabachnick & Fidell, 2013). Logistic regression is also superior to discriminant function analysis because it allows for the use of non-continuous predictor variables (Dattalo, 2010).

Jimerson and Ferguson (2007) used analysis of covariance (ANCOVA) tests to detect if differences existed between groups of elementary school students who were recommended for transition classrooms and retained, recommended for transition classrooms and promoted, retained, and promoted. ANCOVA tests are appropriate when comparing two or more group means, adjusting for covariates thought to influence the outcome (Howell, 2007; McMillan, 2012). ANCOVA tests will not test the degree to which variables are useful to predict an outcome like graduation status, nor will they yield the classification tables produced using logistic regression analysis or discriminant function analysis indicating the accuracy of prediction. As such, their utility is limited when testing models predicting high school graduation status. Similarly, Barrington and Hendricks (1989) employed a series of analysis of variance (ANOVA) tests in their study to compare the means of graduates, dropouts, students who were no longer attending school after four years, and students still enrolled for a fifth year of high school. ANOVA tests are appropriate for comparing means between two or more groups, however they are usually inferior to ANCOVA tests because they do not allow covariates to be considered in the analysis (Howell, 2007; McMillan, 2012).

Implications for the Current Study

Previous studies in the literature have found relationships between academic and nonacademic variables and high school graduation status. If the overall body of knowledge is to

move forward and build on itself, it is important to test predictor variables found to be significant in previous studies together in the same model to explore the extent to which the findings are replicated. Also, although the ability of nonacademic demographic variables to predict graduation status has been explored in many studies, none of the previous empirical work has considered nesting students in their out-of-school communities to explore the predictive nature of these variables hierarchically. This study seeks to learn: (1) the utility of the findings from the studies outlined in this review when they are tested together and (2) the utility of community-level variables in nested models.

III. METHODOLOGY

The purpose of this study was to test two series of models designed to predict high school graduation status using student-level, school-level, and community-level data. This chapter describes the research design and methodology for this study. Included are descriptions of the population, sample, the models that were developed and tested, data analyses, and potential delimitations for the study. This study was guided by the following research questions.

1. To what extent do academic variables predict high school graduation status?
    a. To what extent do individual-level academic variables predict student high school graduation status?
    b. To what extent do school building-level variables predict high school graduation status?
2. To what extent do nonacademic variables predict high school graduation status?
    a. To what extent do individual levels of student engagement predict high school graduation status?
    b. To what extent do demographic variables predict high school graduation status?
    c. To what extent do community-related variables predict high school graduation status?

Research Design

This study was a predictive study using longitudinal administrative data to explore the predictive ability of student-level, school-level, and community-level variables. Secondary data is appropriate for use in this study since the variables included are observed variables that are already collected by the school district for other uses. The research questions in this study were answered with several secondary data analyses outlined in the pages that follow.

Data Sources & Participants

That data that were used to carry out this study were obtained from a single urban school district located in a mid-sized city in the Mid-Atlantic region of the United States. Over 32000 students attend school in the district, and more than seven in ten of these students are eligible for free or reduced price lunch. The student population is ethnically diverse, and a majority of the students identify as African American. Data from the class of 2015 and 2016 graduation cohorts will be combined for this study, yielding a total sample of 4561 participants. The city in which the school district resides is home to a large military installation. Students transferring into the school district after beginning the ninth grade were excluded from the sample. A more detailed description of the sample can be found at the beginning of the next chapter.

Most of data used in this study were obtained from observed measures included in an administrative dataset obtained from a school district. Those that were not were obtained via official dissemination from the United States government by way of the U.S. Census Bureau. However, error can still exist in these data. One possible source of error could result from data entry mistakes. These have been reduced in recent years in the school district included in this study. Since the 2011-2012 academic year, each high school in the district has employed graduation coaches whose mission is to increase graduation rates (Marshall, 2016). A finding

from a 2016 evaluation of the program was that graduation coaches worked to find and correct instances of data entry error. This was a necessary step to take in order to identify students who were at risk of not graduating. Test scores are another possible source of error (Koretz, 2008; McMillan, 2012). By their very nature, standardized test scores are observed scores resulting from true scores and error. Also, subjectivity is a necessary part of student grades and any inconsistencies in grading across students, or across teachers, could result in error as well. Possible threats to validity include participant maturation and attrition (McMillan, 2012). This longitudinal predictive study examined the relationship between a range of academic and nonacademic variables and high school graduation status. However, students could change and grow in ways not captured by the variables not included in this study that impact their eventual graduation status and that must be acknowledged. Also, students who transfer out of the school district are excluded from the analyses. This is a limitation of the study's design. If students who leave the district differ systematically from those who remain in the district, this could be a threat to the validity of the inferences made from the findings.

Predictive Models

Individual-level variables will be tested in two series of models. The first series of models tested academic and nonacademic predictor variables that have been identified in the literature as being significant predictors of high school dropout. The sample was drawn from two graduation cohorts, and they were tested together in each of the models included in this study. Models using eighth and ninth grade predictor variables tested separately to avoid multicollinearity issues. School-level variables were included in these models as well. The second series of models examined the predictive ability of zip code-level variables to approximate impact a student's neighborhood has on his or her graduation status.

The inclusion and exclusion rules that define the sample were the same for both series of models. Students who transferred out of the school district were excluded from the sample, as were those who are coded as being deceased. The state in which the school district resides considers students who transfer into a graduation cohort to be a part of the cohort, and their outcomes count towards the school's graduation statistics. However, much of the literature that used similar administrative data at the school district level excludes students who transfer in to the district after starting the ninth grade. In this study, all models included in the two series of models were tested excluding students who transferred into the district after the ninth grade. This is both consistent with the literature and a pragmatic decision as well. Students transferring into the school district under investigation would have missing data for many of the variables included in the models.

Series one models. The literature review conducted for this study found several individual-level academic and nonacademic, school-level, and demographic variables associated with decreased rates of high school graduation and increased rates of high school dropout. The first series of models answered both research questions investigating the predictive ability of individual-level academic and nonacademic variables, as well as school-building level variables. Eighth and ninth grade in-school variables were tested in separate models to avoid multicollinearity issues.

Predictor variables. The models testing eighth and ninth grade predictor variables included most of the same predictors. The following demographic variables were included in both models: (1) race; (2) gender; (3) eligibility for free or reduced price lunch [FRL]; (4) status as a special needs student; and (5) whether the student is over age for grade. Race was first coded as a categorical variable with four categories: (1) African American; (2) White; (3)

Latino/a; and (4) Other Race. The Other Race category was comprised of students who were coded as Asian, Pacific Islander, Alaskan Native, American Indian, or as having more than one racial/ethnic background. Each of these categories would have been too small to test individually. Four dummy variables were then created, one for each race category. A student was considered overage for grade if he or she was above the age of 14 entering the ninth grade. Students were considered to be special needs students or gifted if he or she had an IEP indicating as such.

Both models also explored the predictive ability of individual-level academic and nonacademic variables including: (1) number of English and math courses failed; (2) math standardized test score; (3) Reading standardized test score; (4) enrollment in Algebra I by the eighth grade; (5) attendance; and (6) mobility. Failure in English and math was operationalized similar to how Neild et al. (2000) and Owens et al. (2001) did in their work. Course failure in middle school was considered as earning the grade of D or F. Also, for the purposes of this study, a student was identified as having failed a course if he or she earned a D or an F in any of the four marking periods during the eighth grade. Similarly, ninth grade course failure was operationalized as having earned an F in a course in either semester. Two standardized tests related to English are administered in the eighth grade, one for Reading and one for Writing. Only the Reading test scores were included in this model. Enrollment in Algebra I was operationalized as a dichotomous variable (students enrolled in the course by the eighth grades vs. students not enrolled in the course). Similar to the work of Cratty (2012), attendance is operationalized as a pair of dichotomous variables. The first attendance variable indicates whether a student has missed more than two weeks of instruction, or ten days of school. The second attendance variable indicates whether a student has missed more than three weeks of

instruction, or fifteen days of school. Mobility was also operationalized as a dichotomous variable. For the eighth grade model, this variable indicated whether a student had changed schools between the sixth and eighth grades. For the model employing ninth grade individual-level variables, student mobility was operationalized to indicate whether a student had changed schools between the sixth and ninth grades, excluding promotion to high school.

The model testing eighth grade predictor variables also explored the predictive ability of school-level variables. Table 5 provides a description of the middle schools in the school districts, including accreditation status and school-wide proficiency rates on standardized tests. There exists a wide range of educational experiences for students at the middle school level in this school district. Six of the nine schools were fully accredited, two were accredited with warning, and one school was denied accreditation. By comparison, 82% of schools are fully accredited statewide and only five percent have been denied accreditation.

Table 5.

Description of Middle Schools included in Series 1-Eighth Grade Model

School	Accreditation Status	English	Math
A	Full	78%	72%
B*	Full	80	77
C*	Full	96	96
D	Denied	73	76
E	Accredited with Warning	77	76
F	Full	84	83
G	Full	79	76
H	Accredited with Warning	73	69
I	Full	81	84
State Benchmark		**75**	**70**

*Note: Figures represent the percent of students in each school that scored Proficient on state-issued standardized tests; * School enrolled grades K-8. These figures are based on the 2010-11 school year and were obtained from the state Department of Education website. No citation accompanies this to ensure anonymity for the participating school district.*

In this model, student assignment to schools was accounted for with clustering. Students who attended more than one school during the eighth grade were assigned to the school in which they were enrolled for more than 100 school days. If the student was not enrolled in a single school for more than 100 days, the school in which he or she finished the school year was selected as the school to which he or she was assigned for the purposes of this study. As a result of this decision, poor attendance was underestimated in some cases. School-level predictors include: (1) school-wide attendance rate; (2) school-wide FRL rate; (3) school-wide suspension rate; (4) school-wide math proficiency rate; and (5) school-wide English proficiency rate. The school-wide suspension rate is a measure of the percent of students who received an out-of-

school suspension during the academic year. Proficiency in math and English is operationalized as the percent of students in the school who scored proficient on the state-issued standardized tests. The full model testing eighth grade predictor variables is as follows.

$$Logit\ (\pi_i) = \beta_0 + \beta_1\gamma' + \beta_2\delta'$$

Where γ' represents a vector of academic and demographic covariates described above, and δ' represents a vector of fixed effects for each of the 9 schools included in the analysis.

The model including ninth grade variables included the same individual-level predictors with two exceptions: (1) course failure is determined only by the number of failing grades that a student receives; a grade of a D is considered passing for the high school grades; and (2) enrollment in Algebra I by the eighth grade is omitted from the ninth grade model and replaced with two additional variables for AP course enrollment. Some students in the dataset took standardized tests for more than one math subject in the ninth grade. There were 226 students in the dataset for whom this was the case. Of the 177 students, 129 also took a Geometry test, 12 students also took a middle school Math test, 19 students also took an Algebra II test, 64 students took an Algebra I test for the second time, and six students also took a version of the Algebra I test designed for students for whom English is a second language. Since Algebra I is the course that is typically prescribed in the curriculum for students in the ninth grade, this was the test that was selected in each case. For instances where two Algebra I tests were taken, the first attempt was selected for inclusion in this study. No student in this dataset passed a second attempt of Algebra I when the test was retaken due to not earning a proficient score on the first attempt. For instances where a student took two versions of an Algebra I test and one version was designed for students for whom English is a second language, the latter was selected for inclusion in this

study. Algebra I is a math subject and one's ability to read English can introduce construct irrelevant error into test scores. Selecting tests designed for English language learners can reduce this error.

The ninth grade model includes four school-level variables for the high schools in the school district in place of the middle school school-level variables. Similar to the treatment of middle school assignment, students who attended more than one school during the ninth grade were assigned to the school in which they were enrolled for more than 100 school days. If the student was not enrolled in a single school for more than 100 days, the school in which he or she finished the school year was selected as the school to which he or she was assigned for the purposes of this study. Table 6 provides a description of the five high schools in the school district, including accreditation status and school-wide test proficiency rates.

Table 6.

Description of High Schools included in Series 1-Ninth Grade Model

School	Accreditation Status	English	Math
A	Full	75%	71%
B	Partial	78	61
C	Full	90	87
D	Full	88	82
E	Full	81	82
State Benchmark		**75**	**70**

Note: Figures represent the percent of students in each school that scored Proficient on state-issued standardized tests. These figures are based on the 2015-16 school year and were obtained from the state Department of Education website. No citation accompanies this to ensure anonymity for the participating school district.

The full model testing ninth grade variables is as follows.

$$Logit\ (\pi_i) = \beta_0 + \beta_1\gamma' + \beta_2\delta'$$

Where γ' represents a vector of academic and demographic covariates described above, and δ' represents a vector of fixed effects for each of the X schools included in the analysis. A complete list the variables used in this study, as well as descriptions of what each represents can be found in Table 7.

Table 7.

Description of Variables Included in the Models

Variable Name	Used in Models	Description
Individual-Level Demographic Variables		
black	S1a, S1b, S2	Dichotomous dummy variable indicating whether a student was classified as “Black” in record data (1=Yes, 0=No)
white	S1a, S1b, S2	Dichotomous dummy variable indicating whether a student was classified as “White” in record data (1=Yes, 0=No)
latino	S1a, S1b, S2	Dichotomous dummy variable indicating whether a student was classified as “Latino” in record data (1=Yes, 0=No)
other_race	S1a. S1b, S2	Dichotomous dummy variable indicating whether a student was classified as “Asian,” “Am. Indian,” “Hawaiian,” or “2 or More” in record data (1=Yes, 0=No) Students not classified as “Black,” “White,” or “Latino” were captured in this variable
male	S1a, S1b, S2	Dichotomous variable indicating how a student’s gender was classified in record data (1=Male; 0=Female)

sped	S1a, S1b, S2	Dichotomous variable indicating whether a student was classified as a special educations student, operationalized by whether or not they had an IEP, "gifted" students excluded (1=SPED; 0=Not SPED)
msmob	S1a	Dichotomous variable indicating whether a student changed schools between grades 6 and 8 (1=Yes; 0=No)
hsmob	S1b, S2	Dichotomous variable indicating whether a student changed schools between grades 6 and 9, excluding promotion to high school (1=Yes; 0=No)
frl	S1a, Slb, S2	Dichotomous variable indicating whether a student was eligible for free or reduced price lunch (1=Yes; 0=No)
overage_8	S1a	Dichotomous variable indicating whether a student was age 14 or older by September 30 of their 8th grade year (1=Yes, 0=No)
overage_9	S1b, S2	Dichotomous variable indicating whether a student was age 15 or older by September 30 of their 9th grade year (1=Yes; 0=No)

Individual-Level Academic Variables

eng8course_pass	S1a	Dichotomous variable indicating whether a student earned the grade of A, B, or C in each marking period in their 8th grade English course (1=Yes; 0=No)
math8course_pass	S1a	Dichotomous variable indicating whether a student earned the grade of A, B, or C in each marking period in their 8th grade math course (1=Yes; 0=No)
eng8_testsc	S1a	Continuous variable representing a student's standardized test score in Reading. This score represents the student's first attempt at the test.
math8_testsc	S1a	Continuous variable representing a student's standardized test score in math. This score represents the student's first attempt at the test.
eng9course_pass	S1b, S2	Dichotomous variable indicating whether a student earned the grade of A, B, C, or D in each semester in their 9th grade English course (1=Yes; 0=No)

math9course_pass	Slb, S2	Dichotomous variable indicating whether a student earned the grade of A, B, C, or D in each semester in their 9th grade math course (1=Yes; 0=No)
math9_testsc	S1b, S2	Continuous variable representing a student's standardized test score in math. This score represents the student's first attempt at the test.
algebra_1	S1a	Dichotomous variable indicating whether a student was enrolled in Algebra I by the 8th grade (1=Yes; 0=No)
AP_1	S1b, S2	Dichotomous variable indicating whether a student was enrolled in at least one Advanced Placement course (1=Yes; 0=No)
AP_3	S1b, S2	Dichotomous variable indicating whether a student was enrolled in 3 or more Advanced Placement course(s) (1=Yes; 0=No)

Individual-Level Nonacademic Variables

absent8_10	S1a	Dichotomous variable indicating whether a student was absent 10 or more days in 8th grade (1=Yes; 0=No)
absent8_15	S1a	Dichotomous variable indicating whether a student was absent 15 or more days in 8th grade (1=Yes; 0=No)
suspend8_ever	S1a	Dichotomous variable indicating whether a student was ever suspended during the 8th grade year (1=Yes; 0=No)
suspend8_3	S1a	Dichotomous variable indicating whether a student was suspended for 3 or more days during the 8th grade year (1=Yes; 0=No)
absent9_10	S1b, S2	Dichotomous variable indicating whether a student was absent 10 or more days in 9th grade (1=Yes; 0=No)
absent9_15	S1b, S2	Dichotomous variable indicating whether a student was absent 15 or more days in 9th grade (1=Yes; 0=No)
suspend9_ever	S1b, S2	Dichotomous variable indicating whether a student was ever suspended during the 9th grade year (1=Yes; 0=No)

suspend9_3	S1b, S2	Dichotomous variable indicating whether a student was suspended for 3 or more days during the 9th grade year (1=Yes; 0=No)

School-Level Variables

sch_engtest	S1a	Continuous variable indicating the percent of 8th grade students at the school who scored proficient or better on their standardized test in Reading
sch_mathtest	S1a	Continuous variable indicating the percent of 8th grade students at the school who scored proficient or better on t their standardized test in math
sch_attend	S1a	Continuous variable indicating the 8th grade attendance rate at the school
sch_suspend	S1a	Continuous variable indicating the 8th grade suspension rate at the school
sch_frl	S1a	Continuous variable indicating the percent of 8th grade students who qualify for free or reduced priced lunch
hs_mathtest	S1b	Continuous variable indicating the percent of 9th grade students at the school who scored proficient or better on their standardized test in math
hs_attend	S1b	Continuous variable indicating the 9th grade attendance rate at the school
hs_suspend	S1b	Continuous variable indicating the 9th grade suspension rate at the school
hs_frl	S1b	Continuous variable indicating the percent of 9th grade students who qualify for free or reduced price lunch

Community-Level Variables

zip_med_income	S2	Continuous variable indicating the median income for the zip code
zip_nonwhite_perc	S2	Continuous variable indicating the percent of residents within the zip code that are not classified as being "White" by the U.S. Census

zip_twoparent_perc	S2	Continuous variable indicating the percent of households with children age 18 or younger in which both parents reside
zip_hsgradrate	S2	Continuous variable indicating the percent of residents in the zip code age 25 or older who have earned a high school diploma
zip_povrate	S2	Continuous variable indicating the percent of residents in the zip code who earn less than the federal poverty rate

Note: The Used in Models column indicates the model(s)in which the variable is used. S1a = Series 1 8th grade model; S1b = Series 1 9th grade model; S2 = Series 2 model nesting students in zip codes

Operationalization of the dependent variable. The dependent variable for the first series of models was dichotomized as those who earn a diploma and those who did not in four years. International Baccalaureate, advanced, standard, and modified diplomas were classified as the same in this study. Students who were still enrolled in school after four years were classified as non-graduates, along with those students who earn a GED. Students classified as dropouts, as well as those classified as having a long-term absence are also considered non-graduates for the purposes of this study. Students who transferred out of the school district, as well as those students who earned Certificates of Completion, and those who were deceased were excluded from analysis. Certificates of Completion are rarely issued in this school district (n = 10) and earning a Certificate of Completion is typically associated with having an IEP with a low-incidence disability. With the exception of the treatment of individuals who earned a Certificate of Completion, this is similar to the Adjusted Cohort Graduation Rate that the Every Student Succeeds Act’s non-regulatory guidance on high school graduation rate calculation suggests using (U.S. Department of Education, 2017). A description of the dependent variable can be found in Table 8.

Table 8.

Description of the Dependent Variable

Outcome	*N* (%)	Recoded Outcome	*N* (%)
IB Diploma	50 (1.1)	Graduate	3051 (65.6)
Advanced Studies Diploma	1224 (26.3)	Non-Graduate	667 (14.3)
Standard Diploma	1631 (35.1)	Excluded from Analysis	935 (20.1)
Modified Diploma	32 (0.7)		
Special Diploma	114 (2.5)		
GED	142 (3.1)		
Dropout	233 (5.0)		
Long-Term Absence	66 (1.4)		
Still Enrolled	245 (5.3)		
Unconfirmed Status	10 (0.2)		
Certificate of Completion	10 (0.2)		
Deceased	9 (0.2)		
Transferred Out	969 (20.8)		

Analyses. The first research question explored the extent to which academic and nonacademic variables predict high school graduation status, at both the individual-level and school-level. To answer this question, two series of logistic regression analyses were performed – one using eighth grade variables and one using ninth grade variables. STATA was used to conduct the analyses. The variables used in each model are displayed in Table 9.

Table 9.

Variables included in Series 1-Eighth and Ninth Grade Models

Eighth Grade Model Type	Variable	Ninth Grade Model Type	Variable
Demographic	white	Demographic	white
	latino		latino
	other_race		other_race
	male		male
	frl		frl
	sped		sped
	overage_8		overage_9
Academic	eng8course_pass	Academic	eng9course_pass
	math8course_pass		math9course_pass
	eng8_testsc		math9_testsc
	math8_testsc		AP_1
	algebra_1		AP_3
Nonacademic	suspend8_ever	Nonacademic	suspend9_ever
	suspend8_3		suspend8_3
	attend8_10		attend9_10
	attend8_15		attend9_15
	msmob		hsmob
School-Level	sch_engtest	School-Level	hs_mathtest
	sch_mathtest		hs_suspend
	sch_suspend		hs_attend
	sch_attend		hs_frl
	sch_frl		

Prior to analyses being conducted, the data were screened to ensure that the requisite assumptions are met for logistic regression analysis. Logistic regression analysis does not require multivariate normality or homoscedasticity, and does not assume a normal distribution of error terms (Tabachnick & Fidell, 2013). However, there were still assumptions to check prior to analysis. The data being used for these analyses are administrative record data obtained from a

school district. Less than one percent of the data included in this dataset are missing. Any records including missing data were excluded from these analyses. Collinearity diagnostics were run to check for multicollinearity, and variance inflation factor (VIF) values were examined to determine if this assumption was violated (Hosmer & Lemeshow, 2000). Potential issues arise when using logistic regression concerning sample size. Maximum likelihood requires that a minimum of ten cases exist for each variable included in the model (Hosmer & Lemeshow, 2000; Tabachnick & Fidell, 2013). The model with the most variables across the two series of models being tested includes 17 predictors, and therefore this is not a concern in this study.

The analyses of the model using eighth grade variables employed logistic regression analysis, clustering at the middle school level. Ninth grade model was analyzed using logistic regression analysis, clustering at the high school level. Each of the two models was built and tested as follows. First the demographic variables were tested. Individual-level academic and nonacademic variables were added to the demographic variables and tested next. Finally, the full models testing all of the articulated academic and nonacademic, as well as school-level variables were tested. Building a model in this manner is advantageous because it demonstrates how the predictive ability of variables shifts as new sets of variables are added to the model.

Series two models. The review of the literature conducted for this study also found several demographic variables associated with decreased rates of high school graduation and increased rates of high school dropout. The second series of models were tested to answer the second research question and investigate the ability of zip code-level variables to predict high school graduation status.

Predictor variables. The second series of models used the same set of demographic variables that were employed in the first series of models including: (1) race; (2) gender; (3) eligibility for free or reduced price lunch [FRL]; (4) disability status; (5) whether the student is over age for grade. Along with demographic variables, the second series of models investigated the predictive ability of community-level variables, which were examined at the zip code level using data obtained from the U.S. Census. Zip codes represent arbitrary boundaries surrounding the establishment of U.S. Post Offices. However, the experiences that exist across zip codes can vary vastly, especially in terms of socioeconomic status. Table 10 provides a description of the 11 zip codes in which students in the school district reside.

Table 10.

Description of Zip Codes included in Series 2 Model

Zip code	Median Income	Non-White	Two-parent Households	Educational Attainment	Poverty Rate
A	$46684	61.4%	55.1%	85.0%	16.6%
B	45903	33.6	48.5	90.0	14.1
C	27375	93.6	24.8	76.4	36.6
D	40238	45.6	60.0	90.4	18.2
E	49812	43.8	59.4	90.8	27.1
F	45029	57.3	46.8	84.4	13.9
G	80625	32.1	77.3	99.4	15.3
H	42802	70.1	35.9	82.2	17.5
I	39505	38.8	46.6	91.4	27.0
J	52234	37.2	50.1	86.4	16.3
K					

Note: Educational attainment reflects the percent of individuals residing within the zip code who have earned a high school diploma or equivalency.

Community-level predictors incorporated in the second series of models include: (1) median household income; (2) percent of non-White residents; (3) percent of families with two-

parent households; (4) educational attainment; and (5) poverty rate. Educational attainment was measured by the percent of residents over the age of 24 who have graduated from high school. This served as an approximate measure of the number of adults living around a student who have achieved the milestone of earning a diploma. The poverty rate is the percent of residents who live below the federal poverty line.

The full multilevel model for the second series of predictors is as follows.

Level-1 (Student-Level) Model:

$$Logit\left(\pi_{ij}\right) = \beta_{0j} + \beta_{0j}\omega'_{ij}$$

Level-1 Variance of Random Effect = $1/\left[\pi_{ij}\left(1-\pi_{ij}\right)\right]$

Level-2 (Zip Code-Level) Model:

$$\beta_{0j} = \gamma_{00} + \gamma_{01}\delta' + u_{0j}$$

Parameterization of Level-2 Random Effect:

$$u_{0j} \sim N(0, \tau_{00})$$

Where ω' represents a vector of academic and demographic covariates described above and in Figure 1, and δ' represents a vector of community (zip code)-level predictors described above and in Figure 3. All Level 1 predictors will be treated as fixed effects at the second level.

The same ninth grade in-school predictors employed in the first series of models were also tested in the second series of models. These include: (1) English and math courses failed; (2) math standardized test score; (4) enrollment in Algebra I; (5) attendance; and (6) mobility.

These variables were operationalized in the same manner in which they were in the first series of models.

The full model for the logistic regression analysis is as follows.

$$Logit\ (\pi_i) = \beta_0 + \beta_1\gamma' + \beta_2\delta'$$

Where γ' represents a vector of academic and demographic covariates described above, and δ' represents a vector of fixed effects for each of the 11 zip codes included in the analysis. The variables included in this model are presented in Table 11**.**

Table 11.

Variables included in Series 2 Models

Type	Variable
Demographic	white latino other_race male frl sped overage_9
Academic	eng9course_pass math9course_pass math8_testsc AP_1 AP_3
Nonacademic	suspend9_ever suspend9_3 attend9_10 attend9_15 hsmob

Zip Code-Level	zip_med_income zip_nonwhite_perc zip_twoparent_perc zip_hsgradrate zip_povrate

Operationalization of the dependent variable. Similar to the first series of models, the dependent variable for the second series of models were dichotomized as those who earn a diploma and those who did not in four years. Students who were still enrolled in school after four years were classified as non-graduates, along with those students who earn a GED. Students classified as dropouts, as well as those classified as having a long-term absence were also considered non-graduates for the purposes of this study. Students who transferred out of the school district, as well as those students who earned Certificates of Completion, and those who were deceased were excluded from analysis.

Analyses. The second research question explored the extent to which nonacademic variables, including community-level variables, predict high school graduation status. To answer this question, a series of multilevel logistic regression analyses were performed. STATA was used to conduct the analyses. Prior to the analyses being conducted, the data were screened to ensure that the requisite assumptions are met for multilevel logistic regression analysis, including screens for missing data and multicollinearity. The analysis employed multilevel logistic regression, nesting students (Level 1) in home address zip codes (Level 2). Models tested using multilevel modeling should have at least ten units at Level 2 (Raudenbush & Byrd, 2002), and since students are nested in 11 zip codes this assumption is met. An interclass correlation coefficient was calculated using Snijders and Bosker's (1999) method to determine the appropriateness of using multilevel modeling for the analysis.

$$ICC = \frac{\tau_{00}}{\tau_{00} + \frac{\pi^2}{3}}$$

The ICC statistic illustrates how much of the variance is explained by Level 2 predictors and is used to justify the use of multilevel models. τ_{00} represents the amount of variance explained at Level 2; $\tau_{00} + \frac{\pi^2}{3}$ represents the total variance explained.

To build the model, first the null model was tested. Demographic variables were added and tested next. Ninth grade academic variables were added to the demographic variables and tested next. Finally, the full model testing all of the articulated demographic and ninth grade academic and nonacademic variables were tested. Logistic regression analysis was also conducted, clustering at the zip code level. For this analysis, models were built in the same manner that they were for the multilevel logistic regression analysis.

Summary

Two series of models were created and tested, examining the ability of demographic, academic, and nonacademic variables to predict high school graduation status. The first series of models explored the predictive ability of school-level variables and tested eighth and ninth grade predictor variables separately. Two separate logistic regression analyses were conducted to analyze these models. The second series of models explored the predictive ability of the same demographic and ninth grade academic and nonacademic variables, this time in conjunction with zip code-level variables. This model was analyzed two different ways. First, a multilevel logistic regression analysis was conducted, nesting students in home address zip code. Second, a logistic regression analysis was run for the purpose of comparing findings across both series of models. The four analyses run across two series of models were appropriate for understanding

the relationship between demographic, academic, nonacademic, school-level, and zip-code level variables and a student's propensity to graduate from high school.

CHAPTER IV – FINDINGS

This chapter outlines the findings resulting from testing the two series of models described in the previous chapter outlining this study's methodology. This chapter begins with descriptive statistics and an analysis of the prerequisite assumptions associated with logistic regression analysis. For each series of models, tables with point estimates, robust standard errors, *p*-values, and confidence intervals is presented, along with a pseudo R-squared value. Classification tables are also presented for each model tested. The chapter closes with an overall summary of the findings.

Descriptive Statistics

Before data analysis began, descriptive statistics were produced for each variable included in the two series of models. Descriptive statistics are provided in Table 12.

Table 12.

Descriptive Statistics for Demographic Variables Included in the Models

Variable	Class of 2015 *N* (%)	Class of 2016 *N* (%)	Total *N* (%)
African American	1158 (60.9)	1195 (62.2)	2353 (61.5)
White	432 (23.7)	412 (22.6)	844 (23.1)
Latino/a	105 (6.7)	103 (6.6)	208 (6.6)
Asian American*	51 (2.8)	50 (2.7)	101 (2.7)
Hawaiian/Pacific Islander*	6 (0.3)	6 (0.3)	12 (0.3)
American Indian*	9 (0.5)	6 (0.3)	15 (0.4)
Two or more races*	87 (4.7)	95 (5.1)	182 (4.9)
Female	887 (50.7)	915 (50.3)	1802 (50.1)
frl	1201 (69.2)	1213 (69.3)	2414 (69.2)
Special Education status	235 (10.2)	280 (11.9)	515 (11.1)
Over age for grade (8)	583 (33.6)	540 (30.8)	1123 (32.2)
Over age for grade (9)	819 (41.2)	771 (39.0)	1590 (40.1)
Changed schools (6-8)	257 (14.6)	404 (22.1)	661 (18.4)
Changed schools (6-9)	348 (20.1)	456 (25.7)	804 (22.9)

*Note: * = Coded as "Other Race" for analysis due to low sample sizes.*

Descriptive statistics for academic variables are displayed in Table 13. All values represent a percent of the cohort or total population, except for the scaled standardized test scores which represent mean values; descriptive statistics for scaled standardized test scores are presented in Table 14. Descriptive statistics for nonacademic variables are presented in Table 15.

Table 13.

Descriptive Statistics for Academic Variables Included in the Models

Variable	Class of 2015 *N* (%)	Class of 2016 *N* (%)	Total *N* (%)
Passed English 8 course*	1120 (64.5)	1052 (60.5)	2172 (62.5)
Passed English 8 test	1171 (82.0)	1118 (79.5)	2289 (80.7)
Passed Math 8 course*	900 (51.7)	850 (48.8)	1750 (50.2)
Passed Math 8 test	1159 (82.9)	796 (58.7)	1955 (71.0)
Passed English 9 course**	1358 (72.3)	1428 (76.7)	2786 (74.5)
Passed Math 9 course**	1009 (53.2)	1007 (53.7)	2016 (53.4)
Passed Math 9 test	656 (64.6)	716 (68.5)	1372 (66.5)
Enrolled in Alg I by 8th gr.	253 (13.4)	168 (9.0)	421 (11.2)
Enrolled in at least 1 AP course	696 (30.4)	794 (35.1)	1490 (32.8)
Enrolled in 3 or more AP courses	465 (20.3)	505 (21.9)	970 (21.1)

*Note: * Passing courses in the 8th grade includes earning a grade A-C in each of four marking periods.*
*** Passing courses in the 9th grade includes earning a grade A-D in each of two semesters.*

Table 14.

Descriptive Statistics for Scaled Standardized Test Scores

Variable	Class of 2015 M (SE)	Class of 2016 M (SE)	Total M (SE)
English 8 test	457.7 (1.72)	454.0 (1.73)	455.8 (1.22)
Math 8 test	432.6 (1.49)	401.2 (1.06)	417.0 (0.96)
Math 9 test	412.8 (1.42)	413.7 (1.32)	413.3 (0.97)

Table 15.

Descriptive Statistics for Nonacademic Variables Included in the Models

Variable	Class of 2015 *N* (%)	Class of 2016 *N* (%)	Total *N* (%)
8th Grade Attendance (10 or more)	555 (31.9)	562 (32.1)	1117 (32.0)
8th Grade Attendance (15 or more)	300 (17.3)	307 (17.5)	607 (17.4)
8th Grade Suspensions (1 or more)	508 (29.3)	504 (28.8)	1012 (29.0)
8th Grade Suspensions (3 or more)	387 (22.3)	384 (21.9)	771 (22.1)
9th Grade Attendance (10 or more)	729 (36.6)	726 (36.7)	1455 (36.7)
9th Grade Attendance (15 or more)	461 (23.2)	440 (22.2)	901 (22.7)
9th Grade Suspensions (1 or more)	416 (20.9)	381 (19.3)	797 (20.1)
9th Grade Suspensions (3 or more)	286 (14.4)	282 (14.3)	568 (14.3)

Note: Figures in parentheses in the variable field represent days missed from school.

The first series of models investigated the predictive ability of school-level variables. Descriptive statistics for middle schools can be found in Table 16; descriptive statistics for high schools can be found in Table 17. The second series of models investigated the predictive ability of zip code-level variables; descriptive statistics for these can be found in Table 5 in Chapter 3.

Table 16.

Descriptive Statistics for Middle School School-Level Variables in Series 1-Eighth Grade Model

School	English proficiency	Math proficiency	Attendance rate	FRL eligibility	Suspension rate
A	86.5%	83.7%	93.7%	70.9%	34.5%
B	81.6	61.3	93.0	59.5	30.7
C	81.2	72.3	93.5	53.9	23.3
D	95.8	95.8	97.4	35.8	6.1
E	72.3	69.1	90.9	84.8	29.2
F	74.9	54.5	92.7	78.2	36.5
G	85.4	81.5	91.9	62.7	32.6
H	78.0	71.3	92.8	74.5	28.1
I	70.3	63.6	92.7	82.0	27.7

Note: Figures in the first two columns represent the percent of students in each school that scored Proficient on state-issued standardized tests. Data represents that of eighth grade students attending each school.

Table 17.

Descriptive Statistics for High School School-Level Variables in Series 1-Ninth Grade Model

School	Math proficiency	Attendance rate	FRL eligibility	Suspension rate
A	41.9%	89.5%	81.6%	31.5%
B	65.2	89.9	60.0	25.1
C	58.3	90.3	73.9	25.4
D	76.1	91.3	50.6	15.9
E	82.0	91.4	61.6	19.8

Note: Figures in the first column represents the percent of students in each school that scored Proficient on the state-issued standardized test. Data represents that of ninth grade students attending each school.

Finally, descriptive statistics were produced for the dichotomous outcome variable – graduates and non-graduates. The same outcome variable was used for each for of the models tested in this study. They are presented in Table 18.

Table 18.

Descriptive Statistics for the Outcome Variable

	N	%
Graduates	3051	82.1%
Non-Graduates	667	17.9

Analysis of Assumptions

Prior to analyses being conducted, the data were screened to ensure that the requisite assumptions were met for logistic regression analysis. Logistic regression analysis does not require multivariate normality or homoscedasticity, and does not assume a normal distribution of error terms (Tabachnick & Fidell, 2013). However, there are still assumptions to check prior to analysis. Most of the data being used for these analyses are derived from administrative record data obtained from a school district. Less than one percent of the data included in this dataset were missing. Any records including missing data were excluded from these analyses. The data used for the zip code-level variables were derived from the United States Census Bureau's (n.d.) web-based archives. No missing data exist in these variables. Collinearity diagnostics were run to check for multicollinearity. Variance inflation factor (VIF) was examined for each model to determine if this assumption was violated. Variables with values greater than five were

considered to be potential violations of this assumption. In the case of the eighth grade model, this included four of the five school-level variables exceeding this threshold and a fifth school-level variable that was approaching a value of five. VIF values for the Series One eighth grade models are presented in Table 19.

Table 19.

VIF Values for Series One Eighth Grade Models

Original Model Variable	VIF	*p*-value	Adjusted Model Variable	VIF
sch_engtest	25.26	.271	sch_frl	4.35
sch_frl	4.09	.213	sch_engtest	4.23
sch_mathtest	9.81	.620	suspend8_ever	3.10
sch_suspend	9.21	.528	suspend8_3	3.07
sch_attend	4.83	.601	absent8_10	1.90
suspend8_ever	3.11	.000	absent8_15	1.83
suspend8_3	3.08	.008	eng8_testsc	1.81
absent8_10	1.91	.000	math8_testsc	1.54
absent8_15	1.83	.000	white	1.48
eng8_testsc	1.81	.000	frl	1.39
math8_testsc	1.58	.459	math8course_pass	1.38
white	1.53	.015	eng8course_pass	1.37
eng8course_pass	1.39	.085	sped	1.22
frl	1.39	.003	overage_8	1.22
math8course_pass	1.38	.000	other_race	1.11
sped	1.23	.000	latino	1.07
overage_8	1.22	.000	algebra_1	1.06
other_race	1.12	.371	gender	1.05
latino	1.09	.052	msmob	1.02
algebra_1	1.08	.809		
gender	1.05	.000		
msmob	1.03	.011		

Variables highlighted in yellow were removed prior to testing the adjusted model.

Three of the five school-level variables were removed for the adjusted model. Of the five school-level variables, these variables had the highest *p*-values. Multicollinearity issues

disappeared when the model was tested again without these three variables. No variable had a VIF value exceeding five in the adjusted model.

Collinearity diagnostics run on the Series One ninth grade models revealed a similar issue; all four school-level variables had values far exceeding the threshold of five. VIF values for the Series One eighth grade models are presented in Table 20.

Table 20.

VIF Values for Series One Ninth Grade Models

Original Model Variable	VIF	*p*-value	Adjusted Model Variable	VIF
hs_suspend	135.26	.000	hs_frl	3.81
hs_attend	93.26	.000	hs_mathtest	3.73
hs_frl	58.02	.002	suspend9_ever	2.74
hs_mathtest	27.89	.000	suspend9_3	2.68
suspend9_ever	2.75	.042	AP_3	2.65
suspend9_3	2.68	.849	AP_1	2.63
AP_3	2.66	.090	absent9_15	1.97
AP_1	2.64	.000	absent9_10	1.94
absent9_15	1.98	.126	math9_testsc	1.69
absent9_10	1.94	.059	math9course_pass	1.54
math9_testsc	1.69	.252	overage_9	1.47
math9course_pass	1.56	.000	white	1.40
overage_9	1.48	.000	frl	1.37
white	1.41	.620	eng9course_pass	1.25
frl	1.37	.095	other_race	1.14
eng9course_pass	1.25	.003	sped	1.14
other_race	1.15	.862	latino	1.05
sped	1.15	.000	hsmob	1.04
latino	1.06	.209	gender	1.04
hsmob	1.05	.331		
gender	1.04	.008		

Note: Variables highlighted in yellow were removed prior to testing the adjusted model.

In the case of the ninth grade school-level variables, the *p*-values were all similarly significant. Among the four variables, hs_attend was the least sensitive. Between the five high schools in the school district, average attendance ranged from 89.5% to 91.4%. The variables with the highest two VIF values were removed from the model. When the adjusted model without them was run, multicollinearity issues were no longer present.

Multicollinearity diagnostics were also run for the Series Two models testing zip code-level variables. The results were similar to the Series One models; zip code-level variables all had VIF values exceeding five. VIF values for the Series Two models are presented in Table 21.

Table 21.

VIF Values for Series Two Models

Original Model Variable	VIF	*p*-value	Adjusted Model Variable	VIF
zip_hsgradrate	4438.78	.000	zip_twoparent_perc	4.90
zip_nonwhite_perc	4055.89	.000	zip_hsgradrate	3.81
zip_medincome	89.43	.000	suspend9_ever	2.74
zip_twoparent_perc	68.42	.000	suspend9_3	2.68
zip_povrate	14.63	.595	AP_3	2.65
suspend9_ever	2.74	.004	AP_1	2.63
suspend9_3	2.69	.628	absent9_15	1.98
AP_3	2.65	.033	zip_medincome	1.95
AP_1	2.64	.000	absent9_10	1.94
absent9_15	1.98	.240	math9_testsc	1.66
absent9_10	1.94	.018	math9course_pass	1.56
math9_testsc	1.69	.092	overage_9	1.48
math9course_pass	1.56	.006	white	1.40
overage_9	1.48	.000	frl	1.37
white	1.40	.792	eng9course_pass	1.25
frl	1.37	.065	other_race	1.15
eng9course_pass	1.25	.004	sped	1.14
other_race	1.15	.881	latino	1.05
sped	1.15	.000	hsmob	1.05
latino	1.06	.464	gender	1.04
hsmob	1.05	.381		
gender	1.04	.000		

Variables highlighted in yellow were removed prior to testing the adjusted model.

Zip code-level poverty rate was removed due to having the highest *p*-value. The percent of residents living in a zip code that were not classified as "White" was also removed; it was the least malleable of the remaining variables. When the adjusted model was run, no variable had a value exceeding five and no multicollinearity issues were present.

Modeling High School Graduation Status

Analyses for each of the three models were conducted and the findings are presented on the following pages. For each model, a table of odds ratios and *p*-values will be presented for each variable included; pseudo R-squared values will also be presented. Classification tables will be presented for each full model as well.

Series one models. Two separate sets of models were tested in the first series, one including eighth grade variables and another including ninth grade variables. For the eighth grade model, Model 1 examined the impact of demographic variables on high school graduation status. The reference category used throughout this study is an African American female. African American students represent almost two-thirds of the student population in the school district used in this study and were selected as the reference for that reason, and male students were arbitrarily coded as "1" for the dichotomous gender variable. Both White (1.46) and students classified as "Other Race" (1.815) had significantly greater odds of graduating from high school than African American or Latino/a students. The lowest odds ratio was associated with a student being overage-for-grade in the eighth grade. Students age 14 or older by September 30 of their eighth grade year had one-fifth the odds of graduating as someone who was on grade level for their age. Table 22 summarizes the Series One eighth grade models estimated, as well as the proportion of the variation in graduation status explained by each model.

Table 22.

Logistic Regression Findings for the Series One Eighth Grade Models

	(1)	(2)	(3)	(4)
Variables	Model 1	Model 2	Model 3	Model 4
white	1.459**	0.726	0.695**	0.700**
	(0.240)	(0.151)	(0.105)	(0.100)
latino	0.879	0.549*	0.546*	0.460**
	(0.261)	(0.189)	(0.174)	(0.168)
other_race	1.815***	1.172	1.171	1.253
	(0.407)	(0.249)	(0.216)	(0.241)
male	0.673***	0.756***	0.754***	0.754***
	(0.0735)	(0.0594)	(0.0593)	(0.0580)
sped	1.600***	4.802***	4.720***	4.756***
	(0.268)	(1.246)	(1.230)	(1.227)
overage_8	0.232***	0.432***	0.432***	0.427***
	(0.0232)	(0.0579)	(0.0569)	(0.0536)
msmob	0.718*	0.667**	0.668**	0.664***
	(0.143)	(0.109)	(0.108)	(0.104)
frl	0.325***	0.549***	0.567***	0.618***
	(0.0456)	(0.103)	(0.107)	(0.110)
eng8course_pass		1.600**	1.562*	1.554*
		(0.372)	(0.380)	(0.378)
math8course_pass		2.032***	2.103***	2.100***
		(0.394)	(0.419)	(0.419)

eng8_testsc		1.011***	1.011***	1.011***
		(0.00141)	(0.00146)	(0.00145)
math8_testsc		1.001	1.001	1.001
		(0.00223)	(0.00224)	(0.00224)
algebra_1		1.030	1.028	1.022
		(0.184)	(0.184)	(0.181)
absent8_10		0.617***	0.626***	0.619***
		(0.0782)	(0.0828)	(0.0845)
absent8_15		0.594***	0.578***	0.589***
		(0.0571)	(0.0570)	(0.0597)
suspend8_ever		0.496***	0.498***	0.499***
		(0.0633)	(0.0671)	(0.0680)
suspend8_3		0.752***	0.756**	0.748***
		(0.0720)	(0.0826)	(0.0808)
sch_engtest			0.00645	0.00700
			(0.0344)	(0.0370)
sch_frl			0.0545	0.0572
				(0)
0b.frl_rec#1.white				1.051
				(0.235)
0b.frl_rec#1.latino				3.165***
				(0)
0b.frl_rec#1.other_race				0.824
				(0.404)
Constant	24.37***	0.0784***	27.17	22.92
	(3.987)	(0.0709)	(149.5)	(124.7)

Observations	2,931	2,741	2,741	2,741
Pseudo R^2	0.131	0.303	0.306	0.307

Note: Odds ratios are reported for each model with robust SE in parentheses.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Model 2 explored the impact of adding eighth grade academic and nonacademic variables to the demographic variables. Students who earned an A, B, or C each marking period in their eighth grade math course had two times the odds of graduating from high school when compared to a student who earned at least one D or F in eighth grade math. Chronic absenteeism was also found to be a strong predictor of high school graduation status. A student who missed three weeks of instruction or more (15+ days) had half the odds of graduating from high school as was a student who missed fewer than 15 days in the eighth grade (OR=0.594). The pseudo R-squared value increased from .131 to .303 as a result of the addition of the academic and nonacademic variables to the model, indicating improved model fit.

Model 3 explored the impact of adding school-level variables to the demographic, academic, and nonacademic variables. Neither of the two variables were found to be significant predictors. Their inclusion in the model did not substantially alter the pseudo R-squared value either (.303 to .306). Model 4 added interaction terms to the model. Based on findings from previous predictive models in the literature, it was hypothesized that there would be an interaction effect from the eligibility for free or reduced lunch (frl) and the race variables. The inclusion of the interaction effects made for better specified models. Only one of the interaction terms included yielded significant findings. Latino/a students who were eligible for free or reduced priced lunch had better odds of graduating from high school compared to students who

were either just Latino/a or just eligible for free or reduced priced lunch. Students classified as being Latino/a had less than half the odds of their African American peers of graduating from high school (OR = .460). Students who were eligible for free or reduced priced lunch also faced substantially reduced odds of graduating when compared to peers from higher socioeconomic status backgrounds (OR = .618). Students who were both Latino/a and eligible for free or reduced price lunch still had reduced odds of graduating when compared to African American peers who were not eligible for the lunch subside (OR = .899); however, their graduation outcomes were improved over peers who were only Latino/a or only eligible for free or reduced price lunch. In this study, however, this could be a function of a small sample size. The school district from which the sample was drawn has a smaller than typical Latino/a population (approximately 6% compared to over 22% nationally). Only 166 students identified as being Latino/a were included in the sample. It is possible that this finding holds true in this context, even if the larger trends suggest that this is a counterintuitive finding. Although the interaction effect was significant, the addition of the interaction effects added little to the overall explanatory power of the model. The pseudo R-squared value remained approximately the same (.306 to .307).

Overall, the greatest explanatory power in the Series One eighth grade models in terms of an improvement in the amount of the variance explained occurred due to the addition of academic and nonacademic eighth grade predictor variables. A classification table was produced for the full eighth grade model; this is presented in Table 23.

Table 23.

Initial Classification Table for Series One Eighth Grade Models

	Observed Graduates *N* (%)	Observed Non-Graduates *N* (%)	Total
Classified Graduates	2217 (96.1)	267 (61.4)	2030
Classified Non-Graduates	89 (3.9)	168 (38.6)	711
Overall	2306	435	2741

The initial classification table had an overall classification rate of 87.01%. The model correctly classified students who graduated from high school over 96% of the time. However, those who did not graduate from high school on time were only correctly classified 38.62% of the time. To craft a more usable classification table, sensitivity and specificity were graphed to arrive at an appropriate probability cutoff point. See Figure 1 for the Series One eighth grade model sensitivity graph.

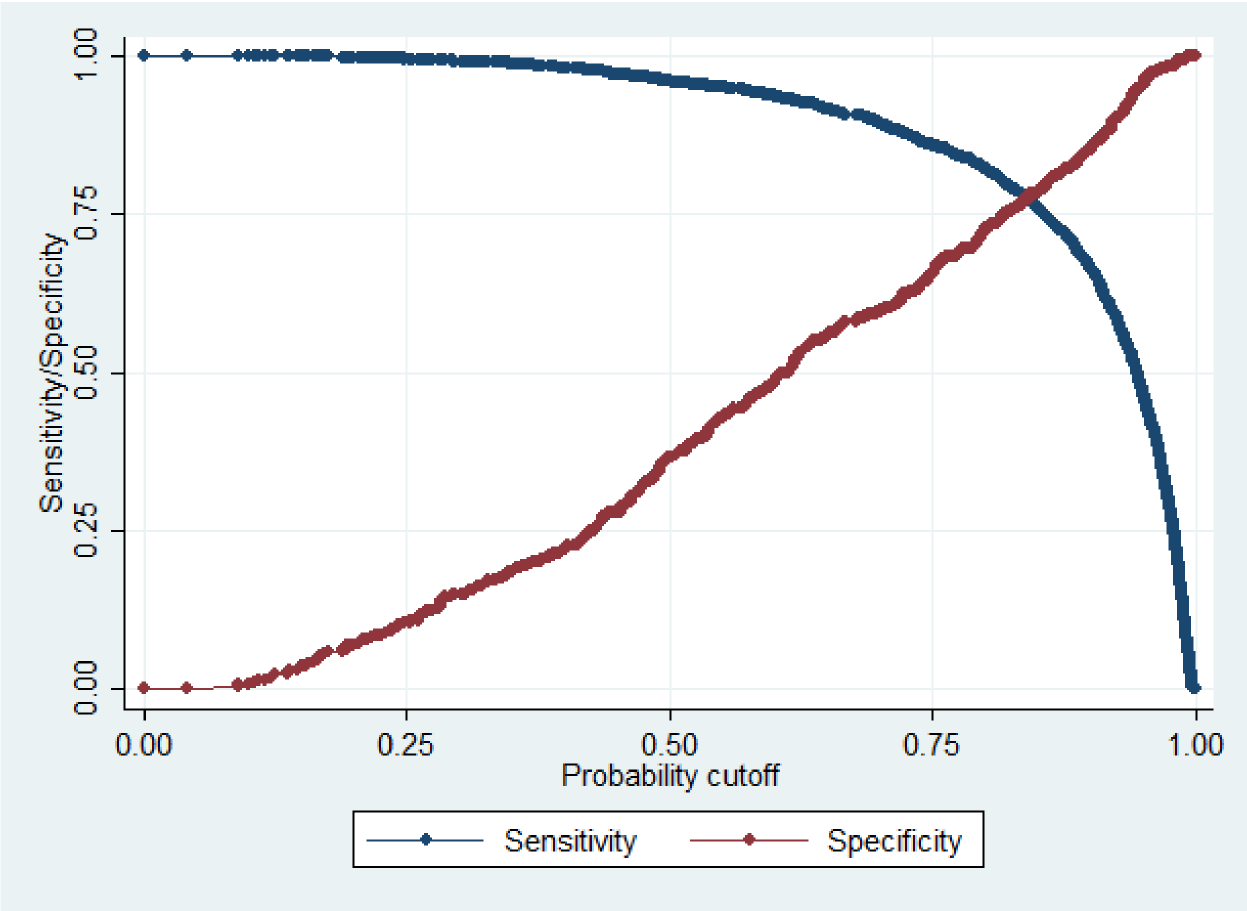

Figure 1. Probability Cutoff for Series One Eighth Grade Models

A new classification table was produced, specifying the cutoff point at 0.8, since that is approximately the point at which sensitivity and specificity meet. The overall classification rate is reduced in the adjusted classification table (80.96%), however specificity – or in the case of this study, the ability to correctly predict non-graduation status – is substantially improved (71.72%). The adjusted classification table is presented in Table 24.

Table 24.

Adjusted Classification Table for Series One Eighth Grade Models

	Observed Graduates *N* (%)	Observed Non-Graduates *N* (%)	Total
Classified Graduates	1907 (82.7)	123 (18.3)	2030
Classified Non-Graduates	399 (17.3)	312 (71.7)	711
Overall	2306	435	2741

For the ninth grade model, Model 1 examined the impact of demographic variables on high school graduation status. The largest odds ratio value was associated with disability status. Students identified as receiving special education services, excluding those identified as "gifted," were more than two times as likely to earn a high school diploma when compared to students who did not receive special education services. When demographic variables alone were considered, male students had approximately two-thirds the odds of female students to graduate from high school (OR=0.677). Students who changed schools between the sixth and ninth grades (excluding for promotion to high school), students who were 15 years old or older on September 30 of their freshman year, and those who were eligible or reduced price lunch all had reduced odds of graduating from high school than students who attended the same middle and high school, were on grade level per their age, and were not eligible for free or reduced price lunch. Table 24 summarizes the Series One ninth grade models estimated, as well as the proportion of the variation in graduation status explained by each model.

Table 25.

Logistic Regression Findings for the Series One Ninth Grade Models

	(1)	(2)	(3)	(4)
Variables	Model 1	Model 2	Model 3	Model 4
white	1.311	0.900	0.883	0.881
	(0.243)	(0.263)	(0.266)	(0.551)
latino	0.886	0.693	0.707	0.463***
	(0.205)	(0.185)	(0.212)	(0.0913)
other_race	1.756**	0.889	0.897	1.211
	(0.474)	(0.709)	(0.729)	(0.958)
male	0.677***	0.736***	0.736***	0.741***
	(0.0469)	(0.0771)	(0.0769)	(0.0840)
sped	2.018***	4.411***	4.427***	4.511***
	(0.370)	(0.846)	(0.823)	(0.791)
overage_9	0.115***	0.320***	0.316***	0.313***
	(0.0111)	(0.0332)	(0.0326)	(0.0332)
hsmob	0.578***	0.814	0.798	0.795
	(0.0980)	(0.189)	(0.193)	(0.194)
frl	0.373***	0.534*	0.531*	1.389
	(0.102)	(0.198)	(0.197)	(0.960)
eng9course_pass		2.372***	2.389***	2.350***
		(0.674)	(0.706)	(0.685)
math9course_pass		1.631***	1.614***	1.576***
		(0.214)	(0.202)	(0.222)

math9_testsc		1.006	1.006	1.006
		(0.00523)	(0.00549)	(0.00521)
AP_1		4.023***	3.994***	4.088***
		(0.756)	(0.708)	(0.737)
AP_3		3.617*	3.674*	3.766*
		(2.712)	(2.772)	(2.739)
absent9_10		0.507*	0.504*	0.495**
		(0.188)	(0.181)	(0.169)
absent9_15		0.583	0.586	0.573
		(0.195)	(0.198)	(0.195)
suspend9_ever		0.544*	0.539*	0.552*
		(0.187)	(0.182)	(0.187)
suspend9_3		1.013	1.021	1.014
		(0.393)	(0.382)	(0.372)
hs_mathtest			0.272	0.225
			(0.298)	(0.256)
hs_frl			0.122	0.109
			(0.249)	(0.227)
0b.frl_rec#1.white				2.626
				(3.684)
Constant	41.31***	1.375	11.57	6.153
	(8.070)	(3.026)	(25.68)	(9.739)
Observations	2,877	1,784	1,784	1,743
Pseudo R^2	0.212	0.385	0.386	0.386

Note: Odds ratios are reported for each model with robust SE in parentheses. The FRL x Latino and FRL x Other Race interaction terms were omitted due to multicollinearity issues.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Model 2 explored the impact of adding ninth grade academic and nonacademic variables to the demographic variables. The odds ratios presented for AP course enrollment demonstrate how much enrolling in an additional two or more AP courses changes a student's odds of graduation, above and beyond the previous level. Students who enrolled in at least one AP course during their high school tenure had four times greater odds of graduating from high school in four years than students who did not enroll in AP courses. Students who were enrolled in three or more AP courses were an additional 3.6 times greater odds of graduating than students who were enrolled in fewer than three AP courses. Students who missed ten or more days of school during the ninth grade and students who were suspended from school at any point during then ninth grade had half the odds of graduate as students who missed fewer than ten days and students who were never suspended. The pseudo R-squared value increased from .212 to .385 as a result of the addition of the academic and nonacademic variables to the model, representing improved model fit. The pseudo R-squared value represents an approximation of the amount of variance explained in analyses that include a binary outcome variable (Raudenbush & Bryk, 2002), such as the analyses included in this study.

Model 3 explored the impact of adding school-level variables to the demographic, academic, and nonacademic variables. Neither of the two variables included were found to be significant predictors, and their inclusion in the model did not substantially alter the pseudo R-squared value (.385 to .386). Model 4 added interaction terms to the model. No significant interaction effects were found, and the inclusion of the interaction effects added no additional explanatory power to the model.

Similar to the eighth grade models, the greatest improvement in the amount of variance explained occurred when academic and nonacademic ninth grade variables were added to the

model. A classification table was produced for the full ninth grade model; this is presented in Table 26.

Table 26.

Initial Classification Table for Series One Ninth Grade Models

	Observed Graduates *N* (%)	Observed Non-Graduates *N* (%)	Total
Classified Graduates	1537 (97.5)	119 (57.2)	1656
Classified Non-Graduates	39 (2.5)	89 (42.8)	128
Overall	1576	208	1784

The initial classification table had an overall classification rate of 91.14%. The model correctly classified students who graduated from high school over 97% of the time. However, those who did not graduate from high school on time were only classified correctly 42.79% of the time. To craft a more useful classification table, sensitivity and specificity were graphed to arrive at an appropriate cutoff point. See Figure 2 for the Series One ninth grade model sensitivity graph.

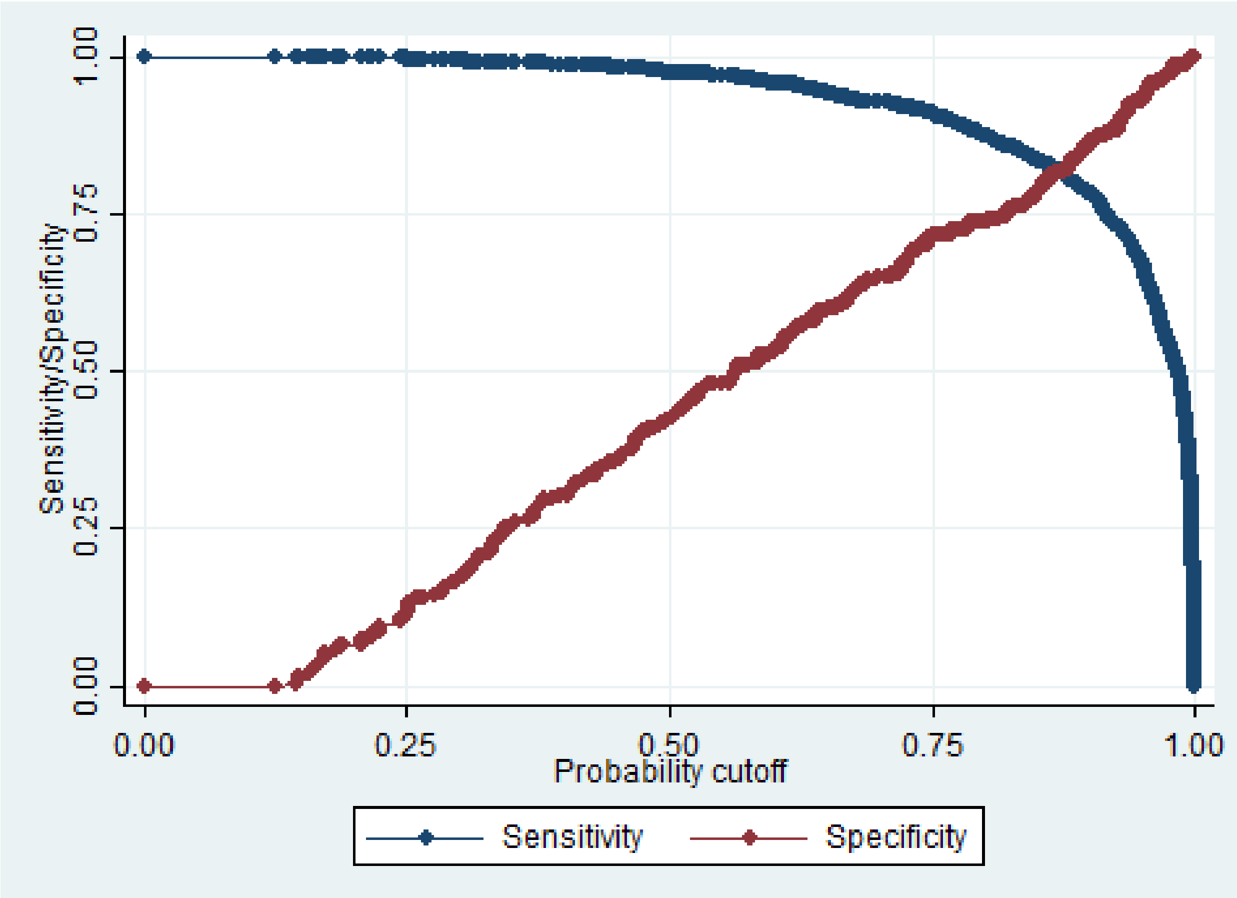

Figure 2. Probability Cutoff for Series One Ninth Grade Models

A new classification table was produced, specifying the cutoff point at 0.8, since that is the approximate point at which sensitivity and specificity meet. The overall classification rate is reduced to 85.76%, however specificity is substantially improved to 73.56%. The adjusted classification table better identifies potential non-graduates and has greater utility for school districts interested in using the findings from predictive models to craft interventions aimed at improving graduation outcomes. The adjusted classification table is presented in Table 27.

Table 27.

Adjusted Classification Table for Series One Ninth Grade Models

	Observed Graduates *N* (%)	Observed Non-Graduates *N* (%)	Total
Classified Graduates	1377 (87.4)	55 (26.4)	1656
Classified Non-Graduates	199 (12.6)	153 (73.6)	128
Overall	1576	208	1784

Series two models. For the second series of models created and tested, two separate analyses were conducted. First, a logistic regression analysis was conducted, similar to the analyses conducted for the Series One models. Second, multilevel logistic regression analysis was conducted, nesting students (Level 1) in home address zip codes (Level 2).

Similar to the Series One models, the logistic regression analysis produced four models. Model 1 examined the impact of demographic variables on high school graduation status. Again, the reference category used is of an African American female student. The findings are identical to those from Model 1 in the ninth grade models. Students receiving special education services had greater odds of graduating from high school than students not receiving special education services. Students who were 15 years old or older by September 30 in the ninth grade had 11% of the odds of graduating as those who were on grade level per their age. Table 26 summarizes the Series Two models estimated, as well as the proportion of the variation in graduation status explained by each model.

Table 28.

Logistic Regression Findings for the Series Two Models

Variables	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 3
White	1.311	0.900	0.865	0.862
	(0.243)	(0.192)	(0.223)	(0.391)
Latino	0.886	0.693	0.747	0.486**
	(0.205)	(0.193)	(0.254)	(0.154)
other_race	1.756**	0.889	0.842	1.137
	(0.474)	(0.779)	(0.793)	(0.971)
Male	0.677***	0.736***	0.748***	0.751***
	(0.0469)	(0.0454)	(0.0536)	(0.0547)
Sped	2.018***	4.411***	4.634***	4.723***
	(0.370)	(1.007)	(1.007)	(0.991)
overage_9	0.115***	0.320***	0.318***	0.317***
	(0.0111)	(0.0208)	(0.0176)	(0.0181)
Hsmob	0.578***	0.814	0.796	0.790
	(0.0980)	(0.204)	(0.202)	(0.206)
Frl	0.373***	0.534*	0.516*	0.582
	(0.102)	(0.177)	(0.181)	(0.227)
eng9course_pass		2.372***	2.379***	2.347***
		(0.652)	(0.698)	(0.680)
math9course_pass		1.631**	1.662***	1.643**
		(0.342)	(0.316)	(0.361)

math9_testsc		1.006	1.008*	1.008*
		(0.00406)	(0.00444)	(0.00409)
AP_1		4.023***	3.905***	3.943***
		(0.901)	(0.725)	(0.756)
AP_3		3.617**	3.550**	3.711**
		(2.192)	(2.119)	(2.136)
absent9_10		0.507**	0.502**	0.496**
		(0.147)	(0.145)	(0.146)
absent9_15		0.583	0.619	0.611
		(0.232)	(0.242)	(0.241)
suspend9_ever		0.544**	0.502***	0.511***
		(0.133)	(0.110)	(0.111)
suspend9_3		1.013	1.099	1.083
		(0.248)	(0.258)	(0.249)
zip_med_income			1.000***	1.000***
			(1.33e-05)	(1.32e-05)
zip_twoparent_perc			1.061***	1.061***
			(0.0174)	(0.0170)
zip_hsgradrate			0.949	0.948
			(0.0328)	(0.0322)
0b.frl_rec#1.white				1.120
				(1.024)
0b.frl_rec#1.other_race				0.431
				(0.393)
Constant	41.31***	1.375	34.41*	40.57**
	(8.070)	(2.153)	(64.15)	(73.07)

Observations	2,877	1,784	1,784	1,743
Pseudo R^2	0.212	0.385	0.399	0.399

Note: Odds ratios are reported for each model with robust SE in parentheses. The FRL x Latino interaction term was omitted due to multicollinearity issues.

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

Model 2 explored the impact of adding ninth grade academic and nonacademic variables to the demographic variables. Findings here are similar to the ninth grade Model 2 findings. Students who avoided earning failing grades in either semester of their ninth grade English course (OR=2.372) and students who avoided earning failing grades in either semester of their ninth grade math course (OR=1.631) substantially increased the odds of graduating from high school in four years when compared to peers who earned failing semester grades in either subject during the ninth grade year. The pseudo R-squared value increased from .212 to .385 as a result of the addition of the academic and nonacademic variables to the model.

Model 3 explored the impact of adding zip code-level variables to the demographic, academic, and nonacademic variables. Two of the three zip code-level variables included in the model were found to be statistically significant predictors of high school graduation. Zip code median income (1.000) and the percent of families in a zip code with two parents (OR=1.061) were both found to be significant predictors. However, median income did not have any practical significance. The proportion of the variance explained increased from .385 to .399 as a result of adding the zip code-level predictors to the model. Model 4 added interaction terms to the model. Similar to the Series One models interaction variables were added, hypothesizing an

interaction between race and free or reduced lunch eligibility. None of the interaction terms were found to be significant, and the pseudo R-squared remained the same at .399.

Similar to the Series One models, the greatest explanatory power in the Series Two models in terms of an improvement in the amount of variance explained came from the introduction of ninth grade academic and nonacademic predictors. A classification table was produced for the full model; this is presented in Table 29.

Table 29.

Initial Classification Table for Series Two Models

	Observed Graduates *N* (%)	Observed Non-Graduates *N* (%)	Total
Classified Graduates	1542 (97.8)	111 (53.4)	1653
Classified Non-Graduates	34 (2.2)	97 (46.6)	131
Overall	1576	208	1784

The initial classification table had an overall classification rate of 91.87%. The model correctly classified students who graduated from high school almost 98% of the time. However, the model failed to correctly classify students who did not graduation from high school more than half of the time. To craft a more useful classification table, sensitivity and specificity were graphed to arrive at an appropriate probability cutoff point. See Figure 3 for the Series Two model sensitivity graph.

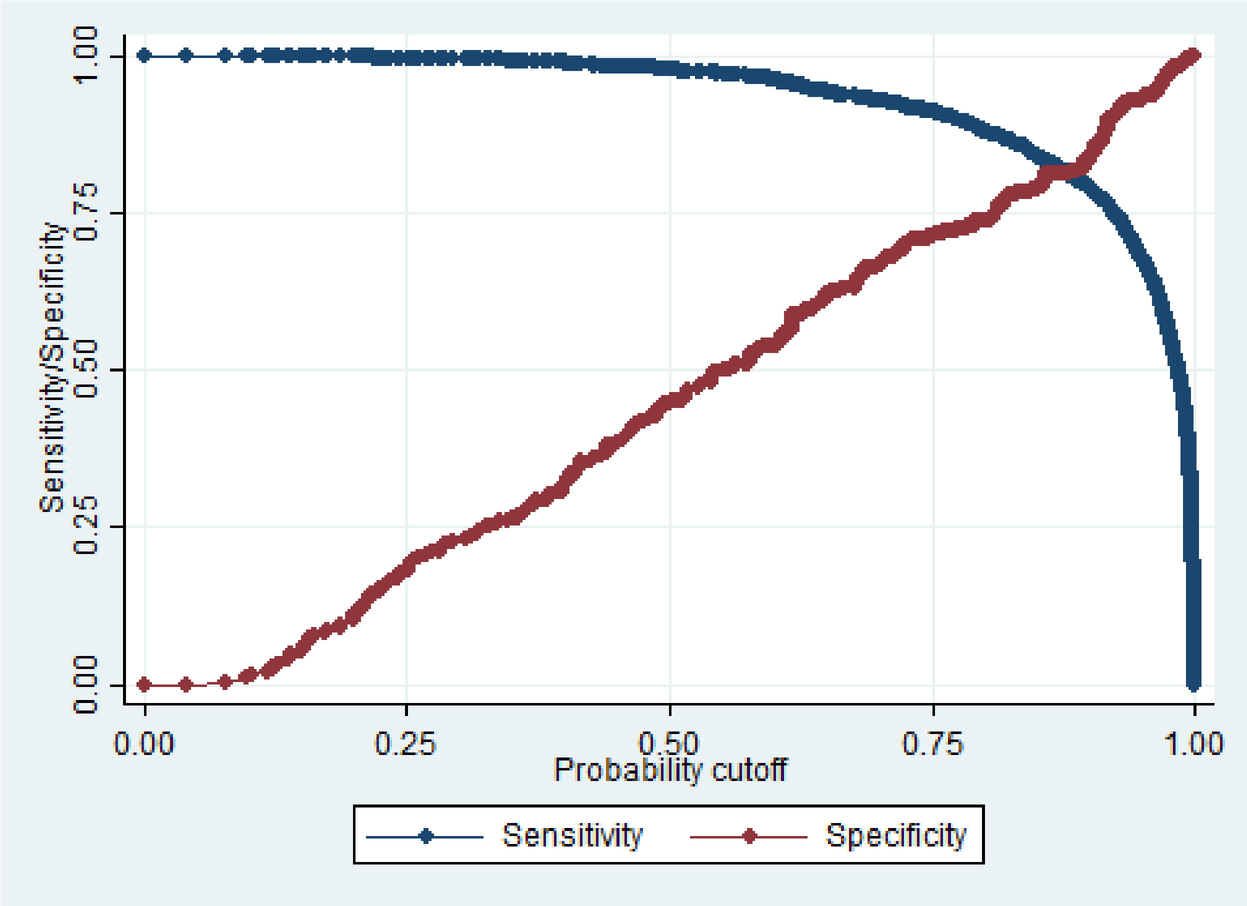

Figure 3. Probability Cutoff for Series Two Models

A new classification table was produced, specifying the cutoff point at 0.8, since that is the approximate point at which sensitivity and specificity meet. The overall classification rate is reduced by approximately six points to 85.99% in the adjusted classification table. However, specificity is substantially improved to 73.56%. The adjusted classification table is presented in Table 30.

Table 30.

Adjusted Classification Tables for Series Two Models

	Observed Graduates *N* (%)	Observed Non-Graduates *N* (%)	Total
Classified Graduates	1381 (87.6)	55 (26.4)	1656
Classified Non-Graduates	195 (12.4)	153 (73.6)	128
Overall	1576	208	1784

The Series Two models were also analyzed using multilevel logistic regression analysis, nesting students (Level 1) in home address zip codes (Level 2). To ascertain whether or not multilevel modeling was appropriate for use with this data, the null model was specified and an intraclass correlation coefficient (ICC) was produced. The ICC for the null model where the outcome variable was whether a student earned a diploma in four years and zip code is the Level 2 grouping variable was .402, indicating that approximately 40% of the variance in student high school graduation status could be explained at the zip code level. Therefore, multilevel modeling is appropriate for use with this data. Table 29 summarizes the Series Two models estimated, and includes AIC values as a measure of incremental model improvement.

Table 31.

Multilevel Modeling Findings for Series Two Models

Variables	(1) Model 1	(2) Model 2	(3) Model 3	(4) Model 4
White	1.399*	0.905	0.865	0.862
	(0.260)	(0.295)	(0.281)	(0.328)
Latino	0.931	0.747	0.747	0.486
	(0.224)	(0.309)	(0.313)	(0.225)
other_race	1.834**	0.872	0.842	1.137
	(0.480)	(0.399)	(0.386)	(0.636)
Male	0.661***	0.735	0.748	0.751
	(0.0763)	(0.138)	(0.141)	(0.142)
Sped	2.094***	4.443***	4.634***	4.723***
	(0.371)	(1.455)	(1.523)	(1.558)
overage_9	0.111***	0.322***	0.318***	0.317***
	(0.0142)	(0.0693)	(0.0684)	(0.0687)
hsmob	0.569***	0.796	0.796	0.790
	(0.0725)	(0.168)	(0.168)	(0.168)
frl	0.362***	0.520**	0.516**	0.582
	(0.0620)	(0.149)	(0.149)	(0.208)
eng9course_pass		2.371***	2.379***	2.347***
		(0.476)	(0.479)	(0.475)
math9course_pass		1.659**	1.662**	1.643**
		(0.359)	(0.360)	(0.359)
math9_testsc		1.007**	1.008***	1.008**
		(0.00300)	(0.00299)	(0.00301)
AP_1		3.797***	3.905***	3.943***
		(1.681)	(1.728)	(1.749)
AP_3		3.556	3.550	3.711
		(2.964)	(2.962)	(3.102)
absent9_10		0.504***	0.502***	0.496***
		(0.127)	(0.127)	(0.126)
absent9_15		0.611*	0.619*	0.611*
		(0.156)	(0.159)	(0.157)
suspend9_ever		0.519**	0.502**	0.511**
		(0.161)	(0.155)	(0.159)
suspend9_3		1.071	1.099	1.083
		(0.368)	(0.377)	(0.372)
zip_med_income			1.000***	1.000***
			(1.98e-05)	(1.98e-05)
zip_twoparent_perc			1.061***	1.061***
			(0.0175)	(0.0176)

zip_hsgradrate			0.949	0.948
			(0.0331)	(0.0333)
0b.frl_rec#1.white				1.120
				(0.798)
0b.frl_rec#1.other_race				0.431
				(0.420)
Constant	34.39***	0.731	34.41	40.57
	(13.30)	(0.897)	(90.67)	(107.5)
Observations	2,877	1,784	1,784	1,743
Number of groups	6	6	6	6

Note: Odds ratios are reported for each with robust SE in parentheses. FRL x Latino interaction is omitted due to multicollinearity issues.

*** p<0.01, ** p<0.05, * p<0.1

In terms of model fit, the AIC values reduce for each subsequent model. Model 1 is an improvement on the null model (3408.07 to 2023.52); Model 2 is an improvement on Model 1 (2023.52 to 819.66); Model 3 is an improvement on Model 2 (819.66 to 814.62). A test of the full model against null model was statistically significant (Wald Chi-Squared = 239.96; $p < .001$).

Two demographic variables were significant predictors in the full model. Students receiving special education services had 4.634 times greater odds of graduating from high school in four years than those who did not receive special education services. Also, similar to findings in the other models in this study, students entering the ninth grade 15 years old or older experienced approximately 30% odds of graduating high school compared to those who entered the ninth grade 14 years old or younger. In terms of academic predictors, the two largest odds ratios were associated with grades in ninth grade English and enrollment in AP courses. Students who passed both semesters of freshman English had twice the odds of graduating when compared to someone who failed one or both semesters (OR=2.379). Likewise, students who enrolled in at least one AP course had almost four times the odds of graduating high school as a

peer who did not enroll in AP coursework (OR=3.905). In this model, enrollment in three or more AP courses was a non-significant predictor. Of the two significant zip code-level predictors, only the percent of two parent families has practical significance. For every one-point increase in the percent of families in a zip code with two parents living at home, a student residing in that zip code has 1.06 times greater odds of graduating from high school. For example, if the percentage of two parent families rose eight points, this finding would suggest that a student in that zip code would have almost 50% greater odds of earning a high school diploma in four years.

Graduation Status Profiles

The findings discussed in this chapter offer insights for evaluating the models that have been tested and created. They also highlight the impact that individual variables have on the relative odds of a student graduating from high school in four years. Two profiles have been constructed to further illustrate the impact that these findings have on a student's propensity to graduate from high school. Profile A is a fictitious African American female. In the first scenario (A.1), Profile A passes her eighth grade math course each marking period, is absent fewer than 10 days during the eighth grade, and was never suspended during the eighth grade. Table 32 represents a partial table for model estimates for the Series One eighth grade full model. Appendices A through C present the complete model estimate tables for each of the three models created and tested in this study.

Table 32.

Model estimates for Series One Eighth Grade Full Model (Abbreviated)

Variable	Coefficient	Robust SE	*p*-value
Intercept	3.1320	5.3290	.565
white	- .3566	.1433	.013
latino	- .7765	.3642	.033
other_race	.2258	.1924	.241
male	- .2820	.0769	.000
math8course_pass	.7420	.1996	.000
absent8_10	- .4799	.1365	.000
suspend8_ever	- .6949	.1362	.000
suspend8_3	- .2906	.1080	.007

Note: Complete table can be found in Appendix A.

Since African American females are captured in the reference category for the Series One eighth grade model, the intercept represents Profile A (3.1320). In A.1 Profile A earned all passing grades in math, .7420 should be added to the intercept to account for that. That produces a logit of 3.874. Predictive probabilities can be calculated using logit values (Raudenbush & Bryk, 2002). A logit of 3.874 would represent a predictive probability of 97.96%. For scenario A.1, Profile A would have an almost 98% probability of graduating from high school. Scenario A.2 is similar to A.1, except in this scenario Profile A did not earn a passing grade in each of marking period in eighth grade math. Since not earning a passing grade in each marking period of eighth grade math is also captured in the reference category, the intercept would be the logit of interest for A.2, yielding a predictive probability of 95.82%. Profile A's probability of graduating from high school was only reduced by two percent due to earning one grade of D or F in eighth grade math. Similar to A.2, Profile A also did not earn a passing grade in each marking

period for eighth grade math in scenario A.3. However, in A.3, Profile A was also absent for at least 10 days during the eighth grade and suspended for at least three days during the same year. Again, not earning a passing grade in each marking period in eighth grade math is already captured in the reference category. However, the excessive absences and days suspended are not. Being absent for 10 or more days in the eighth grade is represented by a logit of -.4799. A logit of -.6949 is associated with having ever been suspended during the eighth grade and a logit of -.2906 is associated with having been suspended for three or more days during the same year. Subtracting these three values from the reference category produces a logit of 1.6656 for Profile A in scenario A.3, yielding a predictive probability of 84.10%. Profile A's probability of graduating high school was reduced by more than 11 percentage points from A.2 to A.3 as a result of missing more than 10 days of school and being suspended for three or more days in the eighth grade. Figure 4 depicts the predictive probabilities for the three scenarios for Profile A in graphical form.

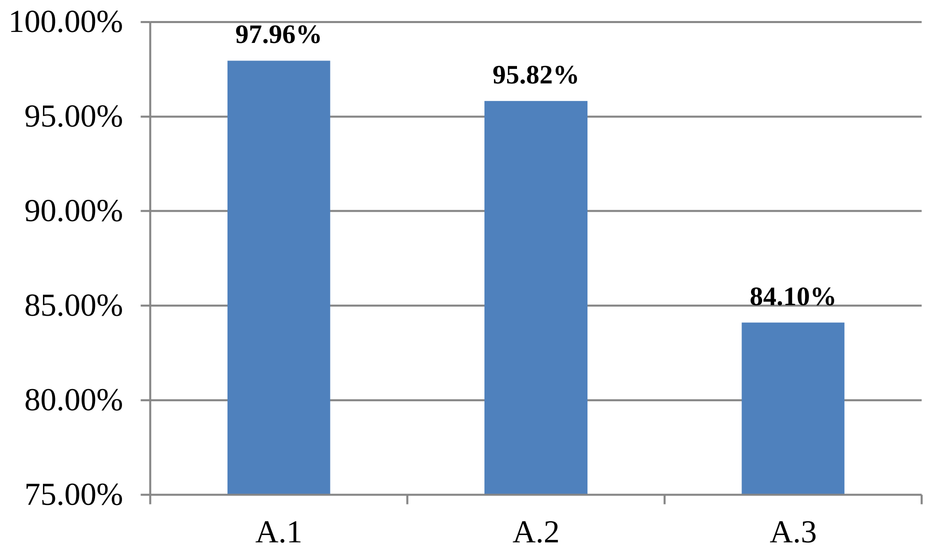

Figure 4. Predictive probabilities for three Profile A scenarios

Note: Profile A portrays an African American female student. A.1 represents a scenario where Profile A earned passing grades each marking period in eighth grade math, was absent fewer than 10 days during the eighth grade, was never suspended during the eighth grade. A.2 represents a scenario similar to A.1, except one where Profile A did not earn passing grades each marking period in eighth grade math. A.3 represents a scenario where Profile A did not earn passing grades in eighth math, was suspended for three days, and was absent for 10 days during the eighth grade.

Profile B is a fictitious Latino male student in the school district. Table 33 represents a partial table for model estimates for the Series One ninth grade full model.

Table 33.

Model estimates for Series One Ninth Grade Full Model (Abbreviated)

Variable	Coefficient	Robust SE	*p*-value
Intercept	1.8170	1.5827	.251
white	- .1268	.6251	.839
latino	- .7694	.1971	.000
other_race	.1917	.7911	.808
male	- .3003	.1134	.008
eng9course_pass	.8543	.2915	.003
math9course_pass	.4551	.1408	.001
AP_1	1.4080	.1802	.000
absent9_10	- .7035	.3425	.040

Note: Complete table can be found in Appendix B.

In the first scenario (B.1), Profile B passes all of his courses, is absent fewer than ten days in the ninth grade, and was never suspended during the same year. The intercept in this model is 1.8170. However, being Latino is associated with a logit of -.7694 and being male is associated with a logit of -.3003. For this student, the starting logit is .7473. Passing each semester of English (.8543) and math (.4551) in the ninth grade increases the logit for scenario B.1 to 2.0567, yielding a predictive probability of 88.66%. Profile B has a probability greater than 88% of graduating from high school in this scenario. Scenario B.2 is similar to B.1, except in this scenario Profile B was enrolled in at least one AP course, which is associated with a logit of 1.4080. The logit associated with scenario B.2 is 3.4647, yielding a predictive probability of 96.97%. Profile B increases his probability of graduating from high school by almost nine percentage points as a result of enrolling in one or more AP courses. For scenario B.3, Profile B fails at least one semester of ninth grade English and math. Not passing both semesters of ninth grade English and math is captured in the reference category. In B.3, Profile B also misses 10 or

more days of school, associated with a logit of -.7035. The total logit for B.3 is .0438, yielding a predictive probability of 51.09%. A student who does not pass both semesters of ninth grade English and math, and misses more than 10 days of school in the ninth grade has a predictive probability of only 51% of graduating from high school in four years. Figure 5 depicts the predictive probabilities for the three scenarios for Profile B in graphical form.

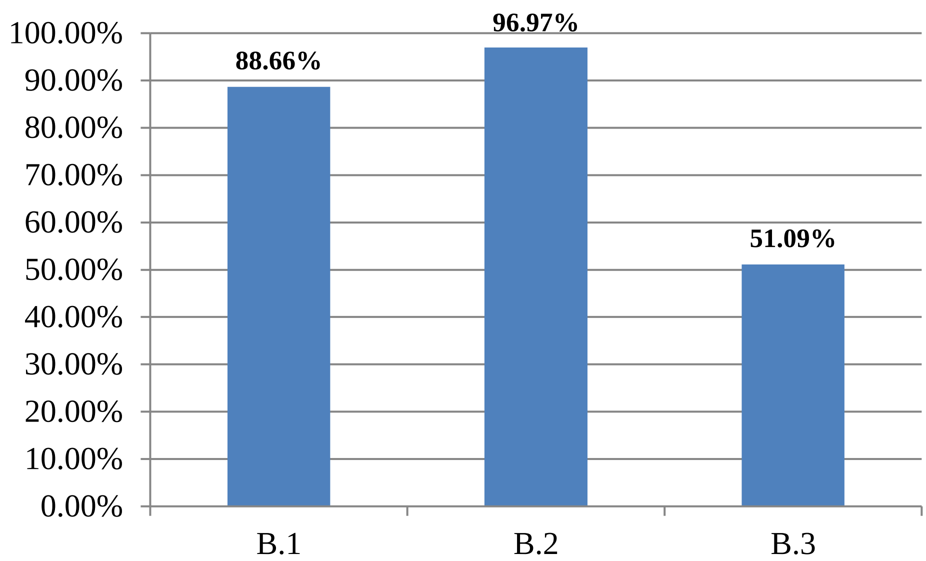

Figure 5. Predictive probabilities for three Profile B scenarios

Note: Profile B portrays a Latino male student. B.1 represents a scenario where Profile B earned passing grades each semester in ninth grade English and math, was absent fewer than 10 days during the ninth grade, and was never suspended during the ninth grade. B.2 represents a scenario similar to B.1, except one where Profile B was enrolled in at least one AP course in high school. B.3 represents a scenario where Profile B failed at least one semester each of ninth grade English and math, and was absent for 10 or more days.

Summary

Two series of models were constructed and tested using logistic regression analysis and multilevel logistic regression analysis. Overall, each of the three logistic regression models

yielded pseudo R-squared values exceeding .30; the Series Two full model produced a pseudo R-squared value of almost .40, which outperforms most models in the literature on high school graduation status predictive models. When specified with a cutoff point of 0.8, classification tables were able to correctly classify students who did not earn a high school diploma in four years between 70% and 80% of the time, which is in line with previous work in the literature. Finally, multilevel modeling was also conducted on the Series Two models, and zip code was found to account for approximately 40% of the variance explained by the models. The addition of demographic, academic and nonacademic, and zip code-level variables led each iteration of the model to improve over the previous iteration.

CHAPTER V – DISCUSSION

This chapter includes a discussion of the findings from the two series of models that were created and tested. Included in this chapter is an overall summary of the results that includes considering the findings within the broader context of the literature on high school graduation status prediction studies, as well as an examination of specific variables and their predictive ability. Two overall patterns emerged from the findings and are explored in greater detail. Finally, implications for future research, practical implications for school districts, and limitations are discussed.

Summary of Findings

The models created and tested in this study were evaluated in two ways. First, the models were evaluated in terms of the amount of the variance that was explained by the predictor variables they included. Of the studies included in the review of the literature for this study, six studies reported pseudo R-squared values and an additional study reported -2 Log Likelihood values, from which a pseudo R-squared value could be calculated (Glynn, 2012). Table 34 presents pseudo R-squared values for each of these studies, nesting the three models tested from the current study within this list. Variables that these studies included different from the current study are identified. Approaches to operationalizing variables that differed are also identified (i.e. Mac Iver & Messel, 2013 operationalized attendance as missing 20 or more days; the current

study operationalized it as missing 10 or 15 days.) Mac Iver and Messel (2013) tested the two cohorts included in their study separately and reported separate pseudo R-squared values for each.

Table 34.

Pseudo R-squared Values for Studies Predicting High School Graduation Status

Study	Differences in Approach to Current Study	Pseudo R^2
Mac Iver & Messel, 2013 2005-2006	Included only 7 predictor variables. Attendance was calculated as having missed 20+ days; current study included variables for 10+/15+ days	.460
Mac Iver & Messel, 2013 2004-2005	Included only 7 predictor variables. Attendance was calculated as having missed 20+ days; current study included variables for 10+/15+ days	.450
Current Study Series 2 (Zip Code Predictors)	**-**	**.399**
Current Study Series 1b (Ninth Grade Predictors)	**-**	**.386**
Cratty, 2012	Included type of school attended, parent education; measured over age for grade in the 3rd grade; included measures of growth on test scores from 3rd to 8th grades	.352
Soland, 2013	Measured courses failed in terms of whether a student had ever failed a course in a given subject; Included teacher prediction variables	.345
Alexander et al., 1997	Included variables for family context, self-image, locus of control, school	.310

	satisfaction	
Current Study Series 1a (Eighth Grade Predictors)	**-**	**.307**
Oh & Reynolds, 2008	Included preschool participation; parent education, birth weight, parental marital status; Included a school-level factor of percent of students above reading level for grade	.275
Neild et al., 2008	Included parent education, parental marital status, student engagement with teachers, student risk-taking behaviors	.270
Tobin & Sugai, 1999	Included two sixth grade predictors – GPA & violent behaviors	.136*

*Note: * Pseudo R^2 not reported; value was calculated based on reported -2 LL values with and without covariates.*

Two of the models created and tested in the current study outperformed all of the models explored in the review of the literature for this study in terms of the amount of variance explained, with Mac Iver and Messel's (2013) Early Warning Indicator model with ninth grade variables being the single exception (pseudo R^2 = .460 & .450). The Series 2 full model yielded a pseudo R-squared value of .399 and the Series 1 ninth grade full model produced a pseudo R-squared vale of .386. The pseudo R-squared value for the Series 1 eighth grade full model (.307) is not as impressive as the other two models in this comparison. However, the Cratty (2012) models and the Mac Iver and Messel (2013) models also tested the predictive ability of eighth grade variables in incomplete iterations of their models. Table 35 compares the four eighth grade models in terms of the amount of variance explained.

Table 35.

Pseudo R-squared Values for Studies Using Eighth Grade Predictors

Study	Pseudo R^2
Mac Iver & Messel, 2013 2005-2006	.340
Current Study Series 1a - (Eighth Grade Predictors)	**.307**
Mac Iver & Messel, 2013 2004-2005	.290
Cratty, 2012	.285
Neild et al., 2008	.270

When compared to other models that tested the predictive ability of eighth grade variables, the Series 1 eighth grade model tested in the current study yielded a pseudo R-squared value outperforming the other models, with the lone exception being one of the Mac Iver and Messel models.

Models were also evaluated in terms of their accuracy in classifying cases. Overall classification rates and non-graduate classification rates were considered in this evaluation, privileging non-graduate classification rates of the two. Although superior overall classification might be interesting from a research perspective, the ability to accurately classify students at risk of failing to complete high school is more is more important in terms of practical implications. Four of the studies included in the review of the literature presented classification tables for their models. Lloyd (1978), McWilliams et al. (2000), and Jimerson (2000) each reported

classification tables for two models in their studies; Owens, Morris, and Lieberman (2001) created and tested 11 models. Table 36 presents the overall classification rates for each of these models. For the Owens et al. study included in this table, only the first model is presented. It yielded the highest overall and highest non-graduate classification rates of the 11 that were tested.

Table 36.

Overall Classification Accuracy for Studies Predicting High School Graduation Status

Study	Model	Overall Classification
McWilliams et al., 2000	Females	90.5%
McWilliams et al., 2000	Males	88.9
Current Study	**Series 2**	**86.0**
Current Study	**Series 1 9th Grade**	**85.8**
Current Study	**Series 1 8th Grade**	**81.0**
Owens et al., 2001	Model 1*	78.4
Lloyd, 1978	Females	77.1
Lloyd, 1978	Males	76.4
Jimerson, 2000	Model 1**	75.0
Jimerson, 2000	Model 2***	75.0

*Note: * Best model of 11 that were tested in terms of overall and dropout classification; ** Model included first and sixth grade predictors; *** Model included first and sixth grade predictors, as well as 16-year-old predictors*

The two models tested in the McWilliams, Everett, and Bass (2000) study both outperformed the models in the current study in terms of the percentage of cases correctly classified overall. This study also relied on data obtained from a single public school district, but from a rural rather than an urban setting. The current models outperformed the other models included in the review of the literature for this study in terms of overall classification rate.

The three models tested in this study correctly classified students as non-graduates between 71% and 74% of the time, or approximately 13 points below the non-graduate classification rate of the McWilliams et al. study. However, the models in this study outperformed models from two other studies as well. Table 37 compares the non-graduation classification rates for the models from this study with those of the other four studies examined in the literature.

Table 37.

Non-Graduate Classification Accuracy for Studies Predicting High School Graduation Status

Study	Model	Non-Graduate Classification
McWilliams et al., 2000	Females	86.7%
McWilliams et al., 2000	Males	86.7
Owens et al., 2001	Model 1*	83.7
Jimerson, 2000	Model 2***	82.0
Current Study	**Series 2**	**73.6**
Current Study	**Series 1 9th Grade**	**73.6**
Current Study	**Series 1 8th Grade**	**71.7**
Lloyd, 1978	Males	69.4
Jimerson, 2000	Model 1**	67.0
Lloyd, 1978	Females	65.7

*Note: * Best model of 11 that were tested in terms of overall and dropout classification; ** Model included first and sixth grade predictors; *** Model included first and sixth grade predictors, as well as 16-year-old predictors.*

Individual Predictors of Graduation Status

The two primary research questions guiding the design and implementation of this study examined the extent to which academic and nonacademic variables were predictive of high

school graduation status. Below, these findings are explored in the context of this study, as well as in the larger context of the literature on high school graduation status.

Academic predictors. Variables representing four of the five categories of academic predictors discussed in the review of the literature in Chapter 2 were included in this study's models. Earning passing grades in each marking period or semester of eighth and ninth grade English and math was a significant predictor of increased odds of graduating from high school; this finding held true across all models tested. The same was true for grade retention; across all models tested, students who were over age for grade faced reduced odds of graduating from high school in four years. These findings are in line with the literature on these predictors; no matter how either variable was operationalized in the studies reviewed in Chapter 2, course passage (or failure) and grade retention were always found to be significant predictors of high school graduation status (Allensworth & Easton, 2005; Mac Iver & Messel, 2013; Saunders, Silver, & Zarate, 2008; Silver, Saunders, & Zarate, 2008; Soland, 2013).

The literature on the predictive ability of standardized test scores is mixed, and the findings in this study are as well (Barrington & Hendricks, 1989; Hernandez, 2011; Mac Iver & Messel, 2013; Neild, Stoner-Eby, & Furstenberg, 2008; Oh & Reynolds, 2008; Parr & Bonitz, 2015). In the Series One eighth grade model, a student's eighth grade English standardized test score was a significant predictor of increased odds of graduating from high school. However, eighth grade math test scores were not significant in any model. Similarly, ninth grade math standardized test scores were significant only in some models. When school-level variables were included with ninth grade predictors, ninth grade math standardized test scores were not found to be significant predictors of high school graduation status. However, when zip code-level variables were included, this was significant. When community-level predictors were

considered, students who earned higher scores on their ninth grade math tests were more likely to graduate from high school than peers who earned lower scores. It is possible that this finding is true just for this sample. It is also possible that when school-level test score predictors are not included in models, test scores have more predictive ability than when school-level test score predictors are included. Test scores were not significant in any model when included as school-level predictors.

The fourth type of academic variable that was included in the model was course enrollment. The studies reviewed in the literature found that enrollment in Algebra I by the eighth grade (Cratty, 2012) and enrollment in AP courses (Goldschmidt & Wang, 1999; Soland, 2013) were both significant predictors of having increased odds of graduating from high school. In the present study, enrollment in AP courses was found to be a significant predictor of increased odds of graduation, and enrollment in three or more AP courses was a significant predictor of increased odds of graduation beyond enrollment in one or two AP courses. However, none of the Series One eighth grade models tested found enrollment in Algebra I by the eighth grade to be a significant predictor. This finding stands in contrast to Cratty's finding that students enrolled in Algebra I by the eighth grade had five times greater odds of graduating than peers who enrolled in Algebra I in later grades.

Nonacademic predictors. Variables representing two of the three categories of nonacademic variables related to student engagement discussed in the review of the literature were included in the models for this study. The literature reviewed on student attendance (Balfanz, Herzog, & Mac Iver, 2007; Mac Iver & Messel, 2013; Rumberger, 2011) and student behavior (Cratty, 2012; Goldschmidt & Wang, 1999; Mac Iver & Messel, 2013; Neild et al., 2008) indicates that both variables, however operationalized, were always significant predictors

of high school graduation status. Overall, the findings in the present study are in line with these trends. Students who were suspended at any time during the eighth or ninth grade faced reduced odds of graduating from high school; this finding held true across all models tested. Similarly, missing 10 or more days of school in the eighth and ninth grade was a significant predictor of reduced odds of graduating from high school in four years. Students who missed fewer than two weeks of instruction and were not suspended from school in the eighth and ninth grades had greater odds of graduating from high school in four years than peers who missed more days and were suspended from school.

Variables representing five of the six categories of nonacademic demographic variables discussed in the review of the literature were included in the models for this study. The models tested in previous studies produced mixed findings regarding the ability of race-related variables to predict high school graduation status. The reference category for race in each of the models tested in this study was African American. This differs from most of the literature that includes more than one categorical variable representing race; in all but one of the studies reviewed, White students represented the reference category. For the Series One ninth grade and Series Two models, no relationship was found between race and graduation status. However, for the Series One eighth grade models, being White or Latino/a was found to be a significant predictor of reduced odds of graduation. African American students had greater odds of graduating from high school in four years than their White or Latino/a peers, echoing Cratty's (2012) findings from her study of a statewide cohort in North Carolina. Half of the studies reviewed in the literature found that male students faced reduced odds of graduating from high school in four years; this finding is in line with the findings from this study. Across all models tested, male students had lower odds of graduating from high school than their female peers.

Socioeconomic status (Alexander, Entwisle, & Horsey, 1997; Battin-Pearson et al., 2000; Hernandez, 2011; Jimerson et al., 2000; Rumberger & Palardy, 2005) and student mobility (Rumberger & Larson, 1998) were found to be significant predictors of high school graduation status no matter how they were operationalized when included in the predictive models included in the review of the literature for this study. In the current study, socioeconomic status was operationalized in terms of eligibility for free or reduced price lunch, and this was found to be a significant predictor of high school graduation status in almost all of the models tested. Students who were eligible for free or reduced price lunch faced reduced odds of graduating from high school in four years when compared to peers who were not eligible. The only models tested where this finding did not hold true were those that included race by socioeconomic status interaction terms.

The findings regarding student mobility were more nuanced in this study than they were in the literature that was reviewed. Students who attended more than one middle school between grades six and eight faced reduced odds of graduating from high school in every model tested. Middle school mobility was a consistently significant predictor of high school graduation status. However, the same was not true for high school mobility. In this study, this variable was operationalized as indicating that a student had changed schools one or more times, excluding promotion from middle to high school. High school mobility was found to be significant as a predictor of high school graduation status in models that only included demographic predictors; the variable yielded non-significant findings in every other model tested. When other academic and nonacademic predictor variables were included in the models, high school mobility was not a significant predictor of graduation status. One possible explanation for this finding is that all students included in the sample had to change schools from eighth to ninth grade, and therefore,

a lack of stability in terms of school setting was not a reality for anyone included in this study. Since everyone changed schools in the year prior, this was not a sensitive predictor variable to include in the model.

Finally, disability status was included as a demographic predictor variable in the models tested in this study. Few of the studies included in this review of the literature included disability status as a predictor variable in their models, and those that did include disability as a predictor yielded a range of findings. In the present study, disability status was operationalized as receiving special education services, with students identified as being gifted excluded. Students receiving special education services had four to five times greater odds of graduating from high school than students who did not receive special education services. This finding is discussed in further detail in the next section of this chapter.

Trends in the Findings

Two overarching trends emerged from the findings based on the models that were tested in this study. First, the implications of using findings from large-scale studies in local contexts are discussed. The two predictor variables with the largest odds ratios are then discussed. Finally, race and high school graduation status are discussed.

Using findings from large-scale studies in local contexts. A tension sometimes exists between generalizability and local specificity. Although most of the findings in this study are in line with the literature on high school graduation status predictors, a few of the findings ran counter to those in the literature reviewed for this study. One example of this would be with the finding in Cratty's (2012) study that students who were enrolled in Algebra I by the eighth grade had five times greater odds of graduating from high school in four years than their peers who

were not enrolled in Algebra I until they were in high school. However, in the present study, enrollment in Algebra I by the eighth grade was a non-significant finding. It is impossible to know exactly why this was the case in the present study, but there is one substantial difference between the two studies. The sample used in Cratty's study was that of a graduation cohort for an entire state. Table 38 compares the sample used in Cratty's study with the sample included in the current study.

Table 38.

Demographics in Cratty (2012) vs. Local Demographics for the Current Study

Demographic	Cratty (2012)	This Study	Delta
White	63.5%	23.1%	- 40.4%
African American	30.5	61.5	+31.0
Latino/a	2.3	6.6	+ 4.3
Other Race*	3.7	8.8	+ 5.1
Poverty Rate**	54.6	69.2	+14.6
Students with Disabilities	15.6	11.1	- 4.5

***** *Operationalized in this study as a student who is not classified as being White, African American, or Latino/a.* ****** *Poverty rate is based on the percent of households who are eligible for free or reduced price lunch.*

The sample in Cratty's study differs from the sample included in the current study substantially in the number of White and African American students included. The sample for the current study has an African American subset that is more than double that of Cratty's sample. At the same time, there is a 40 percentage point gap in the proportion of the sample that is identified as White. The current study also has a sample that is almost 15 percentage points higher in terms of eligibility for free or reduced lunch.

Several other studies included in the review of the literature conducted for this study also used very large sample sizes and used nationally-representative longitudinal datasets for their analyses. Table 39 compares the national population of public school students with the sample included in the current study.

Table 39.

National Demographics vs. Local Demographics for the Current Study

Demographic	National	This Study	Delta
White	52.7%[a]	23.1%	- 29.6%
African American	16.4[a]	61.5	+45.1
Latino/a	22.4[a]	6.6	- 15.8
Other Race*	8.6[a]	8.8	+ 0.2
Poverty Rate**	48.1[b]	69.2	+21.1
Students with Disabilities	12.9[c]	11.1	- 1.8

Note: Figures indicate percent of population who identify with the demographic variable.

***** *Operationalized in this study as a student who is not classified as being White, African American, or Latino/a.* ***** *Poverty rate is based on the percent of households who are eligible for free or reduced price lunch.* [a] *National Center for Education Statistics (2015a)* [b] *National Center for Education Statistics (2015b)* [c] *National Center for Education Statistics (2016b).*

The same three differences exist between the national public school population and the sample used in this study. The lone exception is the difference between the national Latino/a population and the Latino/a population that exists in the present study. The differences that exist between Cratty's study and the present one, or the differences that exist between a nationally-representative sample and the sample included in the current study are worth noting. Context matters, especially in education. Demographic differences do not render the findings from one study irrelevant in a context that is different. However, local policymakers should be aware that

the findings from research conducted with a sample different than that of the local population might not hold true in the local context. That was the case with this study; the findings surrounding Algebra I enrollment for Cratty differed greatly from those arrived at here. Suppose a policymaker read Cratty's work and decided to implement a program in this school district that was designed to increase middle school Algebra I enrollment as a way to increase high school graduation rates. The policymaker might be surprised if the intervention failed to yield the desired results, if he or she had only consulted research that utilized large samples that do not approximate the local school system[1]. This finding provides a caveat for assuming the generalizability of work from one setting to the next.

AP coursework and special education. The largest odds ratios associated with individual predictors included in the models tested in this study were associated with AP course enrollment and disability status. The AP course enrollment finding is consistent with the literature; students enrolled in advanced coursework are more likely than their peers to have excelled in earlier coursework**.** This finding is in line with the literature exploring different academic pathways; students who enroll in more challenging coursework graduate from high school at substantially higher rates than students placed on a less challenging academic track (Oakes & Guiton, 1995; Werblow, Urick, & Duesbury, 2013). An initiative currently exists in the school district from which this sample was drawn that seeks to enroll every student in at least one AP course as a part of their program of study; the findings from this study support the continuation of such an initiative.

[1] The school district from which the sample was drawn implemented an initiative aimed at enrolling students in Algebra I by the eighth grade. However, this initiative was disbanded the year prior to the students included in the sample entering the eighth grade.

The disability status finding is more difficult to interpret. It should be made explicit that it is not recommending that students be unnecessarily enrolled in special education services as a means of increasing graduation rates. However, this finding is deserving of additional discussion. There are at least three possible explanations for this finding. The first possibility is that the special education program in the school district that participated in this study was doing an outstanding job of meeting the needs of the students with disabilities in the school district. The students in this sample with disabilities were getting the services and supports that they needed and they were well-positioned to meet the requirements to earn a high school diploma – whether it was a standard diploma or a modified standard diploma[2]. A second possible explanation is that students were on the modified standard diploma pathway. Students who may have been able to succeed in a standard diploma pathway (see footnote 1) might have been steered towards a track that educators might have deemed more attainable, thus increasing the propensity to graduate in four years. A third possible explanation for the large effect size could be a function of one of the requirements for having an IEP in the first place. In order for a student to be identified as having a disability and obtain an IEP, a parent must give permission for the student to be evaluated. As a result of this, students who have a disability but whose parents do not consent to an evaluation would not be classified as a student with a disability. Therefore, it is possible that a mediating factor that is unaccounted for in this study is parental permission for disability evaluation.

[2] Students in the state from which the sample was drawn are eligible for a modified standard diploma if they have an IEP. Decisions about eligibility for this diploma could be made any time after the eighth grade. Earning a modified standard diploma required fewer credits than a standard diploma. Also, students who earned a modified standard diploma were not required to earn proficient scores on the six standardized tests that were required for a standard diploma. Citation withheld to protect anonymity of the context of this research.

Conclusion

Most of the findings for individual predictors of high school graduation status were in line with the literature on predictors of high school graduation status. However, this was not true for all of the predictor variables that were tested. It is clear that it remains important to consider context when weighing the value of educational research and its application to other settings which may differ in terms of population, policy context, or in other ways. The findings for the models overall indicate that community-level variables are important to consider when modeling high school graduation status. Forty percent of the variance was explained at the zip code level for the Series Two models, and future research on these types of predictive models should continue to consider the settings in which students exist outside of school.

Implications for Future Research

The finding that students with disabilities have four times greater odds than students without disabilities warrants additional study. Findings like these highlight the limits of quantitative research. Much can be learned from quantitative studies; however, qualitative work is often required to develop a deeper understanding of why a relationship exists between variables (Maxwell, 2013; McMillan, 2012). Additional qualitative work should be conducted to develop a better understanding of why students with disabilities were as successful as they were for the graduation cohorts included in this study. This knowledge could help to improve how future models are specified, and help to better explain similar findings from future studies.

Additional qualitative work should also explore the off-diagonals in the classification tables for this study's findings. Of particular interest are the lived experiences of students who graduated from high school, but for whom the models tested here predicted non-graduation

status. Too often the conversation surrounding challenging school contexts is framed in negative terms. This line of inquiry would allow for the exploration of positive outcomes in these contexts. In the models tested in the current study, students who were misclassified were: (1) more likely to be African American; (2) less likely to be White; (3) more likely to be male; (4) more likely to be eligible for free or reduced price lunch, and (5) more likely to have a disability. The tables included in Appendices D through F compare the students in the sample who were correctly classified with those who were misclassified for each model tested. Additional qualitative work might also help to explain the quantitative differences that exist between these two groups.

There were four variables that were initially considered for inclusion in the models tested in this study that were not included in the final models. Two of the variables requested were not available in the administrative data that were collected by the school district – mother's education and participation in extracurricular activities. Mother's education was going to be operationalized as a dichotomous variable – earned a high school diploma or not. Previous studies in the literature suggest that students whose mothers have at least earned a high school diploma are more likely to earn a high school diploma themselves (Battin-Pearson et al., 2000). Also, participation in extracurricular activities was found to be a significant predictor of high school graduation status (McNeal, 1995; Song, Benin, & Glick, 2012). Future research that can capture these variables should include them in predictive models. Significant findings for either predictor could lead to action. Schools could ensure that a range of extracurricular activities exist for students and that they are accessible and provide transportation for participants. Schools with large numbers of students whose parents are not high school graduates themselves

could offer programs for parents to earn their GED, with career counseling services provided as well.

The other two variables that were to be included in the model were English-language learner (ELL) status and military family status. The ELL status variable was collected by the school district and provided for inclusion in this study. However, the ELL population was too small for the sample in this study to warrant inclusion. Future research should continue to include this variable when possible. There is a growing population of individuals in America for whom English is not the first language spoken at home (Zeigler & Camarota, 2014), and there will increasingly be a need for school districts to address this. The school district began collecting data on military family status in 2016, which was too late for the variable to have utility for the present study. However, an opportunity exists to examine the relationship between coming from a military family and high school graduation status. The review of the literature conducted for this study found nothing that examined this relationship. Future work that takes place in school districts located near military installations should consider this as a variable for inclusion in their models.

The addition of the race by socioeconomic status interaction effects created better specified models in the current study. Future predictive studies should also consider adding race by gender by suspension and race by suspension by special education interaction terms as well. A report published by the Legal Aid Justice Center indicates that in more than one-third of Virginia's school districts, between 25% and 40% of African American male students with disabilities were suspended during the 2014-2015 school year (Langberg & Ciolfi, 2016). These numbers are disquieting, and the inclusion of the interaction terms will produce better specified models in future work.

Future work should consider examining the ability of individual-level academic and nonacademic variables, school-level variables, and community-level variables to predict postsecondary educational outcomes. The predictive studies carried out by Mac Iver and Messel (2013) and Soland (2013) included models projecting college matriculation. Future similar predictive studies should also examine college matriculation outcomes, as well as college graduation outcomes.

In this study, students who earned any one of five diploma types were considered to be *graduates*. This was done to accomplish two separate aims, one of which was philosophical and one of which was practical. First, two of the diploma types are associated with students receiving special education services. The students who were classified as non-graduates in this work included students with disabilities. It was only prudent to provide pathways to include students with disabilities as graduates as well. Fifty or fewer students each earned an IB diploma or Modified Standard Diploma in this sample. There would have been too few cases in each cell to explore the ability of variables to predict the type of diploma earned with subsample sizes this small. Future studies with larger sample sizes should consider employing a multinomial outcome that accounts for different diploma types. Table 40 describes the different diploma types, as well as the requirements associated with earning each.

Table 40.

Types of High School Diplomas and Associated Requirements

Diploma Type	Requirements
Standard Diploma	Earn 22 course credits; pass 6 end-of-course standardized tests
Advanced Diploma	Earn 26 course credits; pass 9 end-of-course standardized tests
IB Diploma	Meet all of the requirements associated with the International Baccalaureate program
Modified Standard Diploma*	Earn 20 credits; no end-of-course standardized tests are required
Special Diploma*	Meet requirements of IEP, but do not qualify for other diploma types

*Note: No citation attributed to protect identify of the school district included in this study. * Only students with an IEP are eligible.*

Practical Implications for School Districts

Two implications emerge for school districts from the findings of this study. In line with the themes discussed in an earlier section, a tension exists between generalizability and local specificity. Findings that may hold true in multiple research settings might not in the context of a single school district. For budgetary reasons, it might be beneficial for a school district to conduct small replication studies using its own data to ensure that the findings from other empirical work, which they believe will generalize to their setting, actually do yield similar findings for their students.

This study, along with other similar studies in the literature (e.g. Mac Iver & Messel, 2013), demonstrates that by using data that are already collected for other purposes by school districts, models can be created to identify early warning indicators that school districts can use to craft interventions aimed at improving student outcomes. For example, the findings from the Series One ninth grade models indicate that passing ninth grade English (OR=2.35), passing ninth grade math (OR=1.58), and enrolling in at least one AP course (OR=4.09) are all strong predictors of increased odds of earning a high school diploma in four years. Students who are successful in ninth grade coursework are not only more likely to graduate per the findings in this study, but they are more likely to enroll in AP coursework, which further increases their odds of on-time graduation. Based on these findings, interventions could target middle school English and math instruction to ensure student readiness for success in their ninth grade coursework. Students with the prerequisite knowledge and skills to succeed in ninth grade coursework will be more likely to graduate from high school.

Limitations

There are four limitations associated with this work that are worth noting. First, the variables included in the predictive models created and tested in this study were all obtained from the administrative data from a single school district. Attention to context should be considered before generalizing these findings to other settings. Similarly, all of the predictors included in the models were obtained from administrative data. Other variables that have been found in the literature to be significant predictors of high school graduation status that would not be found in this type of a data set could not be included in the models. For example, empirical work has demonstrated a relationship between teenage alcohol and drug use, and graduation status (e.g. Ellickson, et al., 1996); however, this type of information would not be present in

administrative data from a school district. Finally, approximately one-fourth of the sample was excluded from analysis. This was primarily due to students transferring out of the school district prior to graduation, and their graduation status cannot be known using the available data. The outcomes for these students could differ from the outcomes of students who remained in the school district, and if that were the case, that would limit the inferences that could be made from the findings in this study.

The manner in which the dependent variable was operationalized limits the inferences that can be made from the findings. For the purpose of this study, all diploma types were uniformly considered a *diploma* and students who earned any of the five types of diplomas were uniformly considered to be *graduates.* Table 41 describes the types of diplomas issued in the state from which this sample was drawn. It is possible that differences exist in the experiences of students who earn different types of diplomas, and future research should unpack the outcome variable whenever possible.

Finally, the two graduation cohorts included in the sample used in this study differed substantially in terms of math test proficiency rates. The class of 2015's standardized math test proficiency rate was 82.9% whereas the the proficiency rate for the class of 2016 was only 58.7%. A *proficient* score on a state-issued standardized test in the state from which the sample in this study was drawn is a scaled score of 400 or better. The cohorts were tested together as a combined sample; however, this difference should be noted. When the two cohorts were together, passing every marking period of eighth grade math was found to be a significant predictor variable, whereas eighth grade math test scores were not significant. When the class of 2015 was tested alone, similar findings were reached. However, when the class of 2016 was tested alone, the opposite was found to be true; passing every marking period of eighth grade

math was no longer a significant predictor, but eighth grade math test scores were significant. Enrollment in Algebra I remained non-significant when the two cohorts were tested separately. Students in the class of 2016 had a more equal distribution in terms of math test score outcome, with almost 59% of students scoring at or above the cut score of 400 and just over 40% scoring below that. Fewer than 20% of students in the class of 2015 scored below this threshold. This difference between the two cohorts in terms of eighth grade math achievement could explain the dissimilarity in the associated *p*-values, and this is a limitation that should be considered when making inferences from the findings in this study.

REFERENCES

Abrams, L.M, & Haney, W. (2006). Accountability and the grade 9 to 10 transition: The impact on attrition and retention rates. In G. Orfield (Ed.), *Dropouts in America: Confronting the graduation rate crisis* (pp. 181-206). Cambridge, MA: Harvard Education Press

Alexander, K.L., Entwisle, D.R., & Horsey, C.S. (1997). From first grade forward: Early foundations of high school dropout. *Sociology of Education, 70*, 87-107.

Allensworth, E. M., & Easton, J. Q. (2005). *The on-track indicator as a predictor of high school graduation*. Retrieved from Consortium on Chicago School Research: http://consortium.uchicago.edu/sites/default/files/publications/p78.pdf

Andrew, M. (2014). The scarring effects of primary-grade retention? A study of cumulative advantage in the educational career. *Social Forces*, *93*(2), 653-686. http://www.dx.doi.org/10.1093/sf/sou074

Astone, N.M., & McLanahan, S.S. (1991). Family structure, parental practices, and high school completion. *American Sociological Review, 56*(3), 309-320. Retrieved from: http://www.jstor.org/stable/2096106.

Balfanz, R., Fox, J.H., Bridgeland, J.M., & McNaught, M. (2009). *Grad nation: A guidebook to help communities tackle the dropout crisis.* Washington, D.C.: America's Promise Alliance.

Balfanz, R., Herzog, L., & Mac Iver, D.J. (2007). Preventing student disengagement and keeping students on the graduation path in urban middle-grades schools: Early identification and effective interventions. *Educational Psychologist*, *42*(4), 223-235. http://dx.doi.org/10.1080/00461520701621079

Balfanz, R., & Herzog, L. (2005, March). *Keeping middle grades students on track to graduation: Initial analysis and implications.* Presentation given at the second Regional Middle Grades Symposium, Philadelphia, PA.

Barrington, B.L., & Hendricks, B. (1989). Differentiating characteristics of high school graduates, dropouts, and nongraduates. *The Journal of Educational Research, 82*(6), 309-319. Retrieved from: http://www.jstor.org/stable/40539658.

Battin-Pearson, S., Newcomb, M.D., Abbott, R.D., Hill, K.G., Catalano, R.F., & Hawkins, J.D. (2000). Predictors of early high school dropout: A test of five theories. *Journal of Educational Psychology, 92*(3), 568-582. http://dx.doi.org/10.1037/0022.0663.92.3.568.

Baylis, E., Burgos, T., Burton, T., Davis, M., Gladney, R., Hang, L., ... Overton, D. (2011). *Pushed out: Youth voices on the dropout crisis in Philadelphia*. Retrieved from Youth United For Change: http://youthunitedforchange.org/wp-content/uploads/2013/09/pushed-out.pdf

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: John Wiley & Sons.

Breslow, J. M. (2012). By the numbers: Dropping out of high school. Retrieved from http://www.pbs.org/wgbh/frontline/article/by-the-numbers-dropping-out-of-high-school

Bridgeland, J. M., Dilulio, Jr., J. J., & Morison, K. B. (2006). *The silent epidemic: Perspectiveof high school dropouts*. Retrieved from Civic Enterprises: http://files.eric.ed.gov/fulltext/ED513444.pdf

Bureau of Labor Statistics. (2016, April 28). *College enrollment and work activity of 2015 high school graduates* [Press release]. Retrieved from http://www.bls.gov/news.release/hsgec.nr0.htm

Cavanagh, S.E., Schiller, K.S., & Riegle-Crumb, C. (2006). Marital transitions, parenting, and schooling: Exploring the link between family-structure history and adolescents' academic status. *Sociology of Education, 79,* 329-354.

Christle, C.A., Jolivette, K., & Nelson, C.M. (2007). School characteristics related to high school dropout rates. *Remedial and Special Education, 28*(6), 325-339.

Coleman, J. S., Campbell, E., Hobson, C., McPartland, J., Mood, A., Weinfeld, F., & York, R. (1966). The Coleman Report. *Equality of Educational Opportunity*.

Cratty, D. (2012). Potential for significant reductions in dropout rates: Analysis of an entire 3rd grade state cohort. *Economics of Education Review*, *31*, 644-662. http://dx.doi.org/10.1016/j.econedurev.2012.04.001

Crowder, K., & South, S.J. (2003). Neighborhood distress and school dropout: The variable significance of community context. *Social Science Research, 32,* 659-698. http://www.dx.doi.org/10.1016/S0049-089X(03)00035-8

Dattalo, P. A. (2010). *Discriminant function analysis: An overview*. Unpublished manuscript.

DePaoli, J.L., Balfanz, R., & Bridgeland, J. (2016). *Building a grad nation: Progress and challenge in raising high school graduation rates.* Civic Enterprises: Baltimore, MD.

Doren, B., Murray, C., & Gau, J. M. (2014). Salient predictors of school dropout among secondary students with learning disabilities. *Learning Disabilities Research & Practice*, *29*(4), 150-159.

Dupere, V., Leventhal, T., Dion, E., Crosnoe, R., Archambault, I., & Janosz, M. (2015). Stressors and turning points in high school and dropout: A stress process, life course framework. *Review of Educational Research, 85*(4), 591-629. http://www.dx.doi.org/10/3102/0034654314559845.

Durkheim, E. (1982). *The rules of sociological method*. New York: The Free Press.

Edelbrock, C., & Achenbach, T. M. (1985). Manual for the teacher's report form and teacher version of the child behavior profile. *Burlington, VT: University of Vermont*.

Ellickson, P. L., McGuigan, K. A., Adams, V., Bell, R. M., & Hays, R. D. (1996). Teenagers and alcohol misuse in the United States: By any definition, it's a big problem. *Addiction*, *91*(10), 1489-1503.

Ensminger, M.E., Lamkin, R.P., & Jacobson, N. (1996). School leaving: A longitudinal perspective including neighborhood effects. *Child Development*, *67*(5), 2400-2416. http://www.jstor.org/stable/1131630.

Ensminger, M.E., & Slusarcick, A.L. (1992). Paths to high school graduation or dropout: A longitudinal study of a first-grade cohort. *Sociology of Education, 65*(2), 95-113. Retrieved from: http://www.jstor.org/stable/2112677.

Entwisle, D. R., Alexander, K. L., & Olson, L. S. (2004). Temporary as compared to permanent high school dropout. *Social Forces, 82*, 1181-1205.

Feden, P.D., & Vogel, R.M. (2003). *Methods of teaching: Applying cognitive science to promote student learning.* Boston: McGraw Hill.

Foster, E.M., & McLanahan, S. (1996). An illustration of the use of instrumental variables: Do neighborhood conductions affect a young person's chance of finishing high school? *Psychological Methods, 1*(3), 249-260.

Glynn, P. (2012). *Glynn help documents*. Retrieved from http://staff.washington.edu/glynn.

Goldschmidt, P., & Wang, J. (1999). When can schools affect dropout behavior? A longitudinal multilevel analysis. *American Educational Research Journal*, *36*(4), 715-738.

Hauser, R.M., Simmons, S.J., & Pager, D.I. (2004). High school dropout, race/ethnicity, and social background from the 1970 to the 1990. In G. Orefield (Ed.), *Dropouts in American: Confronting the graduation rate crisis* (pp. 85-106). Cambridge: Harvard Education Press.

Henry, K. L., Knight, K. E., & Thornberry, T. P. (2012). School disengagement as a predictor of dropout, delinquency, and problem substance use during adolescence and early adulthood. *Journal of Youth Adolescence, 41*, 156-166. http://dx.doi.org/10.1007/s10964-011-9665-3

Hernandez, D.J. (2011). *Double jeopardy: How third-grade reading skills and poverty influence high school graduation.* New York: Annie E. Casey Foundation.

Hosmer, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). *Applied logistic regression* (3rd ed.). Hoboken, NJ: John Wiley & Sons.

Howell, D.C. (2007). *Statistical methods for psychology* (6th ed.). Belmont, CA: Thompson Higher Educaiton.

Jacob, B.A., & Lefgren, L. (2004). Remedial education and student achievement: A regression-discontinuity analysis. *The Review of Economics and Statistics, 86*(1), 226-244.

Jimerson, S.R., & Ferguson, P. (2007). A longitudinal study of grade retention: Academic and behavioral outcomes of retained students through adolescence. *School Psychology Quarterly, 22*(3), 314-339. http://dx.doi.org/10.1037/1045-3830.22.3.314.

Jimerson, S. R., Anderson, G. E., & Whipple, A. D. (2002). 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), 441-457. http://dx.doi.org/10.1002/pits.10046

Jimerson, S., Egeland, B., Sroufe, L. A., & Carlson, B. (2000). A prospective longitudinal study of high school dropouts examining multiple predictors across development. *Journal of School Psychology*, *38*(6), 525-549. http://dx.doi.org/10.1016/S0022-4405(00)00051-0

Jimerson, S.R. (1999). On the failure of failure: Examining the association between early grade retention and education and employment outcomes during late adolescence. *Journal of School Psychology*, *37*(3), 243-272.

Jordan, W., Lara, J., & McPartland, J. M. (1996). Exploring the causes of early dropout among race-ethnic and gender groups. *Youth & Society*, *28*(1), 62-94.

Kim, K. (2013). Career trajectory in high school dropouts. *The Social Science Journal*, *50*, 306-312. http://dx.doi.org/10.1016/j.soscij.2013.03.005

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

Koretz, D. (2008). *Measuring up: What educational testing really tells us.* Cambridge: Harvard University Press.

Lagana, M.T. (2004). Protective factors for inner-city adolescents at risk for school dropout: Family factors and social support. *Children & Schools*, *26*(4), 211-220.

Lan, W., & Lanthier, R. (2003). Changes in students' academic performance and perceptions of school and self before dropping out of schools. *Journal of Education for Students Placed at Risk, 8*(3), 309-332. http://dx.doi.org/10.1207/S15327671ESPR0803_2

Langberg, J., & Ciolfi, A. (2016). *Suspended progress.* Legal Aid Justice Center: Charlottesville, VA.

Lee, T., Cornell, D., Gregory, A., & Fan, X. (2011). High suspension schools and dropout rates for Black and White students. *Education and Treatment of Children*, *34*(2), 167-192.

Levin, H. M. (1972). *The costs to the nation of inadequate education* (ED 064467). Washington, DC: Government Printing Office.

Losen, D.J. (2006). Graduation rate accountability under the No Child Left Behind Act and the disparate impact on students of color. In G. Orfield (Ed.), *Dropouts in America: Confronting the graduation rate crisis* (pp. 41-56). Cambridge, MA: Harvard Education Press

Lloyd, D. N. (1978). Prediction of school failure from third-grade data. *Educational and Psychological Measurement*, *38*, 1193-1200. Retrieved from http://epm.sagepub.com.proxy.library.vcu.edu/content/38/4/1193.full.pdf+html

Mac Iver, M. A., & Messel, M. (2012). *Predicting high school outcomes in the Baltimore City Schools*. Washington, DC: Council of the Great City Schools.

Mac Iver, M. A., & Messel, M. (2013). The ABCs of keeping on track to graduation: Research findings from Baltimore. *Journal of Education for Students Placed at Risk*, *18*, 50-67. http://dx.doi.org/10.1080/10824669.2013.745207

Makel, M. C., & Plucker, J. A. (2014). Facts are more important than novelty: Replication in the education sciences. *Education Researcher*, *20*(10), 1-13. http://dx.doi.org/10.3102/0013189X14545513

Mahoney, J.L., & Cairns, R.B. (1997). Do extracurricular activities protect against early school dropout? *Developmental Psychology*, *33*(2), 241-253.

Maxwell, J.A. (2013). *Qualitative research design: An interactive approach* (3rd ed.). Los Angeles: Sage.

McMillan, J.H. (2012). *Educational research: Fundamentals for the consumer* (6th ed.). Boston: Pearson.

McNeal, R.B. (1995). Extracurricular activities and high school dropouts. *Sociology of Education, 68*(1), 62-80. Retrieved from: http://www.jstor.org/stable/2112764.

McWilliams, A.E., Everett, P.C., & Bass, M. (2000). A model of underlying influences on students' decisions to drop out. *Journal of At-Risk Issues, 7*(1), 42-49.

Mitchell, M.L., & Jolley, J.M. (2013). *Research design explained*, (8th ed.). Belmont, CA: Wadsworth.

National Center for Education Statistics. (2015a). *State enrollments by race/ethnicity 2014-15.* Retrieved from https://nces.ed.gov/ccd/elsi/expresstables.aspx?bridge=quickFacts&tableid=19&level=State&year=2014-15

National Center for Education Statistics. (2015b). *Number and percentage of students in prekindergarten through grade 12 living in poverty, by control of school: 2000 through 2013.* Retrieved from https://nces.ed.gov/programs/digest/d14/tables/dt14_102.70.asp

National Center for Education Statistics. (2016a). *High school graduates, by sex and control of school: Selected years, 1869-70 through 2025-26*. Retrieved from https://nces.ed.gov/programs/digest/d15/tables/dt15_219.10.asp

National Center for Education Statistics. (2016b). *Digest of education statistics, 2015* (NCES 2016-014). Retrieved from https://nces.ed.gov/fastfacts/display.asp?id=64

National Research Council. (2011). *High school dropout, graduation, and completion rates: Better data, better measures, better decisions*. The National Academies Press: Washington, D.C.

Neild, R. C., Stoner-Eby, S., & Furstenberg, F. (2008). Connecting entrance and departure: The transition to ninth grade and high school dropout. *Education and Urban Society*, *40*(5), 543-569. http://dx.doi.org/10.1177/0013124508316438

Neild, R.C., & Balfanz, R. (2006). *Unfulfilled promise: The dimensions and characteristics of Philadelphia's dropout crisis, 2000-2005.* Baltimore: Center for Social Organization of Schools, Johns Hopkins University.

Oakes, J., & Guiton, G. (1995). Matchmaking: The dynamics of high school tracking decisions. *American Educational Research Journal, 32*(1), 3-33.

Orfield, G. (2006). Losing our future: Minority youth left out. In G. Orfield (Ed.), *Dropouts in America: Confronting the graduation rate crisis* (pp. 1-12). Cambridge: Harvard University Press.

Ou, S., & Reynolds, A.J. (2008). Predictors of educational attainment in the Chicago Longitudinal Study. *School Psychology Quarterly, 23*(2), 199-229. http://dx.doi.org/10.1037/1045-3830.23.2.199.

Owens, B.R., Morris, J.D., & Lieberman, M.G. (2001). A model to predict high school dropouts in a rural Florida school district. *The Journal of At-Risk Issues, 7*(3), 12-17.

Painter, G., & Levine, D.I. (2000). Family structure and youths' outcomes. *The Journal of Human Resources*, *35*(3), 524-549.

Parr, A. K., & Bonitz, V. S. (2015). Role of family background, student behaviors, and school-related beliefs in predicting high school dropout. *Journal of Educational Research*, *108*, 504-514. http://dx.doi.org/10.1080/00220671.2014.917256

Powell, B., & Steelman, L. C. (1993). The educational benefits of being spaced out: Sibship density and educational progress. *American Sociological Review*, 367-381.

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

Reschly, A.L., & Christenson, S. (2006). Prediction of dropout among students with mild disabilities – A case for the inclusion of student engagement variables. *Remedial and Special Education, 27,* 276-292.

Rumberger, R. W. (2011). *Dropping out: Why students drop out of high school and what can be done about it*. Cambridge: Harvard University Press.

Rumberger, R.W., & Rotermund, S. (2008). *What happened to dropouts from the high school class of 2004?* Stastical Brief 10. Santa Barbara: California Dropout Research Project, University of California. Retrieved from: http://www.cdrp.ucsb.edu/pubs_stetbriefs.htm

Rumberger, R. W. (2006). Why students drop out of school. In G. Orfield (Ed.), *Dropouts in America: Confronting the graduation rate crisis* (pp. 131-155). Cambridge, MA: Harvard Education Press.

Rumberger, R.W., & Palardy, G.J. (2005). Test scores, dropout rates, and transfer rates as alternative indicators of high school performance. *American Educational Research Journal, 42*(1), 3-42.

Rumberger, R. W., & Larson, K. A. (1998). Student mobility and the increased risk of high school dropout. *American Journal of Education, 107*(1), 1-35.

Rumberger, R.W. (1987). High school dropouts: A review of issues and evidence. *Review of Educational Research, 57*(2), 101-121. Retrieved from: http://www.jstor.org/stable/1170232.

Rumberger, R.W. (1983). Dropping out of high school: The influence of race, sex, and family background. *American Educational Research Journal, 20*(2), 199-220.

Saunders, M., Silver, D., & Zarate, E. (2008). *The impact of high schools on student achievement within the Los Angeles Unified School District.* Los Angeles: University of California All Campus Consortium on Research for Diversity.

Silver, D., Saunders, M., & Zarate, E. (2008). *What factors predict high school graduation in the Los Angeles Unified School District.* Santa Barbara, CA: California Dropout Research Project. Retrieved from: http://www.lmri.ucsb.edu/dropouts

Slavin, R.E. (2006). *Educational psychology: Theory and practice* (8th ed.). Boston: Pearson.

Snijders, T., & Bosker, R. (1999). *Multilevel analysis: An introduction to basic and advanced multilevel modeling*. Thousand Oaks, CA: Sage.

Soland, J. (2013). Predicting high school graduation and college enrollment: Comparing early warning indicator data and teacher intuition. *Journal of Education for Students Placed at Risk*, *18*, 233-262. http://dx.doi.org/10.1080/10824669.2013.833047.

Song, C., Benin, M., & Glick, J. (2012). Dropping out of high school: The effects of family structure and family transitions. *Journal of Divorce & Remarriage, 53*, 18-33. http://dx.doi.org/10.1080/10502556.2012.635964.

Stroup, A.L., & Robins, L.N. (1972). Elementary school predictors of high school dropout among Black males. *Sociology of Education, 45*(2), 212-222. Retrieved from: http://www.jstor.org/stable/2112008.

Swanson, C. B. (2006). Sketching a portrait of public high school graduation. In G. Orfield (Ed.), *Dropouts in America: Confronting the graduation rate crisis* (pp. 31-40). Cambridge, MA: Harvard Education Press

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th ed.). Boston: Pearson.

Thorndike, R.M, & Thorndike-Christ, T. (2010). *Measurement and evaluation in psychology and education* (8th ed.). Boston: Pearson.

Tobin, T.J., & Sugai, G.M. (1999). Using sixth-grade school records to predict school violence, chronic discipline problems, and high school outcomes. *Journal of Emotional and Behavioral Disorders*, *7*(1), 40-53.

U.S. Census Bureau (2015). *Annual estimates of the resident population by sex, race alone or in combination, and Hispanic origin for the United States, states, and counties: April 1, 2010 to July 1, 2015.* Retrieved from: https://factfinder.census.gov/faces/tableservices/jsf/pages/productview.xhtml?src=bkmk

U.S. Department of Education (2017). *Every Student Succeeds Act high school graduation rate non-regulatory guidance.* Retrieved from: https://www2.ed.gov/policy/elsec/leg/essa/essagradrateguidance.pdf

U.S. Department of Education (1988). *Eighth grade questionnaire: NELS 88.* Retrieved from: https://nces.ed.gov/surveys/nels88/pdf/01_BY_Student.pdf

Virginia Department of Education (2016). *Graduation requirements*. Retrieved from: http://www.doe.virginia.gov/instruction/graduation/standard.shtml

Warren, J.R., Hoffman, E., & Andrew, M. (2014). Patterns and trends in grade retention rates in the United States, 1995-2010. *Educational Researcher, 43*(9), 433-443. http://dx.doi.org/10.3102/0013189X14563599

Werblow, J., Urick, A., & Duesbery, L. (2013). On the wrong track: How tracking is associated with dropping out of high school. *Equity & Excellence in Education, 46*(2), 270-284. http://www.dx.doi.org/10.1080/10665684.2013.779168.

Werblow, J., & Duesbery, L. (2009). The impact of high school size on math achievement and dropout rate. *The High School Journal, 92*(3), 14-23.

Willingham, D.T. (2009). *Why don't students like school? A cognitive scientist answers questions about how the mind works and what it means for the classroom.* San Francisco: Jossey-Bass.

Zeigler, K., & Camarota, S.A. (2014). *One in five U.S. residents speaks foreign language at home, record 61.8 million.* Retrieved from http://cis.org/record-one-in-five-us-residents-speaks-language-other-than-english-at-home

Zvoch, K. (2006). Freshman year dropouts: Interactions between student and school characteristics and student dropout status. *Journal of Education for Students Placed at Risk, 11*(1), 97-117. http://dx.doi.org/10/1207/s15327671espr1101_6

APPENDICES

Appendix A.

Model estimates for Series One Eighth Grade Full Model

Variable	Coefficient	Robust SE	*p*-value
Intercept	3.1320	5.3290	.565
white	- .3566	.1433	.013
latino	- .7765	.3642	.033
other_race	.2258	.1924	.241
male	- .2820	.0769	.000
sped	1.5594	.2580	.000
overage_8	- .8912	.1255	.000
msmob	- .4094	.1569	.009
frl	- .4813	.1787	.007
eng8course_pass	.4409	.2434	.070
math8course_pass	.7420	.1996	.000
eng8_testsc	.0112	.0014	.000
math8_testsc	.0014	.0022	.531
algebra_1	.0214	.1772	.904
absent8_10	- .4799	.1365	.000
absent8_15	- .5299	.1013	.000
suspend8_ever	- .6949	.1362	.000
suspend8_3	- .2906	.1080	.007
sch_engtest	-4.9618	5.2878	.348
sch_frl	-2.8619	2.1776	.189

Appendix B.

Model estimates for Series One Ninth Grade Full Model

Variable	Coefficient	Robust SE	*p*-value
Intercept	1.8170	1.5827	.251
white	- .1268	.6251	.839
latino	- .7694	.1971	.000
other_race	.1917	.7911	.808
male	- .3003	.1134	.008
sped	1.5064	.1755	.000
overage_9	-1.1603	.1058	.000
hsmob	- .2299	.2446	.347
frl	.3287	.6907	.634
eng9course_pass	.8543	.2915	.003
math9course_pass	.4551	.1408	.001
math9_testsc	.0059	.0052	.258
AP_1	1.4080	.1802	.000
AP_3	1.3261	.7273	.068
absent9_10	- .7035	.3425	.040
absent9_15	- .5562	.3406	.102
suspend9_ever	- .5943	.3385	.079
suspend9_3	.0140	.3668	.969
hs_mathtest	-1.4931	1.1401	.190
hs_frl	-2.2170	2.0876	.288

Appendix C.

Model estimates for Series Two Full Model

Variable	Coefficient	Robust SE	*p*-value
Intercept	3.5385	1.8640	.058
white	- .1448	.2580	.575
latino	- .2916	.3394	.390
other_race	- .1719	.9418	.855
male	- .2906	.0716	.000
sped	1.5334	.2173	.000
overage_9	-1.1469	.0554	.000
hsmob	- .2287	.2535	.367
frl	- .6616	.3516	.060
eng9course_pass	.8668	.2933	.003
math9course_pass	.5083	.1898	.007
math9_testsc	.0079	.0044	.072
AP_1	1.3624	.1856	.000
AP_3	1.2669	.5969	.034
absent9_10	- .6894	.2883	.017
absent9_15	- .4794	.3905	.220
suspend9_ever	- .6900	.2201	.002
suspend9_3	.0943	.2344	.687
zip_med_income	- .0001	.0000	.000
zip_twoparent_perc	.0591	.0164	.000
zip_hsgradrate	- .0524	.0346	.130

Appendix D.

Series 1 Eighth Grade Model – Correctly Classified vs. Misclassified Cases

	Correctly Classified	Misclassified
African American	63.7%	75.1%
White	22.4	12.8
Latino	5.6	6.6
Other Race	8.3	5.5
Male	45.5	55.5
Special Education	7.1	13.0
FRL	63.7	86.6

Note: Values represent the percent of each column's category that apply to the variable to the left. For example, 63.3% of the cases that were correctly classified were for students identified as being African American.

Appendix E.

Series 1 Ninth Grade Model – Correctly Classified vs. Misclassified Cases

	Correctly Classified	Misclassified
African American	64.7%	78.5%
White	22.7	12.6
Latino	3.3	4.4
Other Race	9.3	4.9
Male	43.9	55.1
Special Education	7.1	10.0
FRL	63.2	88.3

Note: Values represent the percent of each column's category that apply to the variable to the left. For example, 64.7% of the cases that were correctly classified were for students identified as being African American.

Appendix F.

Series 2 – Correctly Classified vs. Misclassified Cases

	Correctly Classified	Misclassified
African American	64.9%	77.3%
White	22.5	13.9
Latino	3.4	3.6
Other Race	9.2	5.2
Male	44.3	52.6
Special Education	7.1	10.0
FRL	63.5	86.5

Note: Values represent the percent of each column's category that apply to the variable to the left. For example, 64.9% of the cases that were correctly classified were for students identified as being African American.